PEDAGOGISCH-DIDACTISCH HANDELEN
VAN DOCENTEN IN HET VOORTGEZET ONDERWIJS

YVETTE SOL

Beoordelingscommissie:
Prof.dr. M. Brekelmans (voorzitter)
Prof.dr. P.J. den Brok
Dr. E. Roelofs
Prof.dr. P.R.J. Simons
Prof.dr. M.L.L. Volman

© 2012 Y.B. Sol
ISBN: 978-90-5335-529-9
Druk: RidderPrint, Ridderkerk

PEDAGOGISCH-DIDACTISCH HANDELEN
VAN DOCENTEN IN HET VOORTGEZET ONDERWIJS

Pedagogical behaviour of teachers in secondary education
(with a summary in English)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag
van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit
van het college voor promoties in het openbaar te verdedigen
op dinsdag 8 mei 2012 des ochtends te 10.30 uur

door

Yvette Belinda Sol

geboren op 30 november 1961
te Amersfoort

Promotoren: Prof.dr. K.M. Stokking
Prof.dr. J.D. Vermunt

Een aantal onderzoeken waarvan in dit proefschrift verslag wordt gedaan, is uitgevoerd met financiële steun vanuit de landelijke subsidieregeling Kortlopend Onderwijsonderzoek.

Voorwoord

Als kind was ik al gefascineerd door alles in de wereld om me heen en wilde ik graag het naadje van de kous weten. Het verhaal gaat dat ik menig docent op de basisschool tot wanhoop heb gebracht met mijn aanhoudende reeks vragen. Het gebied waar mijn interesse naar uitging was breed en gevarieerd en spitste zich uiteindelijk aan het einde van mijn middelbare schoolperiode toe op zowel het medisch-biologische als het psychosociale en communicatieve vlak. In een tijd waarin het maatschappelijk gezien vooral verstandig leek een vak te kiezen, leek de opleiding fysiotherapie destijds een goede keuze. Tijdens het uitvoeren van mijn werkzaamheden als fysiotherapeute heb ik een deel van mijn nieuwsgierigheid kunnen bevredigen en niet alleen veel geleerd over diverse aspecten van ziekte en gezondheid, maar ook over het belang van contact en intermenselijke communicatie. Vanuit mijn werk als fysiotherapeute heb ik in contact met collega's en patiënten veel bijzondere momenten beleefd, die mij mede als persoon hebben gevormd.
In de periode dat ik als kinderfysiotherapeute werkzaam was in het speciaal basisonderwijs, verschoof mijn aandacht en interesse voor motorisch leren meer naar leren in het algemeen. Door de zowel positieve als negatieve ervaringen die mijn eigen kinderen en wij zelf daarbij als ouders opdeden in het reguliere basisonderwijs, raakte ik geboeid door de verschillen in het handelen van leerkrachten en de impact daarvan op leerlingen. Dit motiveerde mij in 2000 om de deeltijdopleiding Onderwijskunde te gaan volgen. Dat deze beslissing uiteindelijk tot een loopbaanswitch zou leiden en zou uitmonden in het schrijven van een proefschrift over het handelen van docenten, had ik destijds niet kunnen voorzien. Achteraf gezien is het misschien niet verwonderlijk dat ik dit alsnog ben gaan doen, als je het uitvoeren van een promotietraject opvat als een uitvergrote vorm van nieuwsgierig zijn en precies willen weten hoe het zit.

Dit proefschrift was echter niet tot stand gekomen op basis van vragen vanuit alleen mijn eigen nieuwsgierigheid. De vragen die centraal staan in dit proefschrift zijn gebaseerd op en afgeleid van vragen en opmerkingen van een aantal schoolleiders van scholen in het voortgezet onderwijs die bezig zijn vernieuwingen door te voeren. Voor het tot stand komen van dit proefschrift ben ik de volgende personen veel dank verschuldigd: Henk Zijlstra, Jeanine Vlastuin, Dave Drossaert, Han Doove, Ron Dorreboom, Elly Loman, Loes Lauteslager, Marcel Ter Vergert en Johan Capel. Bedankt voor jullie vragen, aanwijzingen en suggesties, maar ook voor het vertrouwen, de openheid en de plezierige samenwerking gedurende de diverse projecten in de afgelopen jaren.
Graag bedank ik ook de docenten op de scholen die hebben meegewerkt. Ik vond het een voorrecht jullie visies, opvattingen, concrete gedrag en achterliggende intenties in kaart te mogen brengen. Jullie enorme betrokkenheid bij de leerlingen en het onderwijs heeft op mij veel indruk gemaakt. Tevens bedank ik alle leerlingen op de scholen die aan de onderzoeken hebben meegewerkt en hun docenten feedback hebben gegeven.

Gedurende de afgelopen jaren heeft een aantal mensen tijdens diverse soorten bijeenkomsten en contacten op mijn werk met mij meegeleefd en meegedacht en daarmee op heel verschillende manieren ervoor gezorgd dat ik geïnspireerd en gemotiveerd aan het werk kon blijven. Hoewel ik jullie niet allemaal bij name kan en zal noemen, wil ik jullie graag bedanken voor wat jullie voor mij hebben betekend. Dank jullie wel voor jullie steun.

In het bijzonder wil ik op deze plaats mijn beide promotoren, Karel Stokking en Jan Vermunt, bedanken voor de begeleiding bij het tot stand komen van dit proefschrift.

Karel, jouw opmerking destijds bij mijn afstuderen als Onderwijskundige dat je in mij een onderzoeker zag, heeft terugkijkend wellicht onbewust bijgedragen aan bepaalde keuzes die ik in mijn loopbaan heb gemaakt. Na de switch in mijn loopbaan en de overstap naar de Universiteit, waarbij jij een faciliterende rol hebt gespeeld, heb jij steeds voor mij als een mentor gefungeerd, op diverse terreinen. Je was als dagelijks begeleider nauw bij het tot stand komen van het proefschrift betrokken. Ik heb geboft dat je het leuk vond om aan het eind van je loopbaan weer meer contact met het onderwijsveld te hebben, wat erin resulteerde dat je me een aantal malen hebt vergezeld bij de bezoeken aan de scholen. In de manier waarop jij vragen en onderliggende thema's in een breder perspectief kon plaatsen, werd niet alleen jouw brede kennisbasis en ervaring als onderzoeker en adviseur zichtbaar, maar ook jouw enorme betrokkenheid bij de onderwijspraktijk. Ik heb daarbij veel van je kunnen leren. Jouw betrokkenheid bij onderwijs werd voor mij opnieuw zichtbaar bij het ontwikkelen en verzorgen van onderdelen voor de Academische Lerarenopleiding Primair Onderwijs (ALPO), trajecten waarbinnen ik ook als docent intensief met je heb mogen samenwerken. Mede dankzij de kansen die ik kreeg en de ervaringen die ik kon opdoen, heb ik me de laatste jaren niet alleen als onderzoeker, maar ook als adviseur en docent verder kunnen ontwikkelen. Dank je wel voor je vertrouwen, loyaliteit, begeleiding en feedback in de afgelopen jaren. Ik hoop dat we de komende tijd de samenwerking en de inspirerende gesprekken nog verder kunnen voortzetten.

Jan, jij was als tweede promotor iets meer op afstand bij dit traject betrokken. Bedankt voor je goede adviezen en feedback op diverse onderdelen gedurende het proces en bij het tot stand komen van het proefschrift als geheel. Jouw kritische opmerkingen en inhoudelijke suggesties hebben bij de revisie van de verschillende hoofdstukken tot duidelijke verbeteringen geleid.

En last but not least wil ik met nadruk een aantal mensen aan mijn thuisfront bedanken: Walter, Robin, Jorik, mijn ouders en andere familieleden en vrienden: zonder jullie nadrukkelijke steun, en relativeringsvermogen op zijn tijd, had ik dit intensieve traject niet kunnen afronden. Robin en Jorik: jullie zijn ieder als persoon voor mij steeds een grote bron van inspiratie. Ik vind het heel bijzonder dat jullie nu mijn paranimfen willen zijn, omdat het voor mij de cirkel rond maakt in wat wij voor elkaar kunnen en willen betekenen.

Februari 2012 Yvette Sol

Inhoud

1. Inleiding 1

2. Onderzoeken op verzoek van scholen 9
2.1 De aard van de onderzoeken, de rol van de onderzoeker en de deelnemende scholen 9
2.2 De aanleiding, de terugkoppeling en het vervolg per onderzoek 10

3. Pedagogisch-didactisch handelen van docenten in kaart gebracht 15
3.1 Probleemstelling 15
3.2 De dimensie van het pedagogisch-didactisch handelen 16
3.3 Voorwaarden vanuit het beoogde gebruik van het instrument 26
3.4 Opzet en uitvoering van het empirisch onderzoek 30
3.5 Resultaten 40
3.6 De ontwikkeling van een verkorte versie van het instrument 48
3.7 Conclusies en discussie: waarde en gebruiksmogelijkheden van het instrument 52

4. Bijdragen aan validering van de verkorte vragenlijst (PDH) 59
4.1 Inleiding 59
4.2 Kader en vraagstelling 60
4.3 Opzet en uitvoering van het onderzoek 64
4.4 Resultaten 68
4.5 Conclusies en discussie 91

5. Nader onderzoek naar geven van feedback 101
5.1 Inleiding 101
5.2 Feedback in de wetenschappelijke literatuur 102
5.3 Opzet en uitvoering van het onderzoek 116
5.4 Resultaten 126
5.5 Conclusies en discussie 140

6. Nader onderzoek naar leerzaam beoordelen 147
6.1 Inleiding 147
6.2 Relaties tussen taken, beoordelingen en leren: theorie en onderzoek 148
6.3 Opzet en uitvoering van het onderzoek 157
6.4 Resultaten 166
6.5 Conclusies en discussie 190

7. Fundering van gewenst docenthandelen 199
7.1 Inleiding 199
7.2 Opzet en uitvoering van het onderzoek 200
7.3 Resultaten 209
7.4 Conclusies en discussie 238

8. Van huidig handelen naar gewenst handelen 245
8.1 Inleiding 245
8.2 Hebben van een visie op de gewenste ontwikkeling 245
8.3 Leiding geven aan het veranderingsproces 246
8.4 Kunnen omgaan met datafeedback 258
8.5 Conclusie 263

9. Conclusies en discussie 265

Literatuur 283

Samenvatting 297

Summary 307

Curriculum Vitae 317

1 Inleiding

Het voortgezet onderwijs heeft de laatste twee decennia te maken gekregen met een aantal vernieuwingen. Voor een deel werden deze geïnitieerd vanuit de overheid, maar voor een deel ook door schoolleiders zelf in overleg met diverse belanghebbenden (schoolbesturen, docenten, ouders, leerlingen). Scholen maakten daarbij gebruik van de ruimte die zij van overheidswege kregen voor het voeren van eigen onderwijskundig beleid. Deze ontwikkeling heeft ertoe geleid dat scholen onderling zijn gaan verschillen, bijvoorbeeld in de mate waarin zij al dan niet bepaalde vernieuwingen nastreven of waarin zij de mede door henzelf beoogde vernieuwingen inmiddels gedeeltelijk of al geheel hebben doorgevoerd.
Het invoeren van vernieuwingen heeft gevolgen voor allerlei aspecten binnen het onderwijs, niet in de laatste plaats voor het pedagogisch-didactisch handelen van docenten. Dit pedagogisch-didactisch handelen van docenten in vernieuwende scholen staat in dit proefschrift centraal.
Een groep schoolleiders wilde in 2007 meer zicht krijgen op het huidige pedagogisch-didactisch handelen van docenten in relatie tot de door hen beoogde vernieuwingen. Daarnaast wilden zij ook inzicht krijgen in de percepties van het gewenste handelen van leerlingen en docenten. Dit gewenste handelen bleek echter niet altijd even duidelijk beschreven of binnen de context van de school gedeeld (zie ook Sol & Stokking, 2008b en 2010).
Het in kaart brengen van het huidige en het gewenste handelen bleek geen eenvoudige zaak. Instrumenten om het pedagogisch-didactische repertoire van docenten in de volle breedte in kaart te brengen, bleken niet voorhanden.
De verschillende onderzoeken die worden beschreven in dit proefschrift maken onderdeel uit van een serie onderzoeken naar (aspecten van) het huidig en gewenst pedagogisch-didactisch handelen van docenten. Deze onderzoeken zijn uitgevoerd in de periode van 2007 tot en met 2011.

Context van vernieuwingen

Schoolleiders in het voortgezet onderwijs zijn de afgelopen 10-15 jaar steeds meer initiatieven gaan nemen om het onderwijs op hun school te vernieuwen. Zij doen dit naar aanleiding van landelijk beleid, maatschappelijke veranderingen, schoolinterne ontwikkelingen en nieuwe wetenschappelijke inzichten in leren en leeromgevingen. Scholen hebben in deze periode ook ruimte gekregen voor het voeren van een eigen onderwijskundig beleid. Dit wordt ook in toenemende mate van hen verwacht en scholen moeten zich als gevolg daarvan ook steeds meer extern verantwoorden (zie bijvoorbeeld Inspectie van het Onderwijs, 2011).

Vanaf het midden van de 90-er jaren is in het landelijke beleid met betrekking tot het voortgezet onderwijs en in het beleid van scholen en ondersteunende instellingen een aantal vernieuwingen op de agenda gezet. Deze hadden zowel betrekking op wat leerlingen zouden moeten leren als op de manier waarop dat zou moeten gebeuren en op de veranderingen in het curriculum en het handelen van docenten die daarvoor nodig werden geacht. Centraal stond het bevorderen van meer actief en zelfstandig leren, met meer ruimte voor het maken van eigen keuzes, meer leren in samenwerking tussen leerlingen, en meer leren in leergebieden die de grenzen van traditionele schoolvakken overstijgen.

Met deze vernieuwingen werden meerdere doelen tegelijk nagestreefd: motiveren van leerlingen, verbeteren van leerprocessen, en bevorderen dat leerlingen algemene vaardigheden leren die ze tijdens hun verdere loopbaan nodig zullen hebben (zelfstandig problemen oplossen, maken van keuzes, samenwerken, interdisciplinair werken). De maatschappelijke ontwikkelingen die voor deze veranderingen aanleiding waren en de veranderingen die daarop in scholen plaatsvonden, leidden tot toenemende herkenning en erkenning van de verschillen tussen leerlingen. Dit differentiatie-probleem en het daaruit voortvloeiende beroep op de competenties en inzet van docenten om elke leerling optimaal uit te dagen en te begeleiden, zijn tot op de dag van vandaag actueel.

De keuzes die scholen bij het invoeren van vernieuwingen maken, hebben onder meer consequenties voor het te voeren beleid, het inrichten van de leeromgeving en het pedagogisch-didactisch handelen van de docenten in relatie tot de leerlingen. Deze keuzen zijn voor een deel principieel van aard en betreffen verschillende terreinen: conceptueel (welke visies op kennis en leren), pedagogisch (welke ruimte voor leerlingen om hun eigen keuzes te maken en eigen kennis te ontwikkelen), cultureel (welke rollen en verantwoordelijkheden van docenten en leerlingen, hoe om te gaan met toenemende diversiteit), en politiek (aan welke eisen te voldoen, aan wie verantwoording af te leggen) (Windshitl, 2002).

De gemaakte keuzes moeten ook worden geplaatst tegen de achtergrond van de op dat moment overheersende opvattingen over wat goed onderwijs is, zowel maatschappelijk gezien als ook in het wetenschappelijke debat. De opkomst van het (sociaal) constructivisme vanaf 1985 heeft grote invloed gehad op de geldende opvattingen over kennis en leren en op de vraag welke inrichting van de leeromgeving het leren kan ondersteunen. Kennistheoretische en leerpsychologische opvattingen vanuit het (sociaal-)constructivisme hebben sindsdien, naast opvattingen uit de cognitieve psychologie, ook hun weg gevonden naar het onderwijs en hebben een belangrijke rol gespeeld bij het doordenken van en vormgeven van vernieuwende onderwijsconcepten en leeromgevingen.

In 2006 en 2007 werden door wetenschappers vanuit de cognitieve psychologie de eerste kritische kanttekeningen geplaatst bij de manier waarop in het onderwijs constructivistische uitgangspunten op kennis- en leertheoretisch niveau werden vertaald naar didactisch niveau (Kirschner, Sweller & Clark, 2006). De mate van

begeleide instructie bij het aanleren van nieuwe kennis bleek een belangrijk discussiepunt, evenals de mate waarin studenten of leerlingen zelf ontdekkend leren vanuit complexe taken die in betekenisvolle contexten worden aangeboden. In de discussies binnen de wetenschappelijke literatuur werd zichtbaar dat er verschillende visies bestaan op gewenste uitgangspunten en aandachtspunten als het gaat om didactische aspecten in het onderwijs en het inrichten van leeromgevingen (zie onder meer Hmelo-Silver, Golan Duncan & Chinn, 2007; Schmidt, Loyens, Van Gog & Paas, 2007; Sweller, Kirschner & Clark, 2007).

Ook de maatschappelijke context was sinds 2008 aan verandering onderhevig. De commissie Dijsselbloem heeft zich in haar rapport in 2008 onder meer kritisch uitgelaten over het gevoerde beleid in de jaren negentig van de vorige eeuw ten aanzien van de invoering van vernieuwingen in het onderwijs, het ontbreken van een wetenschappelijke onderbouwing daarvoor en de tegenvallende effecten op de onderwijsresultaten (Dijsselbloem, 2008). Kritische geluiden over de onderwijskwaliteit komen ook vanuit de media. Andere ontwikkelingen, zoals de relatief lagere prestaties van Nederlandse leerlingen in de PISA-ranglijsten en de ondermaatse prestaties van PABO-studenten, hebben ertoe geleid dat momenteel de aandacht weer meer wordt gericht op het overbrengen van kennis en dat wordt gestreefd naar betere prestaties. De overheid benadrukt meer dan voorheen het belang van opbrengstgericht werken (Volman, 2011; Inspectie van het Onderwijs, 2011) en de druk op scholen om zich te verantwoorden is alleen maar toegenomen. Naast meer aandacht voor opbrengstgericht werken worden ook bepaalde pedagogische-didactische aspecten binnen het onderwijsleerproces en het handelen van docenten benadrukt (Inspectie van het Onderwijs, 2011). De relevantie van het systematisch in kaart brengen van het huidig en gewenst handelen van docenten in relatie tot de doelen die worden beoogd (vanuit de overheid en vanuit het onderwijsconcept van de school) neemt daarmee alleen maar toe.

Typen vernieuwingen

De aard van de veranderingen die scholen willen doorvoeren als zij hun onderwijs willen vernieuwen, verschilt tussen scholen. De thema's voor vernieuwing die schoolleiders op hun school agenderen, zijn steeds een keuze uit verschillende opties die zich aandienen.

Teurlings, Van Wolput en Vermeulen (2006) onderscheiden bij hedendaagse onderwijsvernieuwingen drie centrale uitgangspunten: een activerende leeromgeving met accent op zelfstandig leren, werken met betekenisvolle en authentieke contexten, en samenwerking tussen leerlingen. Ze beschrijven hoe deze uitgangspunten voortkomen uit bepaalde wetenschappelijke maar vooral ook maatschappelijke ontwikkelingen en ontwikkelingen in het onderwijs zelf. Uit de door Teurlings en collega's bestudeerde onderzoeken blijkt dat waar de genoemde uitgangspunten in praktijk worden gebracht, een aantal positieve effecten op het leren van de leerlingen optreden.

Rondom de invoering van de vernieuwing van de basisvorming is een indeling gemaakt in vier gradaties van vernieuwing, aangeduid als scenario's 1 t/m 4. Hierbij werd gekeken naar de inhoud van het programma en het curriculum, de inrichting van de leeromgeving, de organisatie van de lessen, de leerdoelen, de mate waarin leerlingen keuzes kunnen maken en aandacht wordt geschonken aan zelfstandig leren, en de manier waarop docenten samenwerken (Diephuis & Van Kasteren, 2003). Binnen deze typologie verschillen de eerste twee en de laatste twee scenario's van elkaar in die zin dat de eerstgenoemde proberen de bestaande en traditionele structuur flexibeler te maken, terwijl de laatstgenoemde vertrekken vanuit een heel nieuw idee over de organisatie van het onderwijs, waardoor de vernieuwing bijna alle aspecten van het onderwijs betreft. De verschillende scenario's worden hieronder toegelicht.

Scholen die dichtbij het traditionele onderwijs blijven, maar naast het aanbieden van onderwijs in de schoolvakken leerlingen een paar uur per week laten werken aan vakoverstijgende projecten of thema's behoren binnen deze indeling tot scenario 1. Docenten zijn binnen dit scenario nog georganiseerd in vaksecties en werken vooral op individuele basis samen. Het basisritme van de school wordt slechts doorbroken door ruimte in het rooster voor enkele blokuren (al dan niet wekelijks), waarin leerlingen kunnen werken in projecten.

Scholen die hierin een stap verder gaan, worden getypeerd als behorend tot scenario 2. Deze scholen maken meer tijd in het rooster vrij voor vakoverstijgende projecten dan de scholen in scenario 1, waardoor de verhouding tussen reguliere vakgebonden lestijd en tijd die wordt bestemd voor projecten verschuift ten gunste van het laatste. Er ontstaat zodoende voor leerlingen meer afwisseling in het lesprogramma. Binnen de projecten werken docenten onderling samen. Naast het aanbieden van vakinhoudelijke kennis en -vaardigheden is er ook aandacht voor ontwikkeling van bepaalde vak-overstijgende vaardigheden (bijvoorbeeld het leren samenwerken). Opdrachten binnen de projecttijd worden zoveel mogelijk aangeboden in betekenisvolle contexten.

Scholen die zijn te typeren met scenario 3 hebben ook de structuur van een rooster niet losgelaten, maar geven les in blokuren van 70 tot 90 minuten. Vakken worden geïntegreerd aangeboden in grotere gehelen dan in de scenario's 1 en 2 gebruikelijk is. Dagdelen worden ingeroosterd voor vakoverstijgende projecten. Leerlingen hebben meer keuzemogelijkheden binnen opdrachten of bij de indeling van hun tijd en maken hun eigen planning. Het stimuleren van actief leren van leerlingen wordt belangrijk geacht. Daarbij is ook nadrukkelijk aandacht voor de ontwikkeling van vak-overstijgende vaardigheden. In deze setting werken docenten intensief samen, los van vaksecties, en hebben ze tevens een coachende rol naar de leerlingen.

Bij het meest vergaande scenario (scenario 4) wordt gewerkt met grote groepen leerlingen in stamgroepen en wordt het onderwijs aangeboden in leergebieden en thema's. Docenten hebben meerdere rollen (instructeur/expert, coach, ontwikkelaar) en werken intensief samen in een team. Leerlingen hebben veel keuzemogelijkheden, worden aangesproken op hun eigen verantwoordelijkheid en werken veelal zelfstandig aan taken. Behalve aan cognitieve en affectieve leerdoelen wordt veel aandacht besteed

aan persoonlijke groei en de ontwikkeling van vakoverstijgende vaardigheden en competenties van leerlingen.

Scholen kunnen zich, naast het onderscheid dat in bovenstaande scenario's is bedoeld, ook op een andere manier onderscheiden, bijvoorbeeld als cultuurschool. Ook kan sterk de nadruk komen te liggen op de school als sociale gemeenschap waarin bepaalde vormingsidealen centraal staan, of kunnen relaties met de wereld buiten de school tot voornaam aandachtspunt worden gemaakt.
Scholen combineren al doende vaak verschillende ambities en hun visie op onderwijs blijkt dan veelal te zijn gebaseerd op meerdere bronnen: pedagogisch, cognitivistisch, constructivistisch en motivatiepsychologisch (Sol & Stokking, 2008b; zie ook de hoofdstukken 3 en 7). Als bij leerlingen te bereiken doelen worden vaak genoemd: het bevorderen van zelfstandigheid, verdraagzaamheid, respect en verantwoordelijkheid, en op een meer concreet niveau het bevorderen van vaardigheden voor zelfstandig leren en samenwerken en het maken van eigen keuzes.
Onderzoek in 2008 van de VO-raad bij 159 scholen maakte duidelijk dat van de bevraagde scholen 7% behoorde tot scenario 1, 42% tot scenario 2, 28% tot scenario 3 en 23% tot scenario 4 (VO-raad, 2008).
Elke keuze voor verandering of vernieuwing heeft in meerdere of mindere mate implicaties voor het handelen van docenten en daarom betrekken de meeste schoolleiders hun docenten op de een of andere manier bij de besluitvorming over de richting die de school inslaat en de manier waarop de beoogde vernieuwingen vorm zullen krijgen.
Het is voor docenten daarbij niet altijd duidelijk wat van hen wordt verwacht in de nieuwe situatie (zie hiervoor onder meer Bolhuis, 2003; Sol & Stokking, 2008a en 2010). Een concrete uitwerking van het gewenste pedagogisch-didactisch handelen is vaak niet voorhanden. Gewenst handelen van docenten wordt vaak alleen beschreven in globale termen zoals het vergroten van de aandacht voor het leerproces van de leerlingen, het inspelen op verschillen tussen leerlingen, het bieden van meer keuzemogelijkheden, en het vervullen van een meer begeleidende rol. Wat docenten precies geacht worden te doen in relatie tot de leerlingen in termen van pedagogisch-didactisch handelen blijft dan impliciet en door het ontbreken van de dialoog daarover voor verschillende interpretaties vatbaar.

Vraagstelling en onderzoeksvragen

Onderstaande vraagstelling en onderzoeksvragen zijn geformuleerd als overkoepelend en organiserend kader in dit proefschrift.
De vraagstelling luidt:
In hoeverre en onder welke voorwaarden is het mogelijk het huidige en gewenste pedagogisch-didactisch handelen van docenten in scholen voor voortgezet onderwijs met een al dan niet vernieuwend onderwijsconcept in kaart te brengen, op een

zodanige wijze dat dit wetenschappelijk verantwoord gebeurt en wetenschappelijk relevante inzichten oplevert en tegelijk scholen, schoolleiders, teamleiders en docenten kan ondersteunen bij hun verdere ontwikkeling en vernieuwing?
De onderzoeksvragen zijn:
1 Waaruit bestaat het pedagogisch-didactisch handelen van docenten in het voortgezet onderwijs en hoe kan dit op valide, betrouwbare en bruikbare wijze worden gemeten?
2 In welke mate vertonen docenten dit handelen?
3 Hoe is de variatie in dit handelen en in hoeverre hangt deze samen met kenmerken van docenten (geslacht, leeftijd, ervaring, school) en leerlingen (geslacht, leeftijd, leerjaar, schooltype)?
4 Welk handelen wordt gewenst geacht en door wie en op welke gronden?
5 Wat zijn de verschillen tussen huidig en gewenst handelen en hoe kunnen deze worden overbrugd?

Typering en doelstelling van dit proefschrift

In dit proefschrift zijn aspecten van het handelen van docenten die bijdragen aan (motivatie en) leren geïnventariseerd vanuit wetenschappelijke literatuur. Vervolgens zijn deze aspecten geordend en benut om instrumenten te ontwikkelen om het huidig en gewenst handelen van docenten in kaart te brengen. De gegevens die met de ontwikkelde instrumenten zijn verzameld, zijn teruggekoppeld aan de betrokkenen als vorm van datafeedback. Aan de hand daarvan kon een vergelijking worden gemaakt tussen het geconstateerde handelen en het handelen waarvan blijkens de literatuur bijdragen kunnen worden verwacht aan (motivatie en) leren en konden aanbevelingen worden geformuleerd. Tevens konden de uitkomsten van deze vergelijking in verband worden gebracht met het beoogde handelen op basis van de gewenste vernieuwingen op de scholen.
Het gebruik van de ontwikkelde instrumenten en de terugkoppeling van de gegevens en de gesprekken daarover beogen bij te dragen aan de vergroting van het conceptuele repertoire van betrokkenen, waardoor zij (meer gedifferentieerd) kunnen kijken naar de (eigen) praktijk en woorden tot hun beschikking krijgen om hierover in dialoog te gaan met zichzelf en met elkaar.
Het vergroten van het conceptuele repertoire van schoolleiders en docenten als professionals en het voeren van een inhoudelijke dialoog over gewenst pedagogisch-didactisch handelen heeft tot doel inzicht te bieden in waarom iets wel of niet gewenst kan zijn in relatie tot de doelen die worden nagestreefd. Het beoogt schoolleiders en docenten ondersteuning te bieden bij de afweging welke vernieuwingen met betrekking tot het pedagogisch-didactisch handelen gewenst kunnen zijn of op welke manier het gewenst handelen vervolgens nader geconcretiseerd kan worden. Schoolleiders en docenten kunnen hun afwegingen hierdoor mogelijk ook of meer baseren op een theoretische kennisbasis en niet alleen op empirische generalisaties over wat werkt in de (eigen) praktijk.

Dit proefschrift beoogt door de gekozen invalshoek, de intensieve samenwerking met scholen, de ontwikkelde instrumenten en de interactie met de betrokkenen aan de hand van datafeedback en aanvullende gesprekken met schoolleiders, teamleiders en docenten een bijdrage te leveren aan het overbruggen van de kloof tussen onderwijsonderzoek en de onderwijspraktijk (zie hiervoor ook Broekkamp & Van Hout-Wolters, 2006; Van Tartwijk, 2011). In dit proefschrift wordt niet geprobeerd een verklaring te geven voor het geconstateerde handelen en ook is geen experimenteel onderzoek uitgevoerd.

Opbouw van dit proefschrift

Hoofdstuk 2 gaat in op de aanvragen voor onderzoek die een aantal vernieuwende scholen hebben ingediend en die zijn uitgevoerd gedurende de periode 2007 tot en met 2011 met betrekking tot het in kaart brengen van het pedagogisch-didactisch handelen van docenten.
In hoofdstuk 3 wordt de ontwikkeling en validering beschreven van een instrument om het huidige en gewenste pedagogisch-didactisch handelen van docenten in de volle breedte in kaart te brengen, gezien vanuit het perspectief van zowel de docenten als de leerlingen. Dit onderzoek is uitgevoerd op verzoek van zes scholen.
Hoofdstuk 4 richt zich op het vaststellen van een aantal aspecten van de psychometrische kwaliteit van een op verzoek van de schoolleiders van de hiervoor genoemde scholen gemaakte verkorte versie van het instrument (de PDH genoemd). Tevens wordt ingegaan op de variatie in het handelen tussen scholen en tussen docenten, en de verschillen in antwoorden tussen docenten respectievelijk leerlingen met verschillende persoonlijke kenmerken.
De hoofdstukken 5 en 6 gaan nader in op twee in het instrument onderscheiden taakgebieden van docenten, namelijk het geven van feedback aan leerlingen en het beoordelen van het werk en de prestaties van leerlingen. Deze twee onderzoeken zijn uitgevoerd op verzoek van een aantal schoolleiders die, naar aanleiding van de in hoofdstuk 3 beschreven uitkomsten, behoefte hadden aan verdieping op deze gebieden met het oog op de gewenste ontwikkeling op hun school.
Hoofdstuk 7 rapporteert over een onderzoek dat is uitgevoerd op verzoek van een aantal schoolleiders die aan de hand van de gerapporteerde bevindingen over het pedagogisch-didactisch handelen van docenten op hun scholen constateerden dat dit handelen tussen de docenten op hun scholen onderling sterk varieerde en niet duidelijk overeenstemde met de onderwijsconcepten van hun scholen. Ook waren zij benieuwd naar de percepties van stakeholders buiten de school van het gewenste handelen van docenten en naar een vergelijking tussen de onderwijsconcepten, het landelijke beleid en de wetenschappelijke literatuur.
Hoofdstuk 8 gaat in op een aantal condities die van belang zijn bij processen van verandering, vernieuwing en verbetering van pedagogische-didactische aspecten van het onderwijs en bij benutting van resultaten van onderzoek daarbij. In dit hoofdstuk

wordt tevens verslag gedaan van bevindingen en ervaringen bij een onderzoek naar een veranderingsproces op een school.
In hoofdstuk 9 staan de conclusies van het proefschrift als geheel weergegeven, gevolgd door een discussiegedeelte met daarin onder meer een kenschets van de mogelijke betekenis van de resultaten voor de onderwijspraktijk, enkele beperkingen van de uitgevoerde onderzoeken, de bijdrage aan de wetenschappelijke discussie, en opties voor vervolgonderzoek.

2 Onderzoeken op verzoek van scholen

2.1 De aard van de onderzoeken, de rol van de onderzoeker en de deelnemende scholen

De aard en aanvragen en het subsidiekader van het Kortlopend Onderwijsonderzoek
Vanaf het najaar van 2006 vormde het feit dat een aantal schoolleiders van vernieuwende scholen onvoldoende zicht had op wat docenten feitelijk doen in hun interacties met de leerlingen en ook geen duidelijk inzicht in wat docenten zouden moeten doen volgens hun nieuwe onderwijsconcepten, voor deze schoolleiders aanleiding om onderzoek aan te vragen. Met dit onderzoek wilden zij meer zicht krijgen op het huidige pedagogisch-didactisch handelen van hun docenten en meer inzicht in het gewenste handelen in de specifieke context van hun school. Zij vroegen in opeenvolgende jaren (deels gezamenlijk, in verschillende combinaties van scholen) subsidies aan in het kader van de landelijke regeling voor Kortlopend Onderwijs-onderzoek. Gedurende vier jaar zijn in totaal zes van zulke onderzoeken aangevraagd en gehonoreerd. Deze onderzoeken zijn uitgevoerd door de auteur, werkzaam bij de Universiteit Utrecht, in de periode 2007-2011.
Bij de uitvoering van deze onderzoeken is ernaar gestreefd drie doelen te combineren: a) bruikbare antwoorden verkrijgen op de vragen die de schoolleiders stelden, b) daarbij benutten van een scala aan onderzoeksmethoden en –technieken die geschikt waren om de benodigde gegevens te verzamelen en te analyseren en aan de te stellen wetenschappelijke kwaliteitseisen voldeden, en c) de vraagstelling en de gebruikte instrumenten in elk onderzoek baseren op relevante conceptuele en theoretische kaders. Daarmee waren deze onderzoeken zowel praktijkgericht als wetenschappelijk van aard. Dit droeg tevens bij aan een meer algemene bruikbaarheid van de opbrengsten, dat wil zeggen los van de specifieke contexten van deze scholen. Met deze aanpak werd tevens voldaan aan de eisen van de betreffende subsidieregeling. Van elk onderzoek is een rapport verschenen, dat aan de deelnemende scholen is verstrekt en tevens landelijk is verspreid en beschikbaar gesteld aan het onderwijsveld.

De rol van de onderzoeker
De rol van de onderzoeker binnen deze onderzoeken was niet die van betrokken participant, maar van professionele buitenstaander, dat wil zeggen onafhankelijk, objectief en neutraal. Bij elk onderzoek werd na een fase van verkenning van de problematiek en de vraagstelling in overleg met de betreffende schoolleiders gezocht naar een passende theoretische analyse en inbedding daarvan in de relevante wetenschappelijke literatuur. Vervolgens werden op grond daarvan instrumenten

ontwikkeld en gegevens verzameld op de scholen. Na de fase van analyse en inhoudelijke interpretatie werden de uitkomsten teruggekoppeld als vorm van datafeedback. Daarbij werden aanvullend gesprekken gevoerd met de diverse betrokkenen op de scholen (schoolleiders, teamleiders en docenten) over de interpretatie en relevantie van de uitkomsten voor hun eigen onderwijspraktijk.

De scholen binnen deze onderzoeken

In totaal zijn bij de uitvoering van deze onderzoeken acht scholen voor voortgezet onderwijs betrokken geweest. Zij waren in wisselende combinaties de aanvragers van de onderzoeken, in een aantal gevallen vergezeld van andere scholen die wel aanvrager waren maar niet zelf aan de uitvoering deelnamen.

De deelnemende scholen kunnen als volgt worden getypeerd. Eén school is een vmbo-school, drie scholen zijn brede scholengemeenschappen (van vmbo t/m gymnasium), en vier scholen zijn havo-vwo-scholen. Van deze scholen had één havo-vwo school een meer traditioneel onderwijsconcept (ergens tussen scenario 1 en 2), één havo-vwo school was kortgeleden begonnen met het invoeren van een aantal vernieuwende aspecten (passend bij scenario 3), één havo-vwo school had een afdeling met een sterk vernieuwend onderwijsconcept (scenario 4) maar verder een meer traditioneel concept (tussen scenario 1 en 2), en vijf scholen (de vmbo-school, de vierde havo-vwo school en de drie brede scholengemeenschappen) hanteerden een sterk vernieuwend onderwijsconcept (scenario 4).

In de rapportage over elk onderzoek in de hierna volgende hoofdstukken zijn de deelnemende scholen steeds aangeduid met letters (in willekeurige volgorde), om de privacy van de scholen te waarborgen. Aangezien de deelnemende scholen per onderzoek varieerden maar steeds opnieuw de letters A, B, etc. zijn gebruikt, kunnen de uitkomsten van de scholen die zijn aangeduid met de letters A, B, etc. tussen de hoofdstukken niet met elkaar worden vergeleken, omdat dit steeds andere scholen betreft.

2.2. De aanleiding, de terugkoppeling en het vervolg per onderzoek

De ontwikkeling van een instrument, en een verkorte versie daarvan

De aanleiding

De schoolleiders van zes vernieuwende scholen hadden zelf al gesignaleerd dat sprake was van verschillende opvattingen en gedragingen bij docenten. Het in kaart brengen hiervan leek zinvol en was ook een manier om docenten meer bij het beoogde vernieuwingsproces te betrekken. Naast dat zij meer zicht wilden krijgen op hoe docenten zelf, en eigenlijk ook leerlingen, tegen het beoogde handelen op basis van het vernieuwende onderwijsconcept aankeken, waren zij geïnteresseerd in hoe het

gewenste handelen zich nu zou verhouden tot het huidige handelen en welke verschillen hierbij bestonden tussen docenten, klassen en scholen.
De vraagstelling van de schoolleiders was aanvankelijk gericht op de vaardigheden die worden gevraagd van docenten die werken op scholen met een vernieuwend concept, maar in het kader van het onderzoek dat in het volgende hoofdstuk wordt beschreven, was het niet mogelijk te onderzoeken over welke vaardigheden docenten zouden moeten beschikken en wat ze zouden moeten kunnen, hooguit om in kaart te brengen wat docenten doen en zouden willen doen. Tijdens de eerste gesprekken met de schoolleiders bleek ook dat zij eigenlijk vooral daarin waren geïnteresseerd: het feitelijke gedrag van docenten en wat de betrokkenen (schoolleiders, docenten en leerlingen) zien als wenselijk gedrag.
Een louter op waarneembaar gedrag gerichte benadering leidt gemakkelijk tot het formuleren van enorme aantallen items, die alleen al door hun aantal niet hanteerbaar zijn en bij betrokkenen bovendien kunnen leiden tot de vaststelling dat daarmee niet wordt gedekt waar het eigenlijk om gaat. Op basis van theoretische overwegingen en de wensen van de schoolleiders is uiteindelijk in samenspraak met hen gekozen om te spreken over het handelen van docenten. Het focus van de vraagstelling van de schoolleiders lag op de vormgeving en de inrichting van het onderwijsproces op die momenten dat docenten in interactie zijn met leerlingen. Op basis hiervan is ervoor gekozen een instrument te ontwikkelen dat specifiek is gericht op het in kaart brengen van het algemeen pedagogisch-didactisch handelen van docenten. De ontwikkeling en validering van het instrument wordt beschreven in hoofdstuk 3. Tevens wordt in dit hoofdstuk ingegaan op de ontwikkeling van een verkorte versie van het instrument (de PDH).

De terugkoppeling en het vervolg
De oorspronkelijke versie van het ontwikkelde instrument bestond uit een vragenlijst gebaseerd op een indeling in taakgebieden met per taakgebied een aantal uitspraken.
De uitkomsten van de afnames van de ontwikkelde vragenlijst met taakgebieden bij docenten en leerlingen op zes scholen zoals in dit hoofdstuk beschreven, zijn op maat per school schriftelijk en mondeling teruggerapporteerd en besproken. In het rapport over het onderzoek waarop dit hoofdstuk is gebaseerd, is voor schoolleiders, teamleiders en docenten tevens een aantal gerichte aanbevelingen geformuleerd (Sol en Stokking, 2008b).
De op verzoek daarna ontwikkelde verkorte versie heeft de naam PDH gekregen (van Pedagogisch-Didactisch Handelen). Deze PDH is web-based toegankelijk gemaakt. De schoolleider van één van de scholen die heeft meegedaan aan het in hoofdstuk 3 beschreven onderzoek waarin de schriftelijke vragenlijst is ontwikkeld, wilde in het najaar van 2009 graag een nieuwe afname op de school realiseren en was bereid daarbij de PDH te gebruiken. Over de uitkomsten hiervan is op de school schriftelijk en mondeling teruggerapporteerd. In 2010 bleken de schoolleiders van nog drie andere scholen (van de zes scholen die aan de ontwikkeling van de vragenlijst hadden

meegedaan) belangstelling te hebben voor een afnameronde van de PDH op hun school. Dit is gerealiseerd in de periode van november 2010 tot januari 2011. Ook de uitkomsten hiervan zijn op maat per school schriftelijk en mondeling teruggerapporteerd. Ter verdere validering van de PDH zijn vervolgens op alle inmiddels beschikbare data van de verkorte versie nadere analyses uitgevoerd.
Alle hiervoor beschreven werkzaamheden op de betreffende scholen en van de onderzoekers in relatie tot de verkorte versie PDH (schriftelijke en mondelinge terugrapportage en bespreking op maat per school over de uitkomsten van de schriftelijke vragenlijst met taakgebieden, ontwikkeling van de verkorte versie, invullen daarvan door docenten en leerlingen op vier scholen, schriftelijke en mondelinge terugrapportage en bespreking op maat per school over de uitkomsten, en verdere analyses van alle data) zijn uitgevoerd zonder subsidie. Over deze analyses op de PDH-data en de uitkomsten daarvan, en over de aanpak en ervaringen met de schriftelijke en mondelinge terugrapportages naar en op de scholen, wordt verslag gedaan in hoofdstuk 4.

Nader onderzoek naar het geven van feedback

De aanleiding
Met behulp van het ontwikkelde instrument voor het in kaart brengen van het pedagogisch-didactisch handelen van docenten (zie hoofdstuk 3) heeft een aantal vernieuwende scholen informatie verkregen over het huidige repertoire van de docenten op de onderscheiden taakgebieden. Tijdens de terugkoppeling en de besprekingen hierover vroegen de schoolleiders zich af of docenten tijdens het begeleiden van leerlingen in situaties waarin die zelfstandig aan het werk zijn wel optimaal feedback geven (één van de taakgebieden). De schoolleiders van zeven scholen vroegen om een vervolgonderzoek om meer inzicht te krijgen in de manier waarop docenten feitelijk feedback geven. Zij wilden tevens zicht krijgen op de opvattingen van docenten over feedback in relatie tot het onderwijsconcept van de school, de wensen van de leerlingen, en de beschikbare wetenschappelijke kennis. Dit onderzoek is uitgevoerd in de periode juni 2008 – februari 2009 op vier scholen. Zie hoofdstuk 5 voor de beschrijving en de uitkomsten van het onderzoek.

De terugkoppeling en het vervolg
Na afronding van het onderzoek is aan elk van de deelnemende docenten een persoonlijk verslag toegestuurd over het doorlopen traject met daarin alle resultaten van het onderzoek bij die docent. Daaraan is door de onderzoekers een persoonlijk advies toegevoegd. Via presentaties en workshops op scholen zijn de uitkomsten van het onderzoek als geheel daarnaast gedeeld met schoolleiders, teamleiders en docenten. Tevens is het ontwikkelde materiaal ter beschikking gesteld en is een *job-aid* voor docenten gemaakt met daarop een samenvatting van relevante aspecten van feedback en een lijst met concrete docentactiviteiten op dit taakgebied. In het rapport over het

onderzoek waarop dit hoofdstuk is gebaseerd, is voor schoolleiders, teamleiders en docenten afsluitend een aantal gerichte aanbevelingen geformuleerd (zie voor het volledige rapport Sol & Stokking, 2009).

Nader onderzoek naar leerzaam beoordelen

De aanleiding
In een vervolgonderzoek, opnieuw op verzoek van een aantal schoolleiders en uitgevoerd in 2010, stond het beoordelen als docenttaak en als onderdeel van de leeromgeving en van het onderwijsproces centraal. Docenten laten leerlingen vaak werken aan taken, in de vorm van toetsen of opdrachten, waarvan de resultaten expliciet zullen worden beoordeeld.
De schoolleiders die vroegen om dit vervolgonderzoek wilden graag weten hoe beoordelingssituaties zo kunnen worden ingericht dat deze door leerlingen zoveel mogelijk als leersituaties kunnen worden ervaren en kunnen bijdragen aan het leerproces. Dit onderzoek is uitgevoerd in de periode juni 2010 – maart 2011 op drie scholen. Zie hoofdstuk 6 voor de beschrijving en opbrengsten van het onderzoek.

De terugkoppeling en het vervolg
Na afronding van het onderzoek is aan elk van de deelnemende docenten een persoonlijk verslag toegestuurd over het doorlopen traject met daarin alle resultaten van het onderzoek bij die docent. Daaraan is door de onderzoekers een persoonlijk advies toegevoegd. Via presentaties en workshops op scholen zijn de uitkomsten van het onderzoek gedeeld met schoolleiders, teamleiders en docenten. Tevens is het ontwikkelde materiaal ter beschikking gesteld en is een *job-aid* voor docenten gemaakt met een samenvatting van voor docenten relevante opbrengsten uit het onderzoek. In het rapport over het onderzoek waarop dit hoofdstuk is gebaseerd, zijn voor schoolleiders, teamleiders en docenten afsluitend gerichte aanbevelingen geformuleerd (zie voor het volledige rapport Sol & Stokking, 2011).

Nader onderzoek naar gewenst handelen van docenten

De aanleiding
In een presentatie voor een aantal schoolleiders over een onderzoek naar het pedagogisch-didactisch handelen van docenten (het onderzoek dat is beschreven in hoofdstuk 3) kwam naar voren dat dit handelen en de onderwijsconcepten van de scholen niet altijd duidelijk op elkaar aansluiten. Ook blijken beide niet duidelijk aan te sluiten bij wetenschappelijke inzichten over onderwijs en de vormgeving van leerprocessen die bijdragen aan de motivatie en het leren van leerlingen.
Hierdoor ontstond bij een aantal schoolleiders de behoefte aan verder onderzoek om meer inzicht te verkrijgen in het docenthandelen dat door schoolleiding, teamleiders, docenten, leerlingen en externe *stakeholders* wordt gewenst en de visies, opvattingen

en verwachtingen die daaraan ten grondslag liggen. Daarnaast wilden de schoolleiders een systematische vergelijking tussen dit gewenste docenthandelen en de op hun scholen uitgewerkte onderwijsconcepten, en relevante wetenschappelijke kennis. Tot slot vroeg men zich af hoe een en ander zich verhoudt tot het landelijke onderwijsbeleid.
Het onderzoek is uitgevoerd in de periode april 2009-maart 2010 op vier scholen. Zie hoofdstuk 7 voor de beschrijving en uitkomsten van het onderzoek.

De terugkoppeling en het vervolg
Na afronding van het onderzoek is aan elk van de schoolleiders van de bij de aanvraag betrokken scholen een rapport toegestuurd over het doorlopen traject met alle resultaten van het onderzoek. Daarin zijn tevens verschillende aanbevelingen geformuleerd (zie voor het volledige rapport Sol & Stokking, 2010). Via presentaties zijn de uitkomsten van het onderzoek gedeeld met de schoolleiders. Tevens is het ontwikkelde materiaal ter beschikking gesteld en is een *job-aid* gemaakt met een samenvatting van relevante opbrengsten uit het onderzoek voor schoolleiders en docenten.

Onderzoek naar een veranderingsproces op een school

Op verzoek van een school waar was besloten een bepaalde vernieuwing in te voeren is het veranderingsproces met onderzoek gevolgd, door middel van verheldering van de doelen en herhaalde mondelinge en schriftelijke bevraging van schoolleiding, docenten en leerlingen. De uitkomsten zijn op diverse momenten gedurende het proces teruggekoppeld aan betrokkenen en, mede op basis van relevante wetenschappelijke literatuur, voorzien van aanbevelingen. Dit bleek effecten te hebben op het verdere verloop. Zie hoofdstuk 8, in het bijzonder par. 8.3, voor een beknopte beschrijving.

3 Pedagogisch-didactisch handelen van docenten in kaart gebracht

3.1 Probleemstelling

Een aantal schoolleiders van scholen in het voortgezet onderwijs met een vernieuwend onderwijsconcept had behoefte aan een instrument om het pedagogisch-didactisch handelen van hun docenten in kaart te kunnen brengen. De doelstelling van dit onderzoek was daarmee het ontwikkelen van zo´n instrument. Een oriëntatie op reeds beschikbare instrumenten voor het breed in kaart brengen van het repertoire van docenten in het voortgezet onderwijs op het vlak van pedagogisch-didactisch handelen leerde dat een dergelijk instrument niet voorhanden was. Bestaande instrumenten richten zich op een andere dimensie van het handelen (bijvoorbeeld de Vragenlijst Interpersoonlijk Leraarsgedrag (VIL), Wubbels, Brekelmans & Hooymayers, 1991) of hebben betrekking op meerdere globale dimensies en bijbehorende competenties (bijvoorbeeld de SBL-competenties, Stichting Beroepskwaliteit Leraren, 2004). Andere instrumenten zijn voor een specifieke schoolsituatie ontwikkeld of zijn gericht op het in kaart brengen van bepaalde vakspecifieke handelingen of instructietechnieken. Daarom werd besloten een nieuw instrument te ontwikkelen.
De vraagstelling van het onderzoek luidde als volgt:

1. Hoe kan het pedagogisch-didactisch handelen van docenten in vernieuwende scholen voor voortgezet onderwijs op valide en betrouwbare wijze in kaart worden gebracht?
2. Wat zijn de psychometrische eigenschappen van het ontwikkelde instrument?
3. Aan welke voorwaarden moet een instrument voldoen om op scholen bruikbaar te zijn en voor ontwikkelingsdoeleinden te kunnen worden ingezet?

Gezien de aard van dit onderzoek kan het worden getypeerd als een instrumenteel-nomologisch onderzoek (De Groot, 1971). In een dergelijk onderzoek is de theoretisch-conceptuele analyse van het te meten begrip een onderdeel van het onderzoek (zie 3.2). De schoolleiders die het onderzoek aanvroegen, wilden het te ontwikkelen instrument kunnen inzetten voor ontwikkelingsdoeleinden. Daarom wordt ook expliciet aandacht besteed aan enkele voorwaarden met betrekking tot het beoogde gebruik die aan het instrument moeten worden gesteld (zie 3.3). Daarna worden beschreven de tijdens het onderzoek gevolgde werkwijze en gemaakte keuzen (3.4), de resulterende kenmerken van het ontwikkelde instrument (3.5) en de ontwikkeling van een verkorte versie (3.6).

3.2 De dimensie van het pedagogisch-didactisch docenthandelen

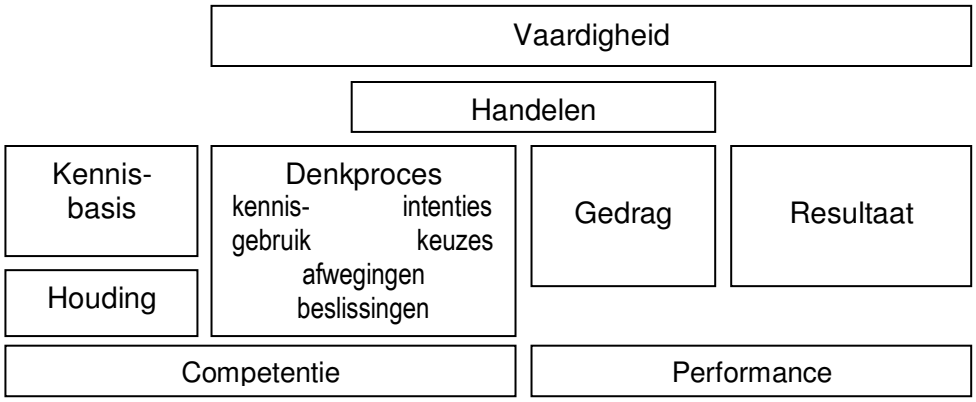
3.2.1 Inleiding

Om een instrument te kunnen ontwikkelen om het huidige en gewenste pedagogisch-didactisch handelen van docenten in kaart te kunnen brengen moet eerst worden vastgesteld wat daar wel en niet onder valt. Dit betreft enerzijds de afbakening van het begrip `handelen` ten opzichte van andere verwante begrippen en van andere aspecten in het takenpakket en het gedrag van docenten. Anderzijds moet worden vastgesteld wat de interne structuur is van het pedagogisch-didactisch handelen zelf. (Zie voor een vergelijkbare aanpak om een abstract begrip te verhelderen: Stoof, Martens, Van Merriënboer & Bastiaens, 2002).
In paragraaf 3.2.2 wordt ingegaan op de afbakening van het begrip handelen ten opzichte van andere begrippen. Daarna wordt het pedagogisch-didactisch handelen afgebakend ten opzichte van andere aspecten in het takenpakket van een docent. Vervolgens wordt in 3.2.3 de interne structuur van het pedagogisch-didactisch handelen in kaart gebracht. Dit heeft geresulteerd in een inhoudelijke indeling in een aantal zogenoemde taakgebieden.

3.2.2 Afbakening van het handelen van docenten ten opzichte van andere begrippen

Om te beginnen kan het begrip “handelen” worden afgebakend ten opzichte van verwante begrippen, zoals vaardigheden, competenties, kennis, maken van keuzen (nemen van professionele beslissingen), performance en gedrag. De hier voorgestelde ordening (zie figuur 3.1) is geënt op het procesmodel voor professioneel handelen zoals beschreven door Roelofs, Nijveldt en Beijaard (2008).
Roelofs en collega’s benadrukken dat met de invoering van het competentiebegrip in het onderwijs minder nadruk komt te liggen op geïsoleerde kenniselementen en gedrag van docenten en meer op integratie van kennis, opvattingen en vaardigheden. In een procesgericht model voor competent handelen van docenten leggen zij een relatie tussen verschillende aspecten die met het begrip “competentie” te maken hebben: de professionele basis (kennis, vaardigheden, opvattingen, persoonlijke eigenschappen), de inschattingen, afwegingen en beslissingen, het gedrag, en de resultaten van dat gedrag (Roelofs, et al., 2008).
De vier elementen in het model van Roelofs en collega’s (2008) komen terug in figuur 3.1, waarin visueel wordt weergegeven hoe de verschillende begrippen zich volgens ons tot elkaar verhouden. De horizontale posities van de begrippen representeren verticaal bekeken hun onderscheid en overlap. Een voorbeeld: het begrip vaardigheid (bovenste rij) omvat zowel denkprocessen (kennisgebruik, afwegingen en beslissingen, intenties en keuzen) in het tweede blokje van de vier blokjes in de derde rij, als het zichtbare gedrag als het resultaat daarvan (dat wil zeggen: of iemand vaardig is wordt

aan al deze drie elementen afgemeten). Het begrip handelen (in de tweede rij) omvat een gedeelte van het tweede blokje in de derde rij (namelijk de intenties en keuzes) en het zichtbare gedrag (het derde element), en dit komt tot uitdrukking in de lengte en plaats van het betreffende balkje.



Figuur 3.1: Het begrip "handelen" in relatie tot andere, verwante begrippen
De horizontale posities van de begrippen representeren verticaal bekeken hun onderscheid en overlap

We lichten het hier voorgestelde gebruik van begrippen nu inhoudelijk nader toe.
Als wordt gesproken over ontwikkeling van docentcompetenties, gaat het om het ontwikkelen van voldoende kennis, vaardigheden en een geschikte houding om adequaat en verantwoord te kunnen opereren als professional in verschillende situaties die zich binnen een onderwijscontext kunnen voordoen. Competenties kunnen daarmee worden beschouwd als het resultaat van een geheel aan kennis, opvattingen en vaardigheden die afhankelijk van de situatie worden omgezet in professioneel handelen (Reynolds, 1992; Roelofs et al., 2008; Shulman, 1987). Het begrip "competentie" refereert daarbij aan wat een persoon weet en kan en wat zichtbaar kan worden onder ideale omstandigheden; het begrip "performance" refereert aan feitelijk zichtbaar gedrag en het mede op basis daarvan behaalde resultaat in een specifieke situatie (Wood, 1987).
Het begrip "vaardigheid" kan verschillend worden opgevat (Stokking, 2001). In het algemeen gaan we ervan uit dat iemand vaardig is als hij een bepaalde handeling of taak snel en nauwkeurig kan uitvoeren en de uitvoering kan aanpassen aan wisselende omstandigheden. Vaardigheid kan worden gedefinieerd in termen van de eisen waaraan het resultaat of product moet voldoen. Een vaardigheid kan ook worden gezien als het (snel, precies, flexibel) uitvoeren van een aantal handelingen in een zekere volgorde om bepaalde problemen op te lossen, antwoorden te genereren of een product te maken. Bij beide onderscheidingen wordt vaardigheid gezien als een taakgerichte prestatie en komt het begrip "vaardigheid" grotendeels overeen met het hierboven beschreven begrip "performance".

Het begrip vaardigheid kan ook worden opgevat als een (verondersteld maar niet direct zichtbaar) vermogen of proces, nodig voor het leveren van bepaalde prestaties. De vaardigheid wordt dan afgeleid uit de waargenomen prestatie (handeling, resultaat of product). Deze bredere opvatting van vaardigheid ligt dicht bij het begrip competentie (Stokking, 2001).
De mate van verwevenheid tussen de begrippen "kennis" en "vaardigheid" is afhankelijk van de gekozen opvatting over vaardigheid, de aard en specificiteit van de taak, de fase in het leerproces en de al dan niet veranderende context. Kennis is bijvoorbeeld nodig bij de ontwikkeling van een vaardigheid (voor het opbouwen van een goed beeld van de uit te voeren taak of reeks handelingen), maar ook bij het kunnen benutten van verkregen feedback over het verloop van de handelingen in relatie tot het beoogde resultaat. Ook op momenten dat vaardigheden reeds geautomatiseerd zijn maar handelingen op problemen stuiten of de situatie of de context verandert, wordt het kunnen beschikken over een kennisbron weer belangrijk om adequaat te kunnen reageren.
Het vermogen om de kennisbasis, houdingsaspecten en gedragingen gecoördineerd in te zetten in specifieke situaties is voorwaardelijk voor adequaat en verantwoord handelen (Dwyer, 1994; Messick 1995a; Wood, 1987). De toegenomen aandacht voor de onderliggende beslisprocessen heeft zich vertaald in een tendens om bij het beoordelen van het handelen van docenten minder de nadruk te leggen op het vertonen van door deskundigen als bekwaam of effectief erkend gedrag en meer op het kunnen verantwoorden van het vertoonde gedrag in uiteenlopende situaties en het kunnen afwegen van alternatieven (Dwyer, 1994; Roelofs et al., 2008; Sol & Stokking, 2010). Daarbij wordt het belangrijk geacht en ligt het voor de hand het handelen en de verantwoording daarvan in verband te brengen met de beoogde doelen en de gevolgen voor de leerlingen (Shuell, 1993; Vermunt & Verloop, 1999).
Bij het inschatten, afwegen, nemen van beslissingen en nastreven van doelen speelt de kennisbasis van docenten een belangrijke rol. Het gaat hierbij vooral om de beschikbare persoonlijke praktijkkennis (Fenstermacher & Richardson, 1993; Verloop & Lowyck, 2009). Deze praktijkkennis is het resultaat van enerzijds formele en informele scholing en anderzijds eigen ervaringen en reflectie op deze ervaringen (Calderhead, 1989; Schön, 1983). Deze kennisbasis omvat zowel bewuste en expliciete kennis en (persoonlijke) opvattingen, als impliciete kennis, intuïties en emoties (Verloop et al., 2009). Al deze elementen kunnen het feitelijk handelen van de docent beïnvloeden. Daarbij kan onderscheid worden gemaakt tussen interactieve cognities en meer algemene cognities.
Interactieve cognities zijn cognities die aanwezig zijn tijdens het lesgeven zelf. Zij bestaan uit een combinatie van denkprocessen van diverse aard (vergelijk Verloop et al., 2009) en worden geactiveerd door bepaalde karakteristieken die in de situatie aanwezig zijn. Naarmate een docent vaker in eenzelfde situatie komt of meer ervaring heeft, krijgt het gedrag vaak een meer geautomatiseerd en routinematig karakter (Fiske

& Taylor, 1991). Dit wordt versterkt doordat tijd voor reflectie in een dergelijke interactieve onderwijssituatie vaak ontbreekt (Dolk, 1997).
Naast deze interactieve cognities spelen ook cognities een rol die niet direct aan bepaald gedrag zijn gekoppeld. Deze worden ook wel aangeduid als opvattingen, subjectieve theorieën, beliefs (Verloop et al., 2009; Verloop en Wubbels, 1994). Deze kunnen van heel diverse aard zijn en zijn meer permanent en op afstand in het denken aanwezig (Luttenberg, 2000). Zij vormen het referentiekader van de docent en fungeren, vaak onbewust en impliciet, als een filter waardoor informatie uit de omgeving of het eigen handelen op een bepaalde manier wordt gepercipieerd, geëvalueerd en eventueel aangepast (Kagan, 1992; Kelchtermans, 2007). Beide typen cognities spelen als onderdelen van de kennisbasis en het referentiekader van de docent een belangrijke rol bij het inschatten, afwegen, keuzes maken en beslissingen nemen over het eigen handelen.
Op basis van de voorgaande overwegingen wordt het handelen van docenten als uitgangspunt genomen voor de ontwikkeling van het instrument en wordt dit handelen als volgt omschreven:
Het handelen van docenten bestrijkt zowel concreet gedrag als ook de verwachtingen, inschattingen, afwegingen en beslissingen die aan dat gedrag ten grondslag liggen en de onderliggende intenties, dat wil zeggen dat wat de docent in het contact met de leerlingen nastreeft.

3.2.3 Pedagogisch-didactisch handelen: afbakening van andere dimensies en interne structuur

Afbakening van andere dimensies in het handelen van docenten

In de literatuur worden de volgende dimensies of perspectieven ten aanzien van het handelen van docenten genoemd: het vakdidactische, het leeractiviteiten-, het interpersoonlijke, het morele en het organisatorische perspectief (Brekelmans, Slegers & Fraser, 2000). Deze dimensies zijn geen onderdelen van het takenpakket van docenten in de zin dat daaruit, met het oog op het realiseren van een bepaald onderwijsconcept, een bepaalde keuze zou kunnen worden gemaakt. Het zijn facetten van het handelen die steeds tegelijk aan de orde zijn. Naast deze dimensies in het handelen wordt in de literatuur ook een aantal soorten kennis en opvattingen van docenten onderscheiden, zoals vakinhoudelijke kennis en kennis over vormen van instructie, curricula, materialen, leerlingen en dergelijke (zie bijvoorbeeld Den Brok, 2011; Shulman, 1986, 1987). Deze kennisbasis (zie figuur 3.1) maakt geen onderdeel uit van het te ontwikkelen instrument, maar kan wel bijdragen aan het proces van het maken van afwegingen en keuzen ten aanzien van het handelen, dat kan uitmonden in intenties die in het handelen tot uitdrukking kunnen komen.
In beleidsgerichte publicaties spreekt men over competenties of bekwaamheden en worden ook bekwaamheidseisen geformuleerd. De Stichting Beroepskwaliteit Leraren

onderscheidt zeven van zulke (domeinen van) competenties: interpersoonlijk, pedagogisch, vakinhoudelijk en didactisch, organisatorisch, samenwerken met collega's, samenwerken met de omgeving, en reflectie en ontwikkeling (Stichting Beroepskwaliteit Leraren, 2004). Bij de opstelling van bekwaamheden en bekwaamheidseisen wordt meestal onderscheid gemaakt tussen enerzijds vakinhoudelijke en vakdidactische kennis, anderzijds pedagogisch-didactisch kwaliteiten (zie bijvoorbeeld Ministerie van OC&W, 1993; Onderwijsraad, 2005a; Van Gennip & Vrieze, 2008).

Dit onderscheid, tussen vakspecifiek en algemeen pedagogisch-didactisch, sluit aan bij het onderscheid dat sommige auteurs maken in vakdidactische kennis en vaardigheden en algemene principes van effectief lesgeven (Bransford, Derry, Berliner & Hammerness, 2005; Shulman, 1987). Bepaalde auteurs bepleiten voornamelijk het belang van specifieke vakinhoudelijke en vakdidactische kennis en vaardigheden (bijvoorbeeld Grossman & Schoenfeld, 2005; Mayer, 2004a en 2004b) en minder of niet dat van algemene pedagogisch-didactische kwaliteiten. De Corte (2000) en Bransford en collega's (2005) benadrukken echter dat effectief onderwijs van docenten vraagt dat zij bij het ontwerpen van leeromgevingen en het beoordelen van leerresultaten niet alleen over vakinhoudelijke kennis beschikken maar ook meer in het algemeen kennis hebben over en inzicht in de volgende vier componenten: a) wat moet worden geleerd, b) hoe leerlingen dit kunnen leren, c) welke leeromgeving dit leren kan bevorderen, en d) hoe kan worden vastgesteld dat en wat er is geleerd. Deze vier componenten bestrijken zowel vakspecifieke als meer algemeen pedagogisch-didactische aspecten en daarmee wordt benadrukt dat beide van belang zijn.

In dit onderzoek is ervoor gekozen het pedagogische en het didactische in samenhang met elkaar te bekijken en een instrument te ontwikkelen waarin beide aspecten aan bod komen. Het pedagogisch-didactische handelen wordt vaak als één geheel beschouwd omdat de pedagogische (ook wel genoemd sociaal-affectieve) dimensie in het onderwijsleerproces als voorwaardelijk wordt gezien voor het stimuleren van het leren en de ontwikkeling van leerlingen (Van Driel, 2006). Ook nemen beide aspecten in onderlinge relatie in de onderwijsconcepten van scholen een belangrijke plaats in (Sol & Stokking, 2010).

Een andere manier om het pedagogisch-didactisch handelen zoals hier bedoeld uit te werken, is naar niveau van specificatie. In de literatuur wordt gewenst docenthandelen op verschillende niveaus uitgewerkt: in termen van rollen, te creëren leeromgevingen, taken, instructiestrategieën en specifieke acties en reacties in interactiesituaties (zoals het stellen van bepaalde soorten vragen, geven van hints, etc.) (Bolhuis, 2003; Brown & Campione, 1996; Collins, Brown & Newman, 1989; De Corte, 2000). Voor het doel van dit onderzoek zijn rollen (bijvoorbeeld: de docent als begeleider) te globaal en instructiestrategieën en typen interacties te specifiek, terwijl het creëren van een bepaalde leeromgeving vraagt om het vooraf ontwerpen, ontwikkelen en organiseren daarvan. Verder is ervan uitgegaan dat goede docenten in hun handelen geen vast patroon laten zien, maar alert en adaptief inspelen op het onderwijsleerproces zoals dit zich ontwikkelt (Shulman, 1987). Een belangrijke voorwaarde daarvoor is het

beschikken over een gedegen kennisbasis en een breed handelingsrepertoire. Dit benodigde handelingsrepertoire in termen van taken is als uitgangspunt genomen bij de ontwikkeling van het instrument.

De interne structuur van het begrip pedagogisch-didactisch handelen en een indeling in taakgebieden

De uitdrukking "pedagogisch-didactisch" wordt veel gebruikt, maar vaak zonder heldere definitie en zonder duidelijke afbakening. In de literatuur zijn hiervoor wel aanwijzingen te vinden. In dit onderzoek is gebruik gemaakt van wetenschappelijke literatuur vanuit drie verschillende invalshoeken: literatuur over algemeen didactische modellen, rapportages over meta-analyses van factoren die de effectiviteit van onderwijs beïnvloeden, en publicaties met theorie en onderzoek naar leren en leeromgevingen. Deze invalshoeken zijn als bronnen benut om nadere specificaties te verkrijgen van de dimensie pedagogisch-didactisch handelen en te komen tot een structuur van "taakgebieden". In de kolommen A, B en C in figuur 3.2 is aangegeven welke aanwijzingen voor een nadere specificatie van de dimensie pedagogisch-didactisch handelen dit heeft opgeleverd. Het onderscheiden van regels in figuur 3.2 was een iteratief proces, waarbij in een volgende kolom bepaalde regels bleken te moeten worden uitgesplitst waardoor in een eerdere kolom sommige aspecten of factoren in meer regels terug dienden te komen. We beschrijven nu de kolommen van links naar rechts.

Ten eerste is een aantal publicaties geraadpleegd (klassieke en meer recente) waarin didactische modellen worden gepresenteerd (De Corte, Geerligs, Lagerwey, Peters & Vandenberghe, 1976; Dochy, 1992; Van Gelder, Oudkerk, Pool, Peters & Sixma, 1973; Merrill, 1971, 2002; Valcke, 2007). De aspecten in die modellen die zijn gerelateerd aan het handelen van docenten zijn weergegeven in kolom A. Deze kolom representeert een min of meer klassieke onderwijskundige indeling (waarbij door het hierboven toegelichte iteratieve proces sommige aspecten in meerdere regels worden benoemd).

Ten tweede zijn reviewstudies en meta-analyses geraadpleegd waarin factoren worden benoemd die bijdragen aan de effectiviteit van het handelen van docenten (Campbell, Kyriades, Muijs & Robinson, 2004; Scheerens, 2007; Seidel & Shavelson, 2007; Shuell,1996). Deze factoren zijn opgenomen in kolom B. De meeste daarvan worden zowel door Scheerens (2007) als door Seidel en Shavelson (2007) genoemd. De onderscheidingen in kolom B blijken grotendeels aan te sluiten bij die in kolom A, waarbij de pedagogische en motivatiepsychologische aspecten van uitdaging en ondersteuning, in kolom B aangeduid met 'motivatie en relatie', bij een aantal onderdelen van kolom A van toepassing blijken te zijn en daarom op verschillende plaatsen in kolom B zijn opgenomen.

	A Aspecten van didactisch handelen in modellen	*B Factoren uit meta-analyses*	*C Aspecten uit theorie en onderzoek naar leren en leeromgevingen*	*D Aspecten in onderwijsconcepten van scholen en in landelijk beleid*	*E Rubrieken zoals voorgelegd aan docenten*	*F Uiteindelijke indeling in taakgebieden*
Ontwikkelwerk, lesvoorbereiding organisatie	Leerdoelen, leerstofaanbod, leermiddelen	Onderwijsconcept. leermateriaal, klassen-organisatie, leertijd	Leerdoelstellingen		Ontwikkelen van materialen en opdrachten, gebruik ICT	
Uitvoering in interactie met leerlingen		Leeromgeving	Pedagogisch leerklimaat Motiverend klimaat	Pedagogisch klimaat (veilig, leerlinggericht)	Pedagogisch en sociaal klimaat	1 Zorgen voor een goed leerklimaat
		Motivatie, relatie		Aandacht voor persoonlijke vorming (zelfstandigheid, verantwoordelijkheid)		2 Bijdragen aan persoonlijke vorming van leerlingen
	Didactisch handelen, instructie, werkvormen	Instructie	Geven van les, uitleg en instructie, Kennisconstructie	Instructie gericht op vergroten van kennis en vaardigheden (gevarieerd, activerend)	Instrueren, activeren, zelfstandig leren	3 Geven van les, uitleg en instructie
			Betekenisvolle context	Aanbieden van kennis in samenhang en betekenisvolle, contextrijke taken	Bijdragen aan samenhang	4 Leerlingen het nut en de samenhang laten zien
	Leeractiviteiten didact. handelen, instructie, werkv.	Activering, leerstrategieën	Krachtige leeromgeving Begeleiden, coachen	Aanbieden activerende leeromgeving	Begeleiden van leerlingen	5 Leerlingen leren zelf opdrachten aan te pakken
		Motivatie, relatie	Begeleiden, coachen kennis en vaardigheden	Aanbieden activerende leeromgeving Leerlingen motiveren en uitdagen Instructie over kennis en vaardigheden	Instrueren, activeren, zelfstandig leren	6 Leerlingen stimuleren en vaardigheden aanleren
	Vaststellen beginsituatie	Adaptief onderwijs	vaststellen beginsituatie Begeleiden, coachen	Begeleiden bij ontwikkeling	Inspelen op verschillen tussen leerlingen	7 Inspelen op verschillen tussen leerlingen
	Groeperingsvorm. Leeractiviteiten	Motivatie, relatie	Bevorderen zelfstandig leren. Begeleiden, coachen. Samenwerken	Leerlingen leren samenwerken, sociale vaardigheden	Samenwerken tussen leerlingen. Begeleiden van leerlingen	8 Leerlingen laten samenwerken en hen daarbij ondersteunen
	Leeractiviteiten	Activering, leerstrategieën, motivatie, relatie	Bevorderen zelfstandig leren. Begeleiden, coachen	Activerende leeromgeving, leerlingen leren zelfstandig te leren, bevorderen zelfsturing, leerling motiveren, uitdagen	Ruimte voor eigen keuzes van leerlingen. Begeleiden leerlingen	9 Leerlingen eigen keuzes laten maken en daarbij ondersteunen
	Leeractiviteiten	Activering, leerstrategieën		Activerende leeromgeving, leerlingen leren zelfstandig te leren, bevorderen zelfsturing	Instrueren, activeren, zelfstandig leren. Begeleiden van leerlingen	10 Leerlingen hun eigen leerproces laten plannen en bewaken
	Volgen van het leerproces	Regulatie en monitoring	Begeleiden, coachen	Begeleiden bij persoonlijke ontwikkeling en kennis en vaardigheden	Volgen van het leren en de ontwikkeling van leerlingen	11 Volgens hoe leerlingen leren en zich ontwikkelen
	Beoordelen	Evaluatie en feedback	Geven van feedback Beoordelen	Feedback geven op werk en gedrag (in gesprek gaan met individuele leerling)	Geven van feedback	12 Geven van feedback
	Toetsing en evaluatie	Evaluatie en feedback	Geven van feedback Beoordelen	Beoordelen van werk en gedrag (resultaten)	Beoordelen	13 Beoordelen van het werk en de prestaties van leerlingen
Samenwerking en professionele ontwikkeling				Samenwerking met collega's	Samenwerken in het docententeam	14a Samenwerken met andere leraren
						14b Samen leerlingen begeleiden
				Professionele ontwikkeling	Reflectie en persoonlijke ontwikkeling	15 Professionele ontwikkeling

Figuur 3.2 Interne structuur van pedagogisch-didactisch docenthandelen: gebruikte bronnen en ontwikkeling van indeling in taakgebieden

Ten derde is een aantal publicaties doorgenomen over leren en leeromgevingen. De daarin gehanteerde onderscheidingen blijken deels van iets andere aard te zijn dan die in de kolommen A en B, maar daaraan wel te kunnen worden gekoppeld. Het resultaat hiervan staat in kolom C. Van boven naar beneden gaat het hierbij om de volgende aspecten en invalshoeken:

- leerdoelen, doelstellingen: gericht op leerlingen en hun leerproces, kennis, vaardigheden, school en maatschappij (Bransford, 2000; Bransford et al., 2005; Windshitl, 2002);
- pedagogisch leerklimaat (Doyle, 1986; Houtveen & Reezicht, 2000; Stevens, Beekers, Evers, Wentzel & Van Werkhoven, 2004) en motiverend klimaat (Blumenfeld, Soloway, Marx, Kraijcik, Guzdal & Palincsar, 1991; Ryan & Deci; 2000);
- geven van les, uitleg en instructie (Brown en Campione, 1996; Elshout-Mohr & van Hout-Wolters, 1995; Elshout-Mohr, Van Hout-Wolters & Broekkamp, 1999; Oostdam, Peetsma & Blok, 2007; Windschitl, 2002) en constructie van kennis (Bransford, 2000; Bransford et al., 2007);
- werken met betekenisvolle contexten (Greeno & Collins, 1996; Windschitl, 2002);
- krachtige leeromgeving, en begeleiden en coachen (dit laatste bleek bij een aantal aspecten in de kolommen A en B van toepassing te zijn) (Bransford, 2000; Bransford et al., 2005; Brown en Campione, 1986; Collins et al., 1989; Veenman, 1992; Windshitl, 2002);
- bevorderen van zelfstandig leren (Bolhuis, 2003; Oostdam et al., 2007; Teurlings et al., 2006);
- stimuleren en benutten van samenwerking tussen leerlingen (Van Boxtel, 2000; Webb, 1997, 2009; Windshitl, 2002);
- formatieve en summatieve beoordeling (Crooks, 1988; Kluger & DeNisi, 1996; Natriello, 1987; Sadler, 1989, Shute, 2008).

In kolom D zijn aanwijzingen weergegeven voor specificatie van het pedagogisch-didactisch handelen vanuit een vierde bron, namelijk de onderwijsconcepten van scholen met een vernieuwend onderwijsconcept en het landelijke beleid. De vraag naar een instrument kwam immers van een aantal schoolleiders van zulke scholen en de inhoud van het instrument diende daarom ook voor hun scholen geschikt te zijn. Gegevens daarover zijn verkregen uit bestaande documenten die door de schoolleiders ter beschikking zijn gesteld. Daarnaast hebben de onderzoekers op drie verschillende scholen een aantal les- en begeleidingssituaties bijgewoond om een beter beeld te krijgen van het handelen van de docenten in concrete situaties. Alle gegevens die op deze manier waren verzameld, zijn geanalyseerd en samengevat en het resultaat daarvan is per school aan de schoolleider voorgelegd (*member-check*, Boeije, 2008) en zijn waar nodig nog aangepast.

De vier bronnen (A t/m D) zijn op drie manieren benut. Ten eerste hebben ze bijgedragen aan het ontwikkelen van een indeling van het pedagogisch-didactisch handelen in een aantal rubrieken. Ten tweede vormden ze bronnen voor het genereren van items per rubriek. Ten derde zijn ze gebruikt voor het soms preciezer formuleren van beschrijvingen van handelingen die zijn gegenereerd door een aantal docenten van de betreffende scholen, die als vijfde bron zijn geraadpleegd. Daarop gaan we nu in.

Omdat concepten en vernieuwingen, zeker op pedagogisch-didactische vlak, pas concreet vorm krijgen in de keuzen en het handelen van docenten, omdat de schoolleiders ook nadrukkelijk zicht wilden krijgen op de verschillen tussen docenten, en omdat formuleringen van pedagogisch-didactisch handelen voor docenten herkenbaar moeten zijn en als relevant moeten worden ervaren, is besloten als vijfde bron ook een aantal docenten van de betrokken scholen bij de ontwikkeling van het instrument te betrekken. Deze aanpak sluit aan bij de gedachtegang van Stoof et al. (2002) dat het voor de levensvatbaarheid van de gebruikte terminologie bij het definiëren en operationaliseren van een dergelijk, niet scherp te definiëren begrip als pedagogisch-didactisch handelen, van belang is de betekenisgeving van betrokkenen zelf in het proces van operationalisering te betrekken.

Het doelgericht benutten van de hiervoor genoemde bronnen voor verdere operationalisering van het begrip pedagogisch-didactisch handelen was niet goed uitvoerbaar zonder het domein van het pedagogisch-didactisch handelen van een indeling en ordening te voorzien. Dit was belangrijk om met betrokkenen in gesprek te kunnen gaan, maar ook om de overeenkomsten en verschillen tussen de onderwijsconcepten goed in beeld te krijgen, zodat een kader kon ontstaan dat het geheel overkoepelde en aan de verschillen recht deed.

Om de gesprekken over de onderwijsconcepten met de schoolleiders goed te kunnen voeren, en ook om docenten te kunnen uitnodigen beschrijvingen van relevante handelingen te genereren, is daarom een aantal rubrieken onderscheiden (kolom E in figuur 3.2). Deze zijn door ons taakgebieden genoemd. Deze zijn op te vatten als aspecten of onderdelen van het pedagogisch-didactisch handelen (zie voor een vergelijkbare aanpak van de operationalisering van begrippen Boesjes-Hommes, 1974). Hoewel de opstelling van deze rubrieken c.q. taakgebieden al zou kunnen worden gezien als een (eerste) resultaat van het ontwikkelproces (en dan pas verderop zou moeten worden beschreven, namelijk in paragraaf 3.4), is het dermate kenmerkend voor de gevolgde ontwikkelaanpak dat we dit hier beschrijven. Ook draagt het onderscheiden van een aantal taakgebieden binnen het pedagogisch-didactisch handelen bij aan verheldering van de betekenis van dit begrip.

Hoewel de onderwijsconcepten van de deelnemende scholen (verwerkt in kolom D) onderling verschilden, bleken zij op vergelijkbare wijze te kunnen worden beschreven door onderscheid te maken tussen beoogde leerdoelen, leeromgevingen en leeractiviteiten, en rollen, taken en activiteiten van docenten. Naast rollen, taken en activiteiten die in documenten en door de schoolleiders expliciet werden benoemd, bevatten de onderwijsconcepten in de beschreven leerdoelen, leeromgevingen en leeractiviteiten ook meer impliciet taken die door docenten zouden moeten worden verricht, om die leerdoelen, leeromgevingen en leeractiviteiten te realiseren. Deze zijn ook meegenomen. Hiermee bestreken de onderwijsconcepten drie van de vier “theories” zoals onderscheiden door De Corte (2000; zie ook Bransford et al., 2005), namelijk de *theory of expertise* (de leerdoelen: dat wat moet worden geleerd), de *theory of learning* (de leeractiviteiten: hoe dit kan worden geleerd), en de *theory of teaching*

(de leeromgevingen en de docentrollen, -taken en –activiteiten: hoe de docent het leerproces kan ondersteunen). De vierde *theory* (de *theory of assessment*) kwam slechts beperkt in de onderwijsconcepten aan de orde. Het beoordelen kan echter ook worden gezien als deel van de leeromgeving en het docenthandelen, dus de *theory of teaching*.
In de *inhouden* van de in de onderwijsconcepten aangeduide leerdoelen, -omgevingen, en -activiteiten en docentrollen, -taken en –activiteiten komen de specifieke *vernieuwingen* naar voren waar de scholen mee bezig zijn. De vernieuwende onderwijsconcepten beogen samengevat het onderwijs meer motiverend te doen zijn (minder saai, meer afwisseling, activering), meer samenhangend (minder versnipperd, door te werken met thema's, projecten, authentieke taken en opdrachten), meer ruimte biedend aan eigen keuzes van de leerlingen (vraaggestuurd) en meer gericht op interactie en samenwerking (tussen docenten en leerlingen en tussen leerlingen).
De op basis van de onderwijsconcepten van de scholen gemaakte indeling van taken van docenten (zie kolom D in figuur 3.2) is aan de schoolleiders voorgelegd. Het daarover gevoerde overleg leidde ertoe dat organisatorische taken (als zijnde niet duidelijk pedagogisch-didactisch van aard) zijn geschrapt, en dat verschillende taakgebieden zijn samengevoegd (vakoverstijgende vaardigheden en brede vorming, instrueren en activeren en zelfstandig leren, persoonlijke vorming en volgen van leren en ontwikkeling, realiseren van differentiatie en inspelen op verschillen, en voeren van gesprekken en coaching). De resulterende indeling in 14 rubrieken (zie kolom E in figuur 3.2) is gebruikt bij het betrekken van docenten bij het genereren van beschrijvingen van handelingen.
De docenten bleken in diverse rubrieken meer, en meer gedifferentieerde activiteiten te onderscheiden dan was voorzien. Op grond daarvan zijn *bijdragen aan persoonlijke vorming*, *geven van les, uitleg en instructie*, en *helpen ontwikkelen van vaardigheden* weer als aparte taakgebieden onderscheiden. Ook is als nieuw taakgebied toegevoegd *leerlingen hun eigen leerproces laten plannen en bewaken*. De beschrijvingen in de rubriek *begeleiden van leerlingen* zijn verdeeld over diverse andere rubrieken (geven van les, uitleg en instructie, leerlingen laten werken aan opdrachten, leerlingen hun eigen leerproces laten plannen en bewaken), omdat zij daarop betrekking hadden. De rubrieken *gebruiken van moderne media en ict*, en *stellen van vragen* zijn geschrapt als te specifiek voor een aparte rubriek (items hierover waren door de docenten geformuleerd bij andere rubrieken). Tot slot is de rubriek *ontwikkelen van opdrachten en materiaal* geschrapt (omdat niet alle docenten dit doen).
Het taakgebied 'professionele ontwikkeling', dat in kolom D is toegevoegd, betreft geen handelen in interactie met leerlingen. Het is in het instrument gehandhaafd omdat de schoolleiders die om de ontwikkeling van het instrument hadden verzocht dit ook in het instrument opgenomen wilden zien. De indeling van het pedagogisch-didactisch handelen in taakgebieden (zie kolom F in figuur 3.2) en de verdere operationalisering daarvan in uitspraken over handelingen heeft plaatsgevonden in nauwe samenspraak met de betrokkenen (schoolleiders en docenten).

Naast het voorafgaande heeft ook een aantal voorwaarden met het oog op het beoogde gebruik van het instrument een belangrijke rol gespeeld bij het vormgeven van het ontwikkeltraject dat is doorlopen om tot een geschikt instrument te komen. Hierop wordt nu ingegaan.

3.3 Voorwaarden vanuit het beoogde gebruik van het instrument

3.3.1 De gebruikscontext en gebruiksdoelen

Een belangrijk aspect bij het adequaat handelen van docenten is dat afstemming ontstaat tussen het eigen referentiekader van de docent en de door hem waargenomen eisen die vanuit de school aan zijn handelen worden gesteld. Dit speelt in elke onderwijssituatie, maar komt nadrukkelijker naar voren bij het invoeren van vernieuwingen op het moment dat de context aan verandering onderhevig is (Luttenberg, Imants, Van Veen & Carpay, 2009).
In een dergelijke veranderende context kunnen bepaalde handelingen van docenten, die eerder breed geaccepteerd waren, als minder adequaat worden gezien, of niet (meer) toereikend zijn in die zin dat een breder of anders ingevuld docentrepertoire wordt verwacht. Dit kan betekenen dat de docent zijn handelingsrepertoire moet aanpassen of andere, aanvullende vaardigheden moet ontwikkelen. Het werken in een dergelijke vernieuwende context vraagt dus van docenten dat zij zich (opnieuw) bezinnen op wat zij doen en moeten doen om eventueel hun handelen (meer) in overeenstemming te brengen met de doelen van de onderwijsvernieuwing. Het vraagt ook dat men met elkaar in gesprek gaat en gezamenlijk betekenis geeft aan wat men zou moeten doen, zodat men een gemeenschappelijk referentiekader ontwikkelt waaraan men zijn eigen handelen kan relateren. Op veel scholen ontbreekt het vaak aan tijd en gelegenheid om met elkaar hierover in gesprek te gaan. Ook ontbreekt het vaak aan een begrippenkader waarmee men in gesprekken tot afstemming kan komen over relevante aspecten van het handelen (Sol & Stokking, 2010).
Dit laatste werd ook door de schoolleiders benadrukt en tijdens het ontwikkelproces kregen zij, in samenspraak met de onderzoekers, duidelijker voor ogen wat ze wilden en voor welk doel het beoogde instrument geschikt zou moeten zijn. Zij wilden een instrument waarmee het huidige en het bij vernieuwend onderwijs passende pedagogisch-didactisch handelingsrepertoire van docenten in kaart kon worden gebracht. Zij bleken met name op zoek naar een instrument dat zou kunnen bijdragen aan het voortgaande proces van verandering en ontwikkeling op hun school en aan het meer betrekken van alle docenten bij het veranderingsproces. Dit sluit aan bij het door Frederiksen en Collins (1989) geïntroduceerde begrip *systemic validity*, dat aanduidt dat het gebruik maken van een instrument bijdraagt aan de in de verdere ontwikkeling van de kennis en vaardigheden die het instrument beoogt te meten en die in de betreffende gebruikscontext gewenst worden geacht.

De wens was dat de uitkomsten die met het instrument worden verkregen door docenten zelf kunnen worden benut voor reflectie, en kunnen bijdragen aan de onderlinge dialoog om het gezamenlijke ontwikkelproces te helpen vormgeven en op den duur meer consensus te bereiken over het gewenste handelen. De schoolleiders wilden zicht krijgen op het gedrag van docenten en de afwegingen die aan dat gedrag ten grondslag liggen en de intenties die docenten daarbij hebben. Zij waren tevens geïnteresseerd in wat docenten belangrijk vinden in de interactie met leerlingen en hoe dit zich verhoudt tot het onderwijsconcept van hun school. Ook wilden zij weten in welke mate docenten hierin onderling verschillen en waar ontwikkelpunten zichtbaar gemaakt kunnen worden in relatie tot de doelen die worden beoogd. Tot slot wilden zij dat het instrument geschikt zou zijn voor externe verantwoording. Indien mogelijk wilden de schoolleiders ook gegevens tussen scholen kunnen vergelijken ten behoeve van benchmarking. Duidelijk werd echter dat het niet de bedoeling was om ook tot standaarden voor het docenthandelen te komen.

3.3.2 Reflectie en gezamenlijke betekenisgeving

In de wetenschappelijke literatuur wordt reflectie op het eigen handelen gezien als een belangrijke voorwaarde voor het leren en de verdere professionalisering van docenten (Calderhead & Gates,1993; Grimmett & MacKinnon, 1992; Russell & Mundy, 1991; Schön 1983, 1987). Zimmerman (2000) maakt, mede in navolging van Bandura (1986), bij zelfreflectie onderscheid tussen drie aspecten: zelfobservatie, zelfevaluatie en zelfbeoordeling. Zelfobservatie refereert aan het observeren van specifieke aspecten van het eigen handelen, de condities die daarbij een rol spelen en de effecten die dat handelen heeft. Zelfevaluatie refereert aan het vergelijken van informatie die door zelfobservatie is verkregen met criteria, een standaard of een doel. Zelfevaluatie kan bijdragen aan het eigen leerproces omdat de uitkomsten aanwijzingen kunnen geven voor verbetering van het handelen. Bij zelfbeoordeling, ook wel zelfassessment genoemd, beoordeelt iemand of zijn eigen handelen, volgens zijn eigen perceptie voldoet aan een gestelde norm. Zelfassessment gaat verder dan zelfevaluatie, omdat men ook een causale betekenis toekent aan de resultaten. Zelfevaluatie en zelfassessment maken beide deel uit van zelfreflectie en zijn nauw verbonden met zelfobservatie (Bandura, 1986; Zimmerman, 2000).
De aanwezigheid van externe criteria en standaarden kan duidelijkheid verschaffen over relevante aspecten van de taak en de eisen waaraan de taakvervulling moet voldoen. Dit geeft iemand de gelegenheid om zijn handelen (voorafgaand, tijdens en na afloop van het uitvoeren van een taak) waar nodig aan te passen, waardoor het vermogen tot zelfsturing kan worden versterkt. Zelfevaluatie en zelfassessment worden daarom gezien als een belangrijke factor bij het vergroten van zelfregulatie bij leren (Boekaerts, 1991; Sadler, 1989; Zimmerman, 2000).
In het onderwijs en in de context van de betrokken scholen ontbreken dergelijke expliciete externe standaarden en is geen norm voorhanden (Sol & Stokking, 2010; zie

hoofdstuk 7). Ook ontbreken vaak zelfs criteria, in de zin van een kader waaraan men zijn eigen handelen kan relateren. Aangezien in dit onderzoek het focus lag op het ontwikkelen van criteria voor zelfobservatie en zelfevaluatie van het huidig en gewenst handelen en het ontwikkelen van standaarden niet aan de orde was, wordt dit laatste aspect verder buiten beschouwing gelaten.

Bij ontbreken van een kader met relevante aspecten van het pedagogisch-didactisch handelen in relatie tot de beoogde doelen, is een docent bij de evaluatie van het eigen handelen voor een groot deel aangewezen op eigen percepties van wat gewenst handelen is. Elke docent ontwikkelt voor zichzelf bepaalde criteria en standaarden (absolute of relatieve) op basis van ervaringen uit het verleden (Bandura, 1977; Boekaerts, 1991). Deze criteria of standaarden kunnen op verschillende manieren zijn ontstaan of worden gevoed: door eigen ervaringen bij het uitvoeren van taken, door het zien en beoordeeld zien worden van gedrag van anderen (gevolgtrekkingen op basis van sociale vergelijking), door gevoelens die ontstaan tijdens het handelen, maar ook door het verkrijgen van feedback en verbale interactie met anderen (Bandura, 1977; Boekaerts, 1991).

Volgens Boud en Falchikov (1989) spelen bij zelfevaluatie twee aspecten een belangrijke rol: internalisering van beoordelingscriteria die een rol spelen bij het werk (c.q. bij het realiseren van bepaalde doelen), en kunnen beoordelen van de mate waarin het eigen handelen aan deze criteria voldoet. Ook volgens Benett (1993) hangt de validiteit van zelfevaluatie af van de vraag of men de belangrijke aspecten van het beoogde docenthandelen heeft geïnternaliseerd en in staat is het eigen handelen te bekijken in het licht van de eigen standaarden.

De referentiekaders van docenten zullen verschillen op grond van eigen opvattingen en ervaringen, de wijze waarop men de beoogde doelen heeft gepercipieerd en de eigen ervaringen heeft geduid, en de mate waarin men uit gevoerde discussies criteria en eventuele standaarden heeft geïnternaliseerd. De uitkomsten van zelfevaluaties en zelfbeoordelingen zullen daardoor tussen docenten niet helemaal vergelijkbaar zijn, maar wel kunnen bijdragen aan het uitwisselen van betekenissen. Voor het opbouwen van meer eenduidige persoonlijk geïnternaliseerde standaarden bij een groep docenten in de context van een vernieuwende school, is een dialoog met anderen nodig. Een kader waarin belangrijke aspecten van het pedagogisch-didactisch handelen zijn geformuleerd in uitspraken over concrete activiteiten kan bij de gewenste dialoog tussen docenten behulpzaam zijn.

Een complicerende factor bij zelfevaluatie is dat wat men denkt of claimt te doen, niet altijd overeenkomt met wat men feitelijk doet, waardoor op basis van zelfevaluatie niet altijd een helemaal realistisch beeld ontstaat van het feitelijke handelen (vergelijk het onderscheid tussen espoused theory en theory-in use, Argyrus & Schon, 1974). Men baseert zich immers op de eigen perceptie van het handelen, tegen de achtergrond van de eigen kennis, ervaringen en concepties. Het is daarom van belang aanvullende informatie te vergaren, bijvoorbeeld door video-opnames van concrete situaties. Daarmee kan worden nagegaan in hoeverre dat wat men denkt te doen en wat men

feitelijk doet met elkaar overeenkomen. Het maken en gebruiken van opnames is echter vaak niet haalbaar. Informatie over hoe het handelen overkomt, kan dan ook worden verkregen aan de hand van de percepties van de leerlingen.
Het betrekken van leerlingen bij het in kaart brengen van het pedagogisch-didactisch handelen sluit aan bij de tendens om ervaringen van leerlingen te benutten bij het vormgeven van de leeromgeving en het realiseren van vernieuwingen. Actieve betrokkenheid van leerlingen bij het onderwijs en bij hun eigen leerproces is zowel een voorwaarde als een doelstelling van het inrichten van een eigentijdse leeromgeving, waarin men leerlingen hun eigen leerproces meer wil laten sturen (Levin, 2000; Stokking & Sol, 2010).
Het erbij betrekken van de percepties van leerlingen is ook om een andere reden van belang. Percepties van leerlingen vormen namelijk een belangrijke schakel in de relatie tussen het handelen van docenten en de leeropbrengsten bij leerlingen (Den Brok, 2001). In belangrijke mate wordt dat wat een leerling leert, bepaald door de manier waarop hij of zij informatie waarneemt, interpreteert en verwerkt (Shuell, 1996). Het is voor docenten zelf, maar ook voor de schoolleiding, van belang om zicht te krijgen op de percepties van leerlingen van het pedagogisch-didactisch handelen, omdat de leerlingen een belangrijke aanvullende bron van informatie zijn over hoe het gedrag van de docenten en hun intenties overkomen en zo van invloed kunnen zijn op de leeropbrengsten bij de leerlingen.
Leerlingen kunnen een andere kijk hebben op de huidige en de gewenste leeromgeving dan de docenten of de schoolleiding. Leerlingen kunnen het handelen van een docent ook anders percipiëren dan de docent zelf. Een extra overweging om leerlingen te betrekken bij het in kaart brengen van het handelen van docenten is dat leerlingen hun ervaring met een docent kunnen baseren op meerdere lessen en ook verschillende docenten meemaken of hebben meegemaakt waardoor zij situaties en docenten met elkaar kunnen vergelijken (Brekelmans, 1989; Den Brok, 2001).
In dit onderzoek ligt het focus op het in kaart brengen van het handelen van docenten waarbij naast hun concrete gedrag ook verwachtingen en intenties een rol spelen. Leerlingen zijn vaak goed in staat om deze afwegingen en intenties van een docent te leren kennen. Verder kan aan leerlingen ook worden gevraagd wat ze graag zouden willen dat de docent doet, waardoor extra inzicht wordt verkregen in het gewenste handelen vanuit hun perspectief. Door over een docent meerdere leerlingen te bevragen en hun antwoorden samen te voegen, kan worden voorkomen dat persoonlijke voorkeuren of een toevallige stemming een (te) grote invloed hebben op de resultaten. Daarnaast kan hierdoor zichtbaar worden in hoeverre de percepties en voorkeuren van leerlingen binnen de totale groep onderling verschillen (Den Brok, 2001).

3.3.3 De aan een instrument te stellen kwaliteitseisen

Een instrument om het pedagogisch-didactisch handelen in kaart te brengen moet valide, betrouwbaar en bruikbaar zijn. Wat betreft de validiteit is het van belang dat het

instrument het domein van het pedagogisch didactisch handelen dekt (construct-validiteit c.q. inhoudsvaliditeit) en dat recht wordt gedaan aan de realiteit van het dagelijks handelen (ook wel ecologische validiteit genoemd). Tevens moet het taalgebruik in het instrument herkenbaar zijn voor docenten en aansluiten bij hun taalgebruik (Beijaard, 1990; Carter, 1990). Indien het instrument ook door leerlingen moet worden gebruikt, moeten formuleringen ook voor hen herkenbaar zijn en aansluiten bij hun taalgebruik.
Het gebruik van een dergelijk instrument kan op zich al bijdragen aan de ontwikkeling van het handelen van docenten en meer inzicht geven in de handelingen die in de betreffende context gewenst zijn. Dit strookt met het door de schoolleiders beoogde gebruiksdoel van het instrument. In navolging van de bijdrage die kennistoetsen in het beoogde leerproces van leerlingen kunnen spelen (Frederiksen & Collins, 1989), kunnen de taakgebieden en items een soortgelijke rol vervullen binnen het leerproces van docenten. We spreken dan van *systemic validity* (Frederiksen & Collins, 1989).
Bij het ontwikkelen van een instrument dat op deze manier zelf ook moet bijdragen aan het beoogde leerproces, speelt een aantal aspecten een rol. Het instrument moet het hele domein van het pedagogisch-didactisch handelen bestrijken. Daar waar taken te onderscheiden zijn, moeten deze afzonderlijk in kaart worden gebracht aan de hand van een beperkt aantal uitspraken die voor betrokkenen herkenbaar zijn. Daarnaast wordt in de literatuur benadrukt dat het met het oog op zelfevaluatie belangrijk is dat de betrokkenen zich de criteria eigen kunnen maken (internalisatie) en feedback krijgen van relevante anderen om hun eigen uitkomsten aan te kunnen relateren en dat herhaalde afname van het instrument kan bijdragen aan het leerproces (Frederiksen & Collins, 1989; Moss, 1992). Andere kwaliteitsaspecten die belangrijk zijn voor de systemic validity (zoals beschikbaarheid van voorbeelden op verschillende niveaus en training van beoordelaars) zijn in dit ontwikkeltraject niet aan de orde.
De validiteit van een instrument kan ook worden onderzocht door de uitkomsten ervan te vergelijken met die van een ander instrument voor hetzelfde construct (congruente validiteit). Aangezien geen andere instrumenten voorhanden bleken te zijn die het pedagogisch-didactisch handelen van docenten op systematische wijze en in de volle breedte in kaart brengen, was deze vorm van validering niet mogelijk. De inhoud van het instrument kan wel worden vergeleken met indelingen vanuit beleidsgerichte kaders die ook pedagogisch-didactische aspecten bevatten (SBL-competenties en Toetsingskader Onderwijsinspectie). Deze zijn benut in figuur 3.2 in kolom D.

3.4 Opzet en uitvoering van het empirisch onderzoek

3.4.1 Instrumentontwikkeling

Inleiding
Na de vaststelling van een indeling in taakgebieden (zie paragraaf 3.2), is een aantal beslissingen in samenhang met elkaar genomen:

- de keuze voor het type instrument (een schriftelijke gesloten vragenlijst);
- de keuze voor één versie voor zowel docenten als leerlingen en voor alle scholen;
- de keuze voor de aard van de vragen en itemformuleringen;
- de keuze voor de werkwijze en stappen bij het genereren en selecteren van items.

De keuze voor het type instrument

Het beantwoorden van de vraag naar verschillen tussen docenten, klassen en scholen vereist een grootschalige dataverzameling en dit kan het meest efficiënt met een schriftelijk en zoveel mogelijk voorgestructureerd instrument. Deze keuze paste ook goed bij de behoefte van de schoolleiders om zich te kunnen verantwoorden, omdat met zo'n instrument meer objectieve en cijfermatige gegevens kunnen worden verkregen en zulke gegevens zich goed voor dat doel lenen.

Op het moment dat uitkomsten moeten kunnen worden benut voor reflectie, betekenisgeving en dialoog door zowel (individuele) docenten als de schoolleiding, is het van belang dat het instrument een conceptueel kader biedt, waarin alle relevante aspecten op een voor betrokkenen herkenbare wijze aan bod komen. Gezien het belang van reflectie op het eigen handelen in het kader van de verdere professionalisering van docenten is het van belang dat het instrument geschikt is om het eigen handelen in kaart te brengen en de mogelijkheid biedt deze uitkomsten te vergelijken met de uitkomsten van anderen. Daarbij staat de betekenisgeving en de mogelijkheid om met elkaar in dialoog te gaan centraal en nadrukkelijk niet het komen tot een expliciete normering.

Op grond van het bovenstaande is gekozen voor de ontwikkeling van een schriftelijke vragenlijst met gesloten vragen, waarbij de docent per uitspraak kan aangeven in welke mate hij of zij bepaalde activiteiten doet en daarnaast hoe belangrijk hij of zij deze vindt.

De keuze voor één versie voor zowel docenten als leerlingen en voor alle scholen

Bevraging van docenten en leerlingen vergde twee vragenlijsten, maar omdat de percepties van beide groepen moesten kunnen worden vergeleken, dienden beide lijsten zoveel mogelijk overeen te komen: dezelfde items, alleen met een verschillend focus bij het invullen (de docenten vullen de lijst in met betrekking tot hun eigen handelen, de leerlingen met betrekking tot het handelen van een docent). Het ging dus in feite om de ontwikkeling van één vragenlijst (afgezien van enkele toe te voegen algemene vragen over persoonlijke gegevens).

De deelnemende scholen zijn alle actief bezig om hun onderwijs te vernieuwen, maar maken daarbij deels verschillende keuzen en verkeren ook in verschillende stadia van ontwikkeling. Tijdens de gesprekken met de schoolleiders was aanvankelijk het idee om van het instrument per school een schoolspecifieke versie te maken, zo goed mogelijk toegesneden op het onderwijsconcept, de stand van het vernieuwingsproces en de actuele informatiebehoefte op die school. Bij het voorleggen van het voorstel met de rubrieken die in het instrument zouden kunnen worden opgenomen (zie 3.2.2),

bleken de schoolleiders echter ook geïnteresseerd in rubrieken die zij zelf aanvankelijk niet hadden genoemd. Vervolgens kwam naar voren dat de scholen alleen onderling kunnen worden vergeleken (hetgeen werd gezien als gewenste vorm van *bench-marking*) voor zover de vragenlijsten overlappen. Toen is besloten toe te werken naar één versie voor alle scholen.

De keuze voor de aard van de vragen en itemformuleringen

Het instrument diende ook te kunnen worden ingezet in het kader van de professionele ontwikkeling van docenten. Het werd daarom wenselijk geacht om zowel te vragen naar het huidige als naar het gewenste handelen (of meer precies: in welke mate de docent het betreffende handelingsaspect belangrijk vindt), omdat verschillen hiertussen kunnen indiceren waar verdere ontwikkeling wenselijk is. In de vragenlijst is dit vormgegeven door beide vragen en de bijbehorende antwoordmodaliteiten naast elkaar te plaatsen, rechts naast de items. Bij elk taakgebied werd boven de items de volgende zin geplaatst: "Geef bij elke uitspraak antwoord op de twee vragen hiernaast". Boven de eerste kolom met antwoordmodaliteiten stond bij de docenten "Dit doe ik nu" en boven de tweede kolom "Dit vind ik belangrijk om te doen". Bij de leerlingen waren beide vragen als volgt geformuleerd: "Past dit bij wat de leraar doet?" en "Wil je dat de leraar dit doet?" In beide versies (voor docenten en leerlingen) was sprake van de volgende vijf antwoordmodaliteiten: helemaal niet; meestal niet; deels niet, deels wel; meestal wel; helemaal wel. Voor deze combinaties van vragen en antwoord-modaliteiten is gekozen omdat zij van toepassing kunnen zijn op alle activiteiten en situaties die in de taakgebieden en items aan de orde komen. De formuleringen van de items waren in beide versies geheel gelijk, in termen van "de leraar". Gekozen is voor formuleringen die voor leerlingen begrijpelijk zijn zonder bij docenten te simpel over te komen.

Het instrument is gericht op het in kaart brengen van het pedagogisch-didactisch handelen van docenten, in termen van zowel concreet waarneembaar gedrag als bijbehorende afwegingen en intenties. Bij de operationalisering van de taakgebieden zijn daarom ook uitspraken opgenomen die aangeven wat de docent probeert te bereiken. Dit kunnen leerlingen vaak goed aangeven, omdat ook zij aan gedrag betekenis toekennen: docenten hebben bepaalde bedoelingen en leerlingen percipiëren en interpreteren het gedrag van hun docenten in termen van zulke bedoelingen. Voorbeelden van zulke items zijn: "de leraar moedigt leerlingen aan te laten zien wat ze kunnen" en "de leraar nodigt de leerlingen uit verbanden met andere vakken te zoeken". Ter verduidelijking van de inhoud van het instrument werd besloten de onderscheiden taakgebieden expliciet in het instrument te noemen en de items per taakgebied aan te bieden.

Tot slot is nogmaals met de schoolleiders besproken of toch niet op de een of andere manier rekening zou moeten en kunnen worden gehouden met het feit dat de scholen zich in verschillende stadia van ontwikkeling en vernieuwing bevinden. Overwogen is de items per taakgebied in verschillende varianten te formuleren, van meer docent-

gestuurd naar meer leerlinggestuurd. Dit bleek echter niet goed realiseerbaar en ook werd bedacht dat het niet nodig is als door adequate formulering van de items en antwoordmodaliteiten de docenten en leerlingen op de betreffende antwoordschaal kunnen aangeven *in welke mate* bepaald handelen voorkomt respectievelijk gewenst is. Daarmee werd het besluit bevestigd om één versie voor alle scholen te maken.

De keuze voor de werkwijze en stappen bij het genereren en selecteren van items
De ontwikkeling van de items is in een aantal stappen verlopen.
Als eerste stap zijn op drie van de zes deelnemende scholen in groepssessies per school door in totaal 16 docenten items gegenereerd (korte beschrijvingen van handelingen van docenten). Zij kregen daartoe eerst een korte uitleg van de doelstelling van het project en de bedoeling van de sessie, en de instructie om op te schrijven wat ze doen, met welke intentie, en hoe ze daarin variëren, en daartoe steeds zinnen te formuleren die beginnen met "ik", gevolgd door een werkwoord (er werd een aantal voorbeelden gegeven, zoals: geef, zeg, bespreek, laat, houd, vertel, stimuleer, vraag, etc.). Voor het noteren van de items ontvingen ze een invulformulier. Dit bestond uit vijf pagina's met daarop de 14 rubrieken (zie figuur 3.2, kolom E) met per rubriek een naam (bijvoorbeeld: Volgen van het leren en de ontwikkeling van leerlingen), een toelichting in de vorm van enkele steekwoorden (denk hierbij aan: …) en zes regels met ruimte voor het noteren van zes items, en een extra pagina met vier rubrieken zonder naam met elk ook weer zes regels voor zes items, waarop eventuele nieuwe rubrieken en items konden worden toegevoegd. Dit resulteerde in een pool van iets meer dan 450 items. Van de mogelijkheid om zelf een rubriek toe te voegen was slechts weinig gebruik gemaakt, en de items die daarbij waren geformuleerd, bleken meestal te kunnen worden ondergebracht bij een van de aangereikte 14 rubrieken (de overige items waren of heel specifiek of onduidelijk).
Deze 450 items zijn vervolgens door de onderzoekers teruggebracht tot 335 items, door het schrappen van dubbele items, vage items, items die niet duidelijk tot een rubriek behoren (ook al waren ze wel in een bepaalde rubriek genoteerd), te specifieke items en te ingewikkelde items. De overgebleven items zijn waar mogelijk meer precies geformuleerd, mede met gebruik van de gemaakte analyse van de onderwijsconcepten van de scholen (op basis van de verkregen documenten, gesprekken met de schoolleiders en observaties van les- en begeleidingssituaties) en de bestudeerde literatuur.
Dit resulteerde in een eerste, ruwe versie van de vragenlijst (14 rubrieken en 335 items). Om bij docenten en leerlingen dezelfde formuleringen te kunnen gebruiken is in de items het woord "ik" vervangen door "de leraar" en de persoonsvorm daarop grammaticaal aangepast. Deze eerste versie is, op de andere drie scholen (de scholen waarvan geen docenten hadden meegewerkt aan het genereren van items), door 6 docenten en 9 leerlingen in aanwezigheid van de onderzoeker bij wijze van proef ingevuld, waarbij de docenten en leerlingen aangaven wat ze niet begrepen, wat volgens hen niet van toepassing was, etc. Ook is bijgehouden hoeveel tijd het invullen

vergde; dit was gemiddeld ruim anderhalf uur. De schoolleiders en docenten op de eerste drie scholen (waar de items waren gegenereerd) kregen deze eerste versie voorgelegd met het verzoek om commentaar.
De docenten die de eerste versie voor proef hebben ingevuld en becommentarieerd (laatste drie scholen) en de schoolleiders en docenten die daarop commentaar hebben gegeven (eerste drie scholen) gaven aan dat de lijst (vrijwel) dekkend is voor de diversiteit van het handelen van docenten. Enkelen gaven aan dat de lijst niet teveel zou moeten worden ingekort, want dan zou geen recht worden gedaan aan het docentschap. Anderen vonden dat er nauwelijks overbodige items in stonden. Een deel van de leerlingen die de eerste versie voor proef hebben ingevuld, gaf aan enkele vreemde woorden niet te kennen (feedback, moderne media, realiteit, concreet). De leerlingen vonden de vragen verder niet vreemd, het ging volgens hen steeds over wat docenten (kunnen) doen. Een deel van de leerlingen besloot de lijst in te vullen met een bepaalde leraar voor ogen, veelal hun mentor, omdat een deel van de handelingen in hun geval met name door de mentor werd uitgevoerd. Daarnaast bleek dat zij de vragen alleen konden invullen met een bepaalde leraar voor ogen, omdat naar hun mening de leraren onderling nogal verschillen in wat ze doen.
Op basis van deze proefinvulling kon het aantal items verder worden teruggebracht tot 270. Dit gebeurde voornamelijk door het schrappen van items die (teveel) op elkaar leken. Voor de bij de proefinvulling betrokken docenten waren dat vaak nog wel betekenisvolle nuanceringen (zij hadden het gevoel dat er nauwelijks overbodige items in stonden en dat het geheel een goede weergave was van mogelijk docentgedrag) maar de betrokken leerlingen hadden het gevoel regelmatig dezelfde vragen te moeten beantwoorden.
Deze versie van 270 items is op vijf van de zes scholen in try-out gegaan. In de introductie in het begin van de vragenlijst werd uitgelegd dat leraren verschillende taken hebben, dat de vragenlijst gaat over een aantal van die taken, dat bij elke taak uitspraken staan over wat leraren kunnen doen, en dat bij elke uitspraak twee dingen kunnen worden aangegeven (hier volgden beide vragen, naar het huidige en het gewenste handelen). De bedoeling van de try-out was om op basis van analyses van de data (op itemnonrespons, spreiding, dimensionaliteit en homogeniteit) te komen tot een definitieve versie van rond 150 items (gemiddeld 10 per taakgebied).
Aan de try-out werd deelgenomen door 17 docenten en 93 leerlingen. Omdat het aantal deelnemende leerlingen wel voldoende was voor factoranalyse per taakgebied, maar het aantal deelnemende docenten niet, werd begonnen met de analyse van de data van de leerlingen. Er bleken geen items te zijn waarbij veel leerlingen geen antwoord hadden gegeven en ook geen items met een heel laag of hoog gemiddelde en/of een heel beperkte spreiding. Een aantal taakgebieden bleek al sterk eendimensionaal (hoge eerste eigenwaarde) en homogeen (hoge itemrestcorrelaties), bij andere taakgebieden was dit (nog) minder het geval. De antwoorden op de vraag naar het gewenste docenthandelen bleken sterker samen te hangen (hogere itemrestcorrelaties en hogere Cronbachs alpha's) dan de antwoorden op de vraag naar het huidige handelen. Op

grond daarvan zijn de analyses geconcentreerd op de data met betrekking tot het huidige handelen en zijn de beslissingen over het schrappen, verplaatsen en/of herformuleren van items in eerste instantie daarop gebaseerd. Deze beslissingen waren gericht op het verkrijgen van homogene schalen die kunnen worden opgevat als valide operationaliseringen van duidelijke constructen (taakgebieden).
In alle gevallen waarin de itemrestcorrelaties (relatief) laag waren, was het item qua formulering en/of inhoud duidelijk afwijkend van de overige items bij de betreffende taak. De meeste van dergelijke items zijn verwijderd, behalve als het een los item betrof dat op zich van belang is (een voorbeeld bij *Samenwerking tussen leerlingen*: 'De leraar laat leerlingen samenwerken aan een taak of opdracht'). Bij beslissingen over schrappen van items op basis van itemrestcorrelaties is ook de factorstructuur betrokken van de factoren op alle items in het betreffende taakgebied. Items die op meer dan één factor laadden of die op geen enkele factor laadden, zijn in veel gevallen ook verwijderd. Bij een beperkt aantal items speelde ook mee dat het item qua inhoud bij nader inzien minder duidelijk paste bij het betreffende taakgebied en/of beter paste bij een ander taakgebied.
Na afronding van de analyses en beslissingen op basis van de data van de leerlingen zijn bij de data van de docenten de schalen gecontroleerd. In alle gevallen bleken de Cronbach alpha's voldoende tot goed. Bij sommige schalen hadden enkele items een te lage itemrestcorrelatie. Dit is echter genegeerd, omdat dit gezien het kleine aantal docenten dat aan de try-out deelnam een toevallig resultaat kon zijn en omdat iteratieve besluitvorming over items door afwisselende analyses op de data van de leerlingen en die van de docenten zou kunnen leiden tot het schrappen van steeds meer items, met een verbrokkelde vragenlijst als resultaat. Een en ander resulteerde in een sterke reductie van het aantal items, van 270 tot ongeveer 140.
Vervolgens is elk taakgebied met bijbehorende resterende items opnieuw inhoudelijk bestudeerd. Bij sommige taken zijn op basis van de bestudeerde literatuur formuleringen van items alsnog aangescherpt en/of items toegevoegd (met name bij de taakgebieden *Geven van les, uitleg en instructie, Leerlingen laten samenwerken en hen daarbij ondersteunen, Geven van feedback,* en *Beoordelen van het werk en de prestaties van leerlingen*), om het betreffende taakgebied inhoudelijk beter te dekken, en bij een deel van de taakgebieden is de aanduiding van het taakgebied aangepast, om beter te passen bij de betreffende items en bij de aanduidingen van de andere taakgebieden.
Taakgebied 14 (Samenwerken met andere leraren) is in twee delen gesplitst, een eerste deel met items die op alle scholen van toepassing kunnen zijn (14a), en een tweede deel (14b) over de situatie waarin meerdere leraren samen een grotere groep leerlingen begeleiden. Deze laatste situatie was niet op alle scholen van toepassing, en in dat geval kon de tweede set items worden overgeslagen. (De leerlingen waren dan klaar met het invullen van de vragenlijst, de docenten dienden nog taakgebied 15 in te vullen, over hun eigen professionele ontwikkeling).

Het eindresultaat was een vragenlijst met 144 items verdeeld over 14 taakgebieden, 12 extra items over (eventueel) gezamenlijke begeleiding door meerdere docenten (taakgebied 14b), en 16 items alleen voor docenten over activiteiten op het vlak van professionele ontwikkeling (taakgebied 15).
Bij de indeling en uitwerking van deze taakgebieden komen zowel taken van docenten aan bod die van oudsher tot het docentenrepertoire behoren als "nieuwe" taken die docenten moeten gaan vervullen bij het realiseren van een vernieuwend onderwijs-concept (zie hoofdstuk 1). We kunnen globaal de volgende indeling maken:
Meer traditionele taken: creëren van een goed leerklimaat; bijdragen aan persoonlijke vorming; geven van les, uitleg en instructie; de samenhang met de dagelijkse leefwereld laten zien; inspelen op verschillen door differentiatie; volgen van de ontwikkeling; geven van feedback; beoordelen van resultaten; afstemmen met andere docenten per vaksectie;
Meer vernieuwende taken: de samenhang tussen vakken laten zien; leerlingen leren opdrachten zelfstandig aan te pakken; aanleren van vaardigheden; inspelen op verschillen door leerlingen keuzes te laten maken en hen te laten samenwerken en dit actief te ondersteunen; leerlingen hun eigen leerproces laten plannen en bewaken en dit actief ondersteunen; bevorderen van het gebruik van feedback; vakoverstijgend samenwerken met andere docenten; professionele ontwikkeling in interactie met collega's.
De meeste vernieuwende taken werden ook vroeger al wel vervuld, maar krijgen tegenwoordig een andere of meer specifieke invulling. Dit geldt vooral voor de taakgebieden 4, 6, 7, 9 en 12. Bij de uitwerking van deze taakgebieden in het instrument is hiermee uiteindelijk rekening gehouden door zowel meer traditionele als meer vernieuwende handelingen op te nemen. Alleen de taakgebieden 5 en 10 lijken min of meer geheel nieuw te zijn, maar ook hierbij kunnen minder en meer vernieuwende handelingen van docenten worden onderscheiden.

3.4.2 Procedure van dataverzameling

Met de schoolleiders is apart overleg gevoerd over de organisatie van de afname van de definitieve vragenlijst. De meesten gaven te kennen dat het onderzoek voor hen vooral een concrete opbrengst krijgt als de vragenlijsten worden ingevuld voor combinaties docent-leerlingen, zodat de gegevens van de leerlingen kunnen worden gekoppeld aan een bepaalde docent. Een aantal scholen heeft echter de organisatie van het onderwijs in termen van uren, ruimtes, vakken, docenten vervangen door een, op elke school weer andere, organisatie in termen van uren, ruimtes, rollen, docenten, waarbij docenten op verschillende momenten en in verschillende situaties (ruimten) verschillende rollen vervullen: expert, mentor, begeleider, etc. Per rol kunnen dan verschillende taken (of delen daarvan) wel of niet van toepassing zijn. Met de schoolleiders is besproken hoe hiermee om te gaan. Eén schoolleider overwoog om leerlingen de lijst te laten invullen voor één specifieke rol. Een nadeel hiervan is echter

dat daarmee geen overall-beeld van een bepaalde persoon wordt verkregen (alleen voor dat deel dat betrekking heeft op de betreffende rol). Het werd bovendien voor alle betrokkenen (leerlingen, docenten, schoolleiders en onderzoekers) niet werkbaar geacht om met een aantal verschillende versies van het instrument te gaan werken (per school per rol). Ook verschillen de namen van de rollen tussen de scholen, verschillen docenten in de manier waarop zij rollen invullen, en weten leerlingen lang niet altijd wie op welk moment welke rol vervult (of geacht wordt te vervullen). Besloten werd dat de (groepen) leerlingen die de lijst gaan invullen, worden geïnstrueerd voor wie (welke persoon) zij de lijst invullen, en dat de zij bij taken die op hun school horen bij een bepaalde situatie of rol die situatie of rol min of meer vanzelf wel voor ogen zouden nemen. De docenten vullen de lijst in voor de betreffende groep leerlingen (bijvoorbeeld: 'onderbouw havo').

De dataverzameling heeft plaatsgevonden in de eerste twee weken van februari 2008. Per school zijn contactpersonen aangewezen. Dit waren op drie van de zes scholen de schoolleiders zelf, op twee scholen een teamleider en op een school een directie-secretaresse. Zij werden vooraf geïnstrueerd en hadden een periode van twee weken om de afname te organiseren en de vragenlijsten uit te delen, te laten invullen door zowel leerlingen als docenten, en weer te verzamelen. Nadrukkelijk is gewezen op het belang van goede communicatie over het onderzoek en het goed invullen van de vragenlijsten. De privacy van docenten was vanaf het begin een belangrijk aandachtspunt, om het ontstaan van weerstand of het invullen van sociaal wenselijke antwoorden te voorkomen. Op alle scholen kregen de docenten de gelegenheid de ingevulde vragenlijsten in gesloten enveloppen aan te leveren.

Op elke school kon in principe het hele docententeam meedoen en daarnaast een 200-tal leerlingen (dit in verband met de beschikbare tijd en budget voor het invoeren van de gegevens). Om een goed beeld te krijgen van elke school werd aanbevolen dat per school tenminste 30 docenten de vragenlijst zouden invullen en tenminste 150 leerlingen.

In de laatste week van februari waren 993 leerling- en 144 docentvragenlijsten ontvangen. Een deel van de leerlingvragenlijsten en enkele docentvragenlijsten bleken zeer onvolledig te zijn ingevuld (voor minder dan de helft) of niet serieus te zijn ingevuld (per taakgebied een verticale streep bij één antwoordmogelijkheid). Deze zijn verwijderd en niet meegenomen bij de verdere verwerking. Tabel 3.1 bevat een overzicht van de bruikbare respons per school: van 900 leerlingen (91 %) en van 141 docenten (98 %).

Tabel 3.1: Aantallen bruikbare vragenlijsten per school

School	*A*	*B*	*C*	*D*	*E*	*F*	*Totaal*
Leerlingen	131	187	70	134	258	120	900
Docenten	26	19	8	9	71	8	141

Aan de scholen waren suggesties gedaan voor manieren waarop de deelnemende leerlingen en docenten zouden kunnen worden geselecteerd en de afname zou kunnen worden georganiseerd (zie boven). Een van de suggesties was om leerlingen de lijst te laten invullen voor de docent die tevens hun mentor is. Deze suggestie is door de meeste scholen overgenomen, met als gevolg dat driekwart van de bruikbare leerlingenlijsten is ingevuld voor een docent/mentor. De 900 leerlingen die de vragenlijst voldoende bruikbaar hebben ingevuld, hebben dat gedaan voor in totaal 151 verschillende docenten, wat dus neerkomt op gemiddeld 6 leerlingen per docent. Van 141 docenten is een bruikbare docentenvragenlijst verkregen. De overlap tussen beide groepen docenten (151 waarvoor ingevuld door de leerlingen en 141 zelf een lijst ingevuld) betreft 85 docenten. Van het totaal van 151 docenten waarvoor leerlingen de lijst hebben ingevuld zijn er 17 docenten waarvoor 10 leerlingen of meer de lijst hebben ingevuld en 35 docenten waarvoor 8 leerlingen of meer de vragenlijst hebben ingevuld.
De deelnemende leerlingen per school kunnen, gezien de aantallen in relatie tot de destijds aanwezige leerlingenpopulatie op de scholen, waarschijnlijk wel representatief worden geacht voor alle leerlingen op de school, maar bij de deelnemende docenten is dit bij de scholen waarvan relatief weinig docenten de lijst invulden minder duidelijk (dit betreft de scholen C, D en F).
De bruikbare respons van de leerlingen is gespreid over alle te onderscheiden schooltypen of niveaus: 35% vmbo, 8% vmbo-t/havo, 16% havo, 23% havo/vwo, 13% vwo, 6% gymnasium. De verhouding jongens-meisjes bij de leerlingen was 40%-60%. De leeftijden liepen uiteen van 12 t/m 18 jaar. De verhouding mannen-vrouwen onder de docenten was 43%-57%. De leeftijden varieerden van 20 tot en met 64 jaar. De aanstellingsomvang liep uiteen van 0,2 tot 1,0 fte (met een gemiddelde van bijna 0,8). De docenten hadden 0 tot 35 jaar ervaring op de betreffende school (gemiddeld 6 jaar) en 0 tot 38 jaar onderwijservaring in totaal (gemiddeld 12 jaar).

3.4.3 Analyses

Op de data van de leerlingen (N=900) over het huidig handelen van de docenten is over alle items per taakgebied een factoranalyse uitgevoerd (principale componentenanalyse met varimax rotatie). Bij het interpreteren van de uitkomsten van de factoranalyse werden de volgende criteria aangehouden: de eigenwaarde van een factor moet groter zijn dan 1, elk item moet hoog laden op één factor (tenminste .50) en de factor moet inhoudelijk goed interpreteerbaar zijn.
Vervolgens zijn, om te bepalen of de items die hoog laden op een factor een betrouwbare schaal vormden, de item-restcorrelaties per taakgebied bepaald en is de betrouwbaarheid berekend met Cronbach's alpha, waarbij werd gestreefd naar een waarde $\geq$.70 en werd .60 aangehouden als minimum. Voor de item-restcorrelaties (r_{ir}) (in SPSS aangeduid met "corrected item-total correlation") werd gestreefd naar .35 en was .20 het minimum. Vervolgens zijn de schalen geconstrueerd, is nagegaan of deze

inhoudelijk goed interpreteerbaar zijn en of zij ook in de data van de docenten tot betrouwbare schalen leiden. De gemiddelden en standaarddeviaties van de nieuwe schaalvariabelen zijn berekend over zowel de gegevens van de leerlingen als die van de docenten, van zowel het huidige als het gewenste handelen.
Uit de factoranalyses en itemanalyses bleek dat bij zes van de 15 taakgebieden ook betekenisvolle subtaakgebieden konden worden onderscheiden. Van de items in deze subschalen zijn de item-restcorrelaties en per subschaal de betrouwbaarheid berekend (Cronbach's alpha). Aanvullend is een factoranalyse uitgevoerd over de geconstrueerde schalen voor de taakgebieden, om na te gaan of sprake is van groepen bij elkaar behorende taakgebieden dan wel dat alle taakgebieden samen één factor vormen.
Vervolgens zijn aanvullende indicaties voor de betrouwbaarheid van het instrument berekend. Ten eerste is per docent waarvoor tenminste 8 leerlingen de vragenlijst hebben ingevuld de jury alpha (interbeoordelaarsbetrouwbaarheid) berekend. Deze maat geeft aan in hoeverre de beoordelingen van de leerlingen die de lijst hebben ingevuld voor dezelfde docent onderling overeenstemmen.
Ten tweede is per docent de correlatie berekend tussen het profiel (de reeks scores op de 14 taakgebieden) van het huidig handelen van de docent zelf en het profiel volgens de gemiddelde scores van de leerlingen bij die docent. Bij de interpretatie van de uitkomsten van de correlaties is ervan uitgegaan dat er sprake is van een zwak verband indien $r \leq .10$, een matig verband indien $r \pm .30$, en een sterk verband indien $r \geq .50$ (Cohen,1992).
Ten derde is nagegaan in hoeverre het instrument discrimineert tussen docenten, door berekening van de intraclassecorrelatie per taakgebied (de ICC(1), zie Lűdke, Robitzsch, Trautwein & Kunter, 2009). De ICC(1) is berekend door middel van enkelvoudige variantie-analyse met docent als random factor. In onderwijskundig onderzoek wordt een ICC(1) van .25 als hoog beschouwd (Mainhard, 2009) en komt een ICC(1) van >.30 niet vaak voor (Lűdke et al., 2009).
Tot slot is per taakgebied door toepassing van de Spearman-Brown formule op de intraclassecorrelatie de betrouwbaarheid geschat van de op docentniveau geaggregeerde leerlingscores (ICC(2)). Een ICC(2) $\geq .85$ wordt gezien als indicatief voor een goede betrouwbaarheid (Lűdke et al., 2009).

Om tegemoet te komen aan de wens van de schoolleiders om het aantal items per taakgebied terug te brengen is tot slot een verkorte versie van de vragenlijst gemaakt, aangeduid met PDH (profiel van het Pedagogisch-Didactisch Handelen van docenten). In dit instrument zijn de gevonden subtaakgebieden als aparte taken opgenomen, waardoor de PDH 22 taken bevat, met in totaal 99 items in de eerste 20 taken die de breedte van het pedagogisch-didactisch handelen van een docent representeren. Bij het reduceren van de oorspronkelijke 144 items tot 99 items is zowel gelet op de itemrestcorrelaties als de inhoud van de items (die per taak deze taak zo goed mogelijk dienden te blijven dekken). Ook zijn de correlaties bepaald tussen de oude schalen (taakgebieden) en de nieuwe schalen (taken).

3.5 Resultaten

3.5.1 Factoranalyses, schaalanalyses en beschrijvende analyses

Op de data van de leerlingen (n=900) over het huidige handelen van de docenten is per taakgebied een factoranalyse uitgevoerd. Tabel 3.2 geeft een overzicht van de uitkomsten. Taakgebied 15 was alleen van toepassing bij de docenten en is bij de factoranalyse buiten beschouwing gelaten. Bij alle taakgebieden waren maximaal twee eigenwaarden ≥1. Bij negen taakgebieden was de tweede eigenwaarde ≥1. Dit betrof de taakgebieden 1, 3, 4, 6, 7, 9, 10, 12 en 13. Op grond van deze uitkomsten is besloten na te gaan of er subtaakgebieden kunnen worden onderscheiden, zie hiervoor ook 3.6.

Tabel 3.2: Aantal items, 1e en 2e eigenwaarde (λ) en cumulatieve R^2 per taakgebied, op de data van de leerlingen (n=900) over het huidige handelen van de docenten

Taakgebied		*# items*	*1e λ*	*R^2*	*2e λ*	*R^2*
1	Zorgen voor een goed leerklimaat	12	5.5	46	1,0	.55
2	Bijdragen aan de persoonlijke vorming van leerlingen	7	3.6	51	0.8	.63
3	Geven van les, uitleg en instructie	11	4.8	44	1.1	.54
4	Leerlingen het nut en de samenhang laten zien	10	4.9	49	1.1	.60
5	Leerlingen leren zelf opdrachten aan te pakken	10	5.2	52	0.8	.60
6	Leerlingen stimuleren en vaardigheden aanleren	11	5.7	52	1.0	.62
7	Inspelen op verschillen tussen leerlingen	10	5.0	50	1.1	.61
8	Leerlingen laten samenwerken en hen daarbij ondersteunen	11	6.1	56	0.9	.64
9	Leerlingen eigen keuzes laten maken en hen ondersteunen	11	5.4	49	1.3	.60
10	Leerlingen hun eigen leerproces laten plannen en bewaken	10	5.0	50	1.2	.61
11	Volgen hoe leerlingen leren en zich ontwikkelen	11	6.7	61	0.8	.68
12	Geven van feedback aan leerlingen	11	6.4	58	1.2	.69
13	Beoordelen van het werk en de prestaties van leerlingen	11	5.5	50	1.1	.60
14a	Samenwerken met andere leraren	8	4.9	61	0.7	.69
14b	Samen een groep begeleiden (in domein of leerplein)	12	6.1	51	0.9	.58
15	Professionele ontwikkeling	16	-	-	-	-

Daarna zijn per taakgebied analyses uitgevoerd om na te gaan in hoeverre de antwoorden op elk item samenhangen met de antwoorden op alle andere items samen (itemrestcorrelaties). Bij de meeste taakgebieden waren alle itemrestcorrelaties ≥.35 (zie tabel 3.3). Bij 11 taakgebieden bleken de itemrestcorrelaties van steeds één item onder de .35 te liggen (maar wel ≥ .24). Bij taakgebied 13 waren bij het huidig handelen volgens de docenten de itemrestcorrelaties van twee items lager dan .35 (maar ≥ .18). De lagere itemrestcorrelatie bij sommige items betroffen vooral de uitkomsten bij de docenten. Bij de leerlingen ging het hierbij slechts om één item van één taakgebied. In eerste instantie is ervoor gekozen de betreffende items bij de docenten niet te verwijderen, omdat zij inhoudelijk betekenisvol zijn, de item-restcorrelaties van diezelfde items bij de leerlingen vrijwel allemaal goed zijn, en de

betrouwbaarheid van de schaal als geheel steeds voldoende is. Ook is per taakgebied de Cronbach's alpha bepaald (zie tabel 3.3). Deze bleek bij de leerlingen steeds hoog tot zeer hoog (≥.83) en bij de docenten voldoende hoog (≥.77), behalve bij taakgebied 13 (huidig handelen bij beoordelen van werk en de prestaties van leerlingen; alpha = .69).

Tabel 3.3: De betrouwbaarheden α en itemrestcorrelaties r_{ir} per taakgebied, bij leerlingen en docenten bij huidig en gewenst handelen

Taakgebied	Leerlingen				Docenten			
	Huidig		Gewenst		Huidig		Gewenst	
	α	r_{ir}	*α*	r_{ir}	*α*	r_{ir}	*α*	r_{ir}
1	.89	≥.29*	.88	≥.49	.81	≥.29*	.84	≥.39
2	.84	≥.41	.84	≥.52	.85	≥.49	.87	≥.57
3	.87	≥.45	.83	≥.38	.78	≥.27*	.81	≥.37
4	.88	≥.45	.88	≥.44	.79	≥.24*	.87	≥.33*
5	.90	≥.57	.88	≥.57	.77	≥.33*	.82	≥.45
6	.91	≥.59	.88	≥.54	.84	≥.40	.88	≥.48
7	.89	≥.57	.87	≥.56	.79	≥.31*	.83	≥.45
8	.92	≥.43	.92	≥.48	.91	≥.52	.92	≥.50
9	.89	≥.46	.90	≥.54	.83	≥.28*	.88	≥.39
10	.89	≥.50	.88	≥.54	.79	≥.25*	.88	≥.45
11	.94	≥.66	.92	≥.66	.87	≥.35	.90	≥.46
12	.93	≥.60	.90	≥.51	.85	≥.47	.90	≥.58
13	.90	≥.50	.88	≥.53	.69	≥.18	.80	≥.32*
14a	.91	≥.63	.89	≥.60	.84	≥.41	.85	≥.49
14b	.91	≥.50	.90	≥.50	.88	≥.44	.92	≥.45
15	-	-	-	-	.84	≥.31*	.88	≥.47

*bij r_{ir} ≤.35 betreft het steeds één item uit de schaal; bij alle andere items is de r_{ir} ≥.35

Tabel 3.4 laat zien dat de gemiddelde scores tussen leerlingen en docenten en tussen huidig en gewenst handelen kunnen variëren afhankelijk van het taakgebied. De standaarddeviatie bij het huidig handelen van docenten varieert per taakgebied tussen .35 en .70, met een gemiddelde van 0,5. Om in de communicatie naar de scholen een eenduidige maat te kunnen gebruiken die kan fungeren als norm (minimum) bij de interpretatie van verschillen tussen twee taakgebieden c.q. schaalgemiddelden, met als doel te voorkomen dat schoolleiders en docenten teveel betekenis toekennen aan kleine verschillen, is de gemiddelde standaardmeetfout berekend van alle paarsgewijze verschilscores tussen de 15 schaalscores van het huidig handelen bij de leerlingen en bij de docenten. Aanvullend is ook de standaardmeetfout berekend voor de schalen met de kleinste en de grootste standaardafwijking. De standaardmeetfout van de verschilscore tussen twee taakgebieden bij de leerlingen bedraagt op de gebruikte schaal van 1 tot 5 gemiddeld 0,35 schaalpunt en varieert van .30 tot .41. De standaardmeetfout bij de docenten bedraagt gemiddeld 0,3 schaalpunt en varieert van .22 tot .35. Om in alle gevallen dezelfde marge aan te houden en daarvoor een afgerond getal te kiezen, is gecommuniceerd aan schoolleiders en docenten dat aan verschillen tussen de scores op twee taakgebieden van kleiner dan 0,5 schaalpunt (globaal 1,5 standaardmeetfout) inhoudelijk geen betekenis moet worden toegekend.

Tabel 3.4: Aantal respondenten, gemiddelden en standaarddeviaties per taakgebied

Taakgebied	*Leerlingen*						*Docenten*					
	Huidig			Gewenst			Huidig			Gewenst		
	N	*gem*	*sd*	*N*	*gem*	*sd*	*N*	*gem*	*sd*	*N*	*gem*	*sd*
1	842	4.0	.63	823	4.5	.47	137	4.2	.35	133	4.6	.32
2	841	3.7	.73	819	4.1	.65	132	3.9	.51	129	4.5	.47
3	845	3.8	.66	825	4.3	.48	133	4.0	.40	131	4.4	.39
4	825	3.3	.82	793	3.8	.76	134	3.4	.51	133	3.9	.57
5	845	3.5	.80	814	4.1	.66	130	3.5	.51	128	4.0	.54
6	833	3.6	.83	817	4.1	.65	127	3.7	.54	128	4.2	.54
7	857	3.5	.82	829	4.1	.64	131	3.7	.52	127	4.1	.52
8	864	3.6	.86	836	4.0	.76	136	3.7	.70	132	4.2	.59
9	820	3.6	.79	788	4.0	.70	133	3.4	.55	126	3.8	.59
10	812	3.7	.77	793	4.1	.67	132	3.6	.50	123	4.0	.59
11	812	3.8	.86	788	4.2	.67	137	4.1	.50	133	4.4	.49
12	804	3.7	.86	787	4.1	.68	137	3.7	.51	134	4.3	.54
13	782	3.6	.79	763	4.0	.68	126	3.6	.42	121	4.0	.49
14a	803	3.9	.84	774	4.3	.65	134	4.3	.47	130	4.6	.41
14b	657	3.8	.74	643	4.3	.58	121	3.8	.52	120	4.6	.45
15							119	3.9	.88	108	4.3	.51

De schaalgemiddelden *per docent* (hier niet weergegeven) bleken meestal te variëren tussen 2,5 en 4,5. Tegen die achtergrond is de variatie *tussen de taakgebieden* in de gemiddelden op het hogere aggregatieniveau over alle 141 docenten (zie tabel 3.4) aanzienlijk. Deze variëren namelijk bij het huidig handelen tussen 3,4 en 4,3 en bij het gewenste handelen tussen 3,8 en 4,6. De gemiddelden *op de taakgebieden* over alle leerlingen variëren tussen 3,3 en 4,0 bij het huidig handelen en tussen 3,8 en 4,5 bij het gewenst handelen. Deze verschillen zijn iets minder groot dan bij de docenten maar wel betekenisvol. Zowel bij de leerlingen als bij de docenten verschillen de gemiddelden van het huidig handelen en het gewenst handelen bij alle taakgebieden significant ($p<.001$). Zowel bij het huidig handelen als bij het gewenst handelen verschillen de gemiddelden bij de leerlingen en bij de docenten bij negen van de 15 taakgebieden significant ($p<.05$). Bij het huidig handelen betreft dit de taakgebieden 5, 6, 8, 12, 13 en 15, bij het gewenst handelen de taakgebieden 4, 5, 6, 7, 10 en 13.
De gemiddelden op de schalen 1-14a in de data van de leerlingen verschillen significant tussen de zes scholen, behalve bij de schaal beoordelen ($F(5,776) = 1,917$, $p = .089$). Paarsgewijs vergeleken (post-hoc) zijn er op de meeste schalen wel bepaalde significante verschillen tussen twee of meer scholen. Specifiek kijkend naar school 6 (de enige scenario 1/2 school tussen verder alleen scenario 4 scholen, zie hoofdstuk 1) blijkt dat 15 van de 70 verschillen (14 schalen x 5 scholen) significant zijn, bij de helft van de schalen. Bij twee schalen (4 (nut) en 6 (vaardigheden)) is het gemiddelde van school 6 het laagste van alle scholen. Bij de andere 12 schalen ligt het gemiddelde van school 6 binnen het interval tussen het laagste en het hoogste gemiddelde. De aantallen docenten met volledige gegevens waren op een deel van de scholen te gering (minder dan 10) om deze analyses ook op de data van de docenten te doen.

De factoranalyse op de geconstrueerde schalen per taakgebied liet zien dat sprake is van één dominerende factor. Zie tabel 3.5. Een factoranalyse op de data van de docenten leidde tot twee factoren (eigenwaarden: 5,2 en 4,1), met een cumulatief percentage verklaarde variantie van 66%. Een factoranalyse op basis van de data van de leerlingen leverde slechts één factor op met een eigenwaarde van 10 en een percentage verklaarde variantie van 71%. De hierna uitgevoerde schaalanalyses leverden de volgende uitkomsten op:

- docenten, taakgebieden 4-14a: Cronbach's alpha is .94
- docenten, taakgebieden 1-3: Cronbach's alpha is .83.
- docenten, taakgebieden 1-14a: Cronbach's alpha is .94
- leerlingen, taakgebieden 1-14a: Cronbach's alpha is .97

Tabel 3.5: Uitkomsten factoranalyse over de taakgebieden (ladingen ≥ .50 vet gezet)

Taakgebied	Factorladingen bij de docenten (na rotatie (varimax))		Factorladingen bij de leerlingen (zonder rotatie)
	1e factor	*2e factor*	*1e factor*
1	.17	**.88**	**.78**
2	.24	**.85**	**.78**
3	.31	**.82**	**.82**
4	**.56**	.23	**.79**
5	**.57**	**.61**	**.86**
6	**.65**	**.56**	**.91**
7	**.62**	.46	**.85**
8	**.66**	**.51**	**.85**
9	**.81**	.24	**.87**
10	**.71**	.16	**.84**
11	**.59**	**.58**	**.88**
12	**.73**	.32	**.86**
13	**.79**	.15	**.88**
14a	**.67**	.41	**.84**

Naar aanleiding van deze uitkomsten beschouwen we het gemiddelde over alle 14 taakgebieden als een zinvolle maat voor het totale betreffende repertoire van docenten, met daarbij de volgende aanvullende argumenten:

- de eerste factor bij de docenten omvat bijna alle taakgebieden (11 van de 14);
- bij de berekeningen per docent van de interbeoordelaarsbetrouwbaarheid, de correlatie tussen het profiel van de scores op de schalen volgens de docent en het profiel volgens de leerlingen, en de intraclassecorrelatie, vormen de data van de leerlingen het uitgangspunt;
- de gegevens van de leerlingen zijn primair, aangezien de invloed van het handelen van docenten op de leerlingen verloopt via de perceptie van dit handelen door de leerlingen;
- vergelijkingen tussen docenten en leerlingen vergen dat in beide groepen eenzelfde maat wordt gebruikt.

Alle betreffende 14 taakgebieden vormen samen een betrouwbare schaal, zowel bij de docenten als bij de leerlingen en zowel voor het huidig als het gewenst handelen. De gegevens van 32 docenten en 366 leerlingen bleken bruikbaar voor deze berekeningen. Bij de docenten is de Cronbach's alpha bij de overallschaal van het huidig handelen .93 (alle rir ≥.53). De score op deze schaal varieert van 2,7 tot 4,4 (gemiddelde 3,8 en standaarddeviatie .38). De Cronbach's alpha bij de overallschaal van het gewenste handelen is .94 (alle rir ≥.55). De score op de deze schaal varieert van 3,5 tot 5 (gemiddelde 4,3 en standaarddeviatie .38). Bij de leerlingen is de Cronbach's alpha bij de overallschaal van het huidig handelen .97 (alle rir ≥.69). De score op deze schaal varieert van 1,3 tot 4,9 (gemiddelde 3,7 en standaarddeviatie .69). De Cronbach's alpha bij de overallschaal van het gewenste handelen is .96 (alle rir ≥.55). De score varieert van 2 tot 5 (gemiddelde 4,3 en standaarddeviatie .51).
De gemiddelden op de overallschaal per school in de data van de leerlingen variëren van 3,4 (de scholen 1 en 6) tot 3,8 (school 2). Deze variatie is significant (Anova: F(5, 525) = 4,223, p=.001). Paarsgewijs vergeleken (post-hoc) is geen van de verschillen tussen twee scholen significant. Er is dus geen school die significant afwijkt van een of meer van de andere vijf scholen.
De aantallen docenten met volledige gegevens waren op een deel van de scholen te gering (minder dan 10) om deze analyses ook op de data van de docenten te doen.

3.5.2 De gevonden subtaakgebieden

Bij de ontwikkeling van de vragenlijst was de selectie van de items erop gericht tot betrouwbare schalen per taakgebied te komen. Uit tabel 3.3 blijkt dat dit voor alle taakgebieden ook is gelukt. Uit de factoranalyses per taakgebied bleek echter dat bij negen taakgebieden de tweede eigenwaarde ≥ 1 was. Bij zes taakgebieden was het geroteerde factorpatroon goed te interpreteren en was sprake van een "simple structure" waarbij elk item op precies één factor laadde met een factorlading ≥ 0.5. In tabel 3.6 is als voorbeeld de factormatrix van taakgebied 9 opgenomen. Hieruit blijkt dat binnen dit taakgebied twee groepen kunnen worden onderscheiden van handelingen die relatief vaak samengaan. Deze groepen zijn *sub*taakgebieden genoemd.
Bij de drie andere taakgebieden, waarbij de 2^{e} eigenwaarde in de factoranalyse ≥ 1.0 was (respectievelijk 1.0, 1.1 en 1.1.), konden geen inhoudelijk goed interpreteerbare subtaak-gebieden worden onderscheiden. Dit waren de taakgebieden 1 (zorgen voor een goed leerklimaat), 3 (geven van les, uitleg en instructie) en 13 (beoordelen van het werk en de prestaties van leerlingen).
De zes taakgebieden waarbij inhoudelijk wel goed interpreteerbare subtaakgebieden konden worden onderscheiden, zijn: 4 (leerlingen het nut en de samenhang laten zien), 6 (leerlingen stimuleren en vaardigheden aanleren), 7 (inspelen op verschillen tussen leerlingen), 9 (leerlingen eigen keuzes laten maken en hen daarbij ondersteunen), 10 (leerlingen hun eigen leerproces laten plannen en bewaken) en 12 (geven van feedback aan leerlingen).

Tabel 3.6: Voorbeeld van geroteerde factormatrix met twee factoren (taakgebied 9)

Items van taakgebied 9: Leerlingen eigen keuzes laten maken en hen daarbij ondersteunen	Factorladingen	
	1e factor	*2e factor*
De leraar biedt leerlingen meer mogelijkheden waaruit ze kunnen kiezen	**.59**	.46
De leraar zet de voor- en nadelen van verschillende keuzes op een rij	**.73**	.29
De leraar laat leerlingen bij opdrachten zelf een aanpak bepalen	.33	**.66**
De leraar laat leerlingen bij opdrachten zelf een planning maken	.13	**.69**
De leraar laat leerlingen zelf kiezen of ze met anderen willen samenwerken	.24	**.70**
De leraar laat leerlingen zelf kiezen met wie ze willen samenwerken	.16	**.77**
De leraar bespreekt met de leerlingen de redenen voor bepaalde keuzes	**.74**	.34
De leraar laat de leerlingen verantwoordelijkheid dragen voor hun eigen keuzes	.32	**.58**
De leraar vraagt de leerlingen of hun keuzes nog leiden tot problemen	**.79**	.24
De leraar laat leerlingen uitwisselen wat ze hebben gekozen en waarom	**.83**	.21
De leraar laat leerlingen uitwisselen wat hun ervaringen zijn met hun keuzes	**.85**	.15

Dat het nu juist de zes genoemde taakgebieden zijn waarbij subtaakgebieden kunnen worden onderscheiden, kan komen doordat bij de andere taakgebieden minder gevarieerde items zijn opgenomen en dat zou dan eventueel kunnen worden gezien als een beperking van de vragenlijst. Omdat voor alle taakgebieden het ontwikkelproces op dezelfde manier verliep en de docenten en schoolleiders die bij de ontwikkeling en try-out waren betrokken regelmatig aangaven de items min of meer dekkend te achten voor het totale professionele handelen van docenten, lijkt deze verklaring echter niet direct voor de hand te liggen.

Een andere interpretatie is dat de genoemde zes taakgebieden onderdeel zijn van het proces van onderwijsvernieuwing en van verbreding van het repertoire waarover docenten moeten beschikken, zoals we dat de laatste 15 jaar, ook internationaal, kunnen waarnemen. Als docenten nog bezig zijn zich het handelen op een taakgebied eigen te maken, ligt het immers voor de hand dat zij daarbinnen vooralsnog eigen accenten leggen, gebaseerd op hoe zij gewend zijn te werken en op hun persoonlijke professionele opvattingen. Dat kan dan in de antwoorden van de leerlingen tot uitdrukking komen, doordat zij het handelen van hun docenten op bepaalde delen van het taakgebied in hogere mate waarnemen en ervaren dan op andere delen.

De twee subtaakgebieden die bij elk van de zes genoemde taakgebieden kunnen worden onderscheiden, betreffen in alle gevallen groepen van docenthandelingen die in een bepaald opzicht binnen dat taakgebied *meer traditioneel* zijn respectievelijk *verder gaan* (meer vernieuwend zijn). Zie het overzicht in figuur 3.3. De betrouwbaarheid van deze subtaakgebieden bleek voldoende tot goed (de Cronbach alpha's lagen bij de docenten tussen .65 en .84 en bij de leerlingen tussen .79 en .90).

Taakgebied	*Subtaakgebied a, minder vernieuwend of gemakkelijker*	*Subtaakgebied b, meer vernieuwend of moeilijker*
4 leerlingen het nut en de samenhang laten zien	Ingaan op de dagelijkse realiteit en op toepassingsmogelijkheden	Leggen van relaties met andere vakken
6 leerlingen stimuleren en vaardigheden aanleren	Wekken van interesses bij de leerlingen	Aanleren van vaardigheden
7 inspelen op verschillen tussen leerlingen	Diverse manieren om als docent in te spelen op verschillen	Inspelen op verschillen door leerlingen zelf, via eigen keuzes bij opdrachten
9 leerlingen laten kiezen en hen daarbij begeleiden	Overlaten van keuzes aan de leerlingen	Actief ondersteunen van leerlingen bij het maken van keuzes
10 leerlingen het eigen leerproces laten plannen en bewaken	Overlaten van plannen en bewaken van het leerproces aan leerlingen	Actief ondersteunen van leerlingen hierbij
12 geven van feedback aan leerlingen	Diverse manieren om feedback te geven	Bevorderen dat leerlingen feedback gebruiken

Figuur 3.3: Interpretaties van de subtaakgebieden binnen zes taakgebieden

Tabel 3.7 toont de gemiddelden van het huidig handelen vanuit het perspectief van de leerlingen en dat van de docenten op de 12 subtaakgebieden. Per taakgebied vertegenwoordigt het tweede subtaakgebied (b) handelingen die verder gaan (meer vernieuwend zijn of moeilijker zijn) dan de handelingen in het eerste subtaakgebied (a). Per taakgebied zien we bij bijna alle meer vernieuwende subtaakgebieden (b) lagere gemiddelden. De verschillen tussen de minder en de meer vernieuwende subtaakgebieden blijken significant ($p \leq 0.01$), zowel bij de docenten als bij de leerlingen.

Tabel 3.7: Verschillen in huidig handelen bij zes taakgebieden tussen beide subtaakgebieden

Taakgebied	Subtaak a		Subtaak b					
	M	*SD*	*M*	*SD*	*n*	*df*	*t*	*p*
Docenten								
4	3.4	.506	3.1	.745	134	133	10.669	<.001
6	3.6	.569	3.5	.649	127	126	4.242	<.001
7	3.7	.516	3.2	.693	131	130	13.074	<.001
9	3.8	.547	3.4	.677	133	132	8.441	<.001
10	3.9	.600	3.6	.589	132	131	3.362	≤.001
12	3.8	.494	3.4	.732	137	136	15.280	<.001
Leerlingen								
4	3.4	.833	3.1	1.03	825	824	13.201	<.001
6	3.6	.903	3.5	.899	833	832	2.665	.008
7	3.7	.844	3.2	.969	857	856	19.083	<.001
9	3.7	.790	3.4	.943	820	819	13.512	<.001
10	3.9	.801	3.6	.873	812	811	12.616	<.001
12	3.8	.863	3.4	1.60	804	803	16.724	<.001

3.5.3 Aanvullende indicaties voor de betrouwbaarheid van het instrument

Op verschillende manieren zijn aanvullende indicaties voor de betrouwbaarheid van het instrument verkregen. Daarvoor is een gedeelte van de data gebruikt, namelijk de gegevens van die docenten waarvoor de scores van 8 of meer leerlingen beschikbaar waren op de 14 taakgebieden. Dit bleek bij 35 van de 141 docenten het geval (met gemiddeld 10 leerlingen). Dit betreft 25% van alle docenten en 40% van alle leerlingen. Bij deze docenten is de jury alpha (interbeoordelaarsbetrouwbaarheid) voor de leerlingen berekend. De jury alpha's bleken bij 25 van de 35 docenten >.50 en bij 17 docenten ≥ .70. Bij 10 docenten blijkt er weinig overeenstemming te zijn (jury alpha <.50).
Aanvullend is, na spiegeling van de betreffende datamatrix om de hoofddiagonaal, per docent de correlatie berekend tussen het profiel (in de zin van de reeks scores op de 14 taakgebieden) van het huidig handelen volgens de docent zelf en het profiel volgens de gemiddelde scores van de leerlingen bij die docent. Bij 21 docenten waren beide gegevens beschikbaar (zie tabel 3.8). De betreffende correlatie bleek bij 9 docenten sterk tot zeer sterk (≥ .50 oplopend tot .83), bij 3 docenten matig sterk (≥ .40 en < .50), bij 4 docenten matig (≥. 30 en < .40) en bij 5 docenten zwak tot matig (< .30). Deze uitkomsten laten zien dat de mate waarin het profiel van de docent correleert met het gemiddelde profiel volgens zijn leerlingen sterk varieert tussen docenten. De gemiddelde correlatie tussen de beide profielen over alle 21 docenten is .43.

Tabel 3.8 De correlaties tussen het profiel van het huidig handelen op de taakgebieden volgens de docent zelf en het profiel volgens de gemiddelde scores van de leerlingen bij die docent

Docent	*r*	*p*	*Docent*	*r*	*p*
1	.37	.18	12	.40	.14
2	.14	.62	13	.43	.11
3	.53	.03*	14	.74	< .01*
4	.53	.04*	15	.72	< .01*
5	.41	.13	16	.15	.59
6	.36	.19	17	.83	< .01*
7	.71	< .01*	18	-.07	.81
8	.36	.19	19	.51	.05*
9	.27	.33	20	.51	.06
10	.61	.01*	21	.31	.26
11	.20	.47			

*: Significant bij α= 0.05

Als indicatie van de mate waarin onderscheid kan worden gemaakt tussen docenten is ten aanzien van het huidig handelen van de docenten de intraclassecorrelatie berekend bij de groep van 35 docenten (zie tabel 3.9). Bij alle taakgebieden is de ICC(1) > .30, bij 10 taakgebieden ≥ .35 en bij 4 taakgebieden > .40. Dit betekent dat bij alle taakgebieden het instrument in ruim voldoende mate onderscheid kan maken tussen docenten. Het percentage verklaarde variantie varieert van 31% (bij de taakgebieden 4

en 8) tot 50% (taakgebied 1). De ICC(2), als indicator van de betrouwbaarheid van de per docent over de leerlingen geaggregeerde scores, is bij alle taakgebieden hoog.
Ook voor de overallschalen zijn de ICC(1) en de ICC(2) bepaald. Bij de overallschaal voor het huidige handelen is de ICC(1): .50 en de ICC(2): .91. Dit betekent dat 50% van de variantie in de scores van de leerlingen samenhangt met het feit dat de leerlingen in groepen (van gemiddeld 10 leerlingen) de vragenlijst hebben ingevuld voor één bepaalde docent. De betrouwbaarheid van de op de docentniveau geaggregeerde scores van leerlingen is hoog.
Bij de overallschaal voor het gewenste handelen is de ICC(1): .10 en de ICC(2): .53. Dit betekent dat de leerlingen vooral individueel verschillen in de mate waarin zij een breed repertoire van pedagogisch-didactisch handelen van een docent als gewenst beschouwen en dat de betrouwbaarheid van het gemiddelde op docentniveau laag is.

Tabel 3.9: Intraclassecorrelatie (ICC(1)) en betrouwbaarheid (ICC(2)) per taakgebied, huidig handelen

Taakgebied	*ICC(1)*	*ICC(2)*
1 Zorgen voor een goed leerklimaat	.50	.91
2 Bijdragen aan de persoonlijke vorming van leerlingen	.37	.86
3 Geven van les, uitleg en instructie	.41	.87
4 Leerlingen het nut en de samenhang laten zien	.31	.82
5 Leerlingen leren zelf opdrachten aan te pakken	.36	.85
6 Leerlingen stimuleren en vaardigheden aanleren	.37	.85
7 Inspelen op verschillen tussen leerlingen	.36	.85
8 Leerlingen laten samenwerken en hen daarbij ondersteunen	.33	.83
9 Leerlingen eigen keuzes laten maken en daarbij ondersteunen	.35	.84
10 Leerlingen hun eigen leerproces laten plannen en bewaken	.34	.84
11 Volgen hoe leerlingen leren en zich ontwikkelen	.35	.84
12 Geven van feedback aan leerlingen	.40	.87
13 Beoordelen van het werk en de prestaties van leerlingen	.32	.83
14a Samenwerken met andere leraren	.46	.90

3.6 De ontwikkeling van een verkorte versie van het instrument

Na afloop van het onderzoek is op verzoek op een aantal van de deelnemende scholen per school voor schoolleiding en teamleiders een presentatie verzorgd over de opbrengsten. Bij de bespreking van de uitkomsten bleken zowel de schoolleiders als de teamleiders betekenis te kunnen geven aan de uitkomsten van de analyses in die zin dat zij het handelen van de docenten (zoals in het instrument weergegeven door docenten zelf en de betreffende leerlingen) inhoudelijk konden duiden in het licht van de ontwikkeling binnen de school als geheel en binnen de verschillende teams in het bijzonder. Meerdere schoolleiders uitten bij deze gelegenheid de behoefte aan een versie van het instrument met minder items (sneller afneembaar) en/of een versie die digitaal afneembaar is (flexibel ook in delen afneembaar). Zij overwogen een dergelijk

instrument regelmatig in te zetten voor monitoring en evaluatie van verandering (via herhaalde metingen) en als basis voor gesprekken over functioneren en professionele ontwikkeling.
Naar aanleiding hiervan is een verkorte versie gemaakt, waarbij tegelijkertijd de zes taakgebieden waarbij sprake was van twee subtaakgebieden (zie figuur 3.3) zijn gesplitst, waardoor het instrument in totaal 22 taken ging bevatten: 1-20 voor de oorspronkelijke 1-14a, 21 voor 14b (met meer docenten een grotere groep leerlingen begeleiden) en 22 voor 15 (professionele ontwikkeling).
Voor de eerste 20 taken werd gestreefd naar een reductie van de oorspronkelijke 144 naar maximaal 100 items, dus gemiddeld 5 per taak. Bij de gesplitste taakgebieden behoefde daartoe vaak slechts één item te worden geschrapt, bij de overige (niet-gesplitste) taakgebieden moest steeds ongeveer de helft van de items komen te vervallen. Gehandhaafd zijn steeds de items met een relatief hoge itemrestcorrelatie en een vlot leesbare en inhoudelijk duidelijke formulering, die samen de betreffende taak zo goed mogelijk dekken.
Daarnaast is de formulering van alle items van taak 8 (vaardigheden aanleren) aangepast: “Bij de leraar leer je …” is veranderd in “De leraar leert je …” omdat deze formulering beter een activiteit uitdrukt. Ook is de formulering van alle items van taak 21 (samen met andere leraren een groep leerlingen begeleiden) aangepast. In de oude versie waren alle items in meervoud gezet: “Ze zijn benaderbaar …”, maar omdat leerlingen de vragenlijst invullen voor een bepaalde docent is dit veranderd in “De leraar is benaderbaar …”.
Verder zijn bij 12 items, verspreid over de taken, een of twee woorden weggelaten of veranderd, om de bedoelde inhoud zo mogelijk nog iets duidelijker te doen uitkomen. Een voorbeeld van een item waarbij een of meer woorden zijn verwijderd is: “De leraar zorgt ervoor dat leerlingen elkaar niet voor schut zetten *of afmaken*.” (‘of afmaken’ is verwijderd). Een voorbeeld van een item waarbij woorden zijn veranderd is: “De leraar laat leerlingen bij opdrachten zelf de planning bepalen” was aanvankelijk “De leraar laat leerlingen bij opdrachten zelf een planning maken”.
Al met al betreffen deze veranderingen rond 20% van de items in de verkorte versie. Het overgrote deel van de geselecteerde items is dus ongewijzigd gebleven. Bij taken waarbij door het weglaten van items de volgorde van de items minder logisch zou kunnen overkomen, is die volgorde aangepast.
De nieuwe versie heeft 99 items voor de taken 1-20, 6 items voor taak 21 (samen met andere leraren een groep leerlingen begeleiden; alleen in te vullen indien dit op de school van toepassing is) en 12 items voor taak 22 (professionele ontwikkeling; alleen in te vullen door docenten). Van die 99 items zijn er drie nieuw; deze zijn op inhoudelijke gronden toegevoegd bij de taken 8, 21 en 22, om die taken nog iets beter te dekken. Een voorbeeld bij taak 8 (leerlingen vaardigheden aanleren) is: “De leraar leert je hoe je een goed verslag kunt schrijven.”

Tabel 3.10 geeft een overzicht van de omzetting van de oude in de nieuwe versie. De uitkomsten hebben uiteraard betrekking op de items waarvan data beschikbaar waren: in totaal 98 items voor de taken 1-20, 5 voor taak 21 en 11 voor taak 22.
De meeste nieuwe taken correleren hoog tot zeer hoog met de (sub)taakgebieden waaruit ze zijn ontstaan (meestal .95 of hoger) (zie tabel 3.10). Bij de taken 5, 12 en 18 is deze correlatie met de oude subtaakgebieden 1, doordat daar geen items zijn geschrapt. Bij de taken 8, 14 en 15 zijn de correlaties wat lager. Bij taak 8 komt dit doordat in het betreffende subtaakgebied meerdere items zijn geschrapt en bij de taken 14 en 15 doordat uit beide taken één item naar de andere taak is verplaatst omdat dit daar inhoudelijk bij nader inzien beter past.
Bij de leerlingen variëren de gemiddelde schaalscores over alle leerlingen tussen de schalen bij het huidige handelen van 3,1 tot 4,0 en bij het gewenste handelen tussen 3,6 en 4,5. Bij de docenten variëren de gemiddelde schaalscores over alle docenten tussen de schalen bij het huidige handelen van 3,0 en 4,2 en bij het gewenste handelen tussen 3,7 en 4,6.
Om na te gaan of de eerder gestelde norm van 0,5 schaalpunt voor het interpreteren van verschilscores tussen twee taakgebieden ook gebruikt kan worden als norm voor het interpreteren van verschilscores tussen twee taken, is ook voor de uitkomsten op de taken de gemiddelde standaardmeetfout berekend, over het huidig handelen op de eerste 20 taken, bij zowel de docenten als de leerlingen. Aanvullend is de standaardmeetfout berekend bij de taken met de kleinste respectievelijk grootste standaardafwijking. De gemiddelde standaardmeetfout van de verschilscores bedraagt bij de docenten .43 schaalpunt (variërend van .32 bij de kleinste standaardafwijking tot .54 bij de grootste) en bij de leerlingen .55 schaalpunt (variërend van .41 bij de kleinste standaardafwijking tot .56 bij de grootste). Ook bij de verkorte versie is een norm van minimaal 0,5 schaalpunt voor het interpreteren van een verschil tussen de gemiddelden op twee taken dus goed bruikbaar.
Al met al kunnen we stellen dat de verkorte versie in hoge mate vergelijkbaar is met de uitgebreide versie en de oorspronkelijke inhouden van de taakgebieden nog steeds goed dekt. De schaalbetrouwbaarheden zijn voldoende tot goed, bij docenten en bij leerlingen (zie tabel 3.10). In hoofdstuk 4 wordt verslag gedaan van analyses op de data met betrekking tot de verkorte versie.

Tabel 3.10: Samenstelling verkort instrument voor het verkrijgen van een profiel van het Pedagogisch-Didactisch Handelen van docenten (PHD). Oude versie (taakgebieden en subtaakgebieden) en Nieuwe versie (verkort): correlaties en betrouwbaarheden

Oude versie (k = aantal items)			Nieuwe versie (k = aantal items)			Correlatie oud-nieuw**	Cronbach's alpha voor handelen nu	
nr	*k*	*omschrijving*	*nr*	*k**	*omschrijving*		*docenten*	*leerlingen*
1	12	Zorgen voor een goed leerklimaat	1	6	Zorgen voor een goed leerklimaat	.95	.67	.82
2	7	Bijdragen aan de persoonlijke vorming van leerlingen	2	5	Bijdragen aan de persoonlijke vorming van leerlingen	.97	.77	.67
3	11	Geven van les. uitleg en instructie	3	6	Geven van les. uitleg en instructie	.96	.66	.80
4	10	Leerlingen het nut en de samenhang laten zien						
4a	*6*	*Ingaan op dagelijkse realiteit en toepassingsmogelijkheden*	4	4	Leerlingen het nut van de leerstof laten zien	.95	.66	.79
4b	*4*	*Relaties met andere vakken*	5	4	Leerlingen wijzen op verbanden tussen vakken	1	.84	.87
5	10	Leerlingen laten werken aan opdrachten	6	6	Leerlingen leren werken aan opdrachten	.97	.66	.84
6	11	Leerlingen stimuleren en vaardigheden aanleren						
6a	*5*	*Wekken van interesses bij de leerlingen*	7	5	Leerlingen stimuleren en activeren	.97	.76	.83
6b	*6*	*Aanleren van vaardigheden*	8	5	Leerlingen vaardigheden aanleren	.81	.69	.74
7	10	Inspelen op verschillen tussen leerlingen						
7a	*5*	*Diverse manieren om als docent in te spelen op verschillen*	9	4	Inspelen op verschillen tussen leerlingen	.97	.67	.80
7b	*5*	*Inspelen op verschillen via eigen keuzes bij opdrachten*	10	4	Inspelen op verschillen door middel van opdrachten	.98	.68	.81
8	11	Leerlingen laten samenwerken en hen daarbij ondersteunen	11	4	Leerlingen laten samenwerken en daarbij ondersteunen	.91	.77	.77
9	11	Leerlingen keuzes laten maken en hen daarbij ondersteunen						
9a	*5*	*Overlaten van keuzes aan de leerlingen*	12	5	Leerlingen eigen keuzes laten maken	1	.65	.69
9b	*6*	*Actief ondersteunen van leerlingen bij maken van keuzes*	13	5	Leerlingen ondersteunen bij het zelf maken van keuzes	.99	.78	.87
10	10	Leerlingen hun eigen leerproces laten plannen en bewaken						
10a	*5*	*Overlaten v. plannen en bewaken v.h. leerproces aan de lln.*	14	5	Leerlingen hun eigen leerproces laten sturen	.68	.78	.85
10b	*5*	*Actief ondersteunen van leerlingen hierbij*	15	5	Leerlingen ondersteunen bij het sturen van hun leerproces	.73	.73	.77
11	11	Volgen hoe leerlingen leren en zich ontwikkelen	16	5	Volgen wat leerlingen leren en hoe zij zich ontwikkelen	.98	.79	.89
12	11	Geven van feedback aan leerlingen						
12a	*7*	*Diverse manieren om feedback te geven*	17	6	Geven van feedback aan leerlingen	.99	.81	.83
12b	*4*	*Bevorderen dat leerlingen feedback gebruiken*	18	4	Bevorderen dat feedback wordt gebruikt	1	.78	.86
13	11	Beoordelen van het werk en de prestaties van leerlingen	19	6	Beoordelen van het werk en de prestaties van leerlingen	.95	.64	.84
14	8	Samenwerken met andere leraren	20	5	Open staan voor leren en samenwerken	.97	.78	.87
14x	12	Meer leraren/begeleiders. in een grote ruimte	21	6	Samen met andere leraren een groep lln. begeleiden	.96	.82	.84
15	16	Professionele ontwikkeling	22	12	Professionele ontwikkeling	.97	.77	n.v.t.

* Bij de taken 8, 21 en 22 waren 4, 5 resp. 11 items voor de verkorte versie geselecteerd. Bij elk van deze drie taken is daarna één nieuw item toegevoegd.
** In de nieuwe versie is één item verplaatst van taak 15 naar taak 14.

3.7 Conclusies en discussie: waarde en gebruiksmogelijkheden van het instrument

Conclusies

Ontwikkeltraject

Om tot een valide en betrouwbaar instrument te komen waarmee het pedagogisch-didactisch handelen van docenten op scholen voor voortgezet onderwijs in kaart kan worden gebracht en dat op scholen bruikbaar is voor ontwikkelingsdoeleinden, is een werkwijze gevolgd waarin uitkomsten van wetenschappelijke theorie en onderzoek, inzichten en wensen uit in onderwijsbeleid en onderwijspraktijk en psychometrische analyses zijn gecombineerd.

Het begrip handelen van docenten is afgebakend ten opzichte van verwante begrippen zoals vaardigheid, gedrag en competentie. De dimensie van het pedagogisch-didactisch handelen is afgebakend ten opzichte van andere dimensies van het handelen van docenten, zoals het interpersoonlijke en het vakdidactische. In een iteratief proces van het bestuderen van wetenschappelijke literatuur en documenten van scholen en landelijk beleid, het genereren van een groot aantal items door een aantal docenten van vernieuwende scholen met behulp van een voorlopige indeling in rubrieken, en het ordenen, selecteren en herformuleren van deze items in het licht van de literatuur en documenten, is de interne structuur van het pedagogisch-didactisch handelen uitgewerkt in een 15-tal taakgebieden met in totaal rond 150 items. De ontwikkelde schriftelijke vragenlijst is ingevuld door 141 docenten en 900 leerlingen op een zestal scholen. Uit analyses op de verkregen data bleek dat bij zes van de taakgebieden een betekenisvolle onderverdeling kan worden gemaakt tussen een meer en een minder vernieuwend subtaakgebied. Verdere analyses hebben geleid tot een verkorte versie met in totaal 22 taken en rond 100 items.

Indicaties voor validiteit

Doordat de interne structuur van het begrip pedagogisch-didactisch handelen is geoperationaliseerd op basis van wetenschappelijke literatuur vanuit drie verschillende invalshoeken (literatuur over algemene didactische modellen, rapportages over meta-analyses van factoren die de effectiviteit van onderwijs beïnvloeden, en publicaties met theorie en onderzoek naar leren en leeromgevingen), mag ervan uit worden gegaan dat de indeling in taakgebieden die hieruit is voortgekomen het domein van het pedagogisch-didactisch handelen in voldoende mate dekt. Dit betekent dat door de gekozen aanpak en uitvoering van het ontwikkeltraject de construct- en inhouds-validiteit van het instrument als voldoende mogen worden beschouwd. Dit wordt onderstreept door de reacties van de schoolleiders en docenten gedurende verschillende momenten tijdens het ontwikkeltraject, bij de afname van het instrument en tijdens de besprekingen bij de terugkoppeling van de gegevens. Zij gaven aan dat volgens hen geen items behoefden te worden geschrapt en ook niet iets ontbrak. De uitkomsten van

de try-out maakten bij de terugkoppeling aan de schoolleiders en teamleiders in hun ogen betekenisvolle verschillen zichtbaar in scores op de taakgebieden tussen teams en tussen scholen. Ze konden deze verschillen duiden op grond van hun kennis en inzicht in het feitelijk handelen van de docenten in de verschillende teams. Docenten gaven aan de onderscheiden taakgebieden en de items daarbinnen ter herkennen en als min of meer dekkend te zien. Zij vonden de in de items beschreven handelingen betekenisvol in relatie tot het beoogde handelen van docenten in de vernieuwende context op hun scholen.
Daarnaast zijn ook aanwijzingen uit de onderwijsconcepten van de scholen en uit het landelijk beleid betrokken bij de operationalisering van het begrip pedagogisch-didactisch handelen en de indeling in taakgebieden. Dit heeft er mede toe bijgedragen dat met het instrument een breed handelingsrepertoire in kaart kan worden gebracht.
In de volgende stap van de operationalisering van de taakgebieden werden docenten en leerlingen betrokken bij het genereren van items, zodat recht zou worden gedaan aan het dagelijks handelen van docenten. Deze aanpak had ook tot doel te bevorderen dat de formuleringen van de items voor docenten en leerlingen herkenbaar zouden zijn en zouden aansluiten bij hun taalgebruik, hetgeen ook bijdraagt aan de validiteit van de met het instrument te verkrijgen gegevens (ecologische validiteit).
De taakgebieden en de items per taakgebied kunnen worden beschouwd als mogelijke criteria (‘content standards’) voor het gewenste handelingsrepertoire van docenten. Het beantwoorden van de vragen in het instrument kan daardoor bijdragen aan het proces van internalisatie van die criteria, waardoor docenten hun eigen handelen met die criteria kunnen gaan vergelijken. Door dit leereffect kan een docent met het instrument mogelijk een meer valide uitkomst verkrijgen dan waar hij zelf wellicht in eerste instantie toe in staat is. Samen met de mogelijkheid om feedback van anderen te verkrijgen en te benutten om de verkregen uitkomsten aan te relateren, draagt ook dit bij aan de validiteit van het instrument (*systemic validity*).

Indicaties voor betrouwbaarheid
Om een indicatie te krijgen van de betrouwbaarheid van het instrument zijn aanvankelijk diverse analyses uitgevoerd op de uitkomsten op basis van de indeling in taakgebieden. Met schaalanalyses kon worden vastgesteld dat elk taakgebied een betrouwbare schaal vormt. Met factoranalyses per taakgebied werden inhoudelijk betekenisvolle subschalen gevonden.
Vervolgens zijn aanvullende indicaties voor de betrouwbaarheid van het instrument berekend. De berekening van de interbeoordelaarsbetrouwbaarheid maakte zichtbaar dat bij de helft van de hierbij betrokken docenten de scores van de leerlingen die de vragenlijst voor dezelfde docent hebben ingevuld in vrij hoge mate overeenstemmen, bij bijna een kwart van de docenten redelijk en bij ruim een kwart niet. De correlatie tussen het profiel van de scores op de taakgebieden volgens de docent en het profiel volgens de gemiddelde scores van de leerlingen bij die docent varieerde sterk tussen docenten. Bij rond 40% van de docenten werd een sterk verband gevonden, bij nog

eens 35% een matig tot matig sterk verband, en bij het resterende kwart weinig of geen verband. Beide hiervoor beschreven indicatoren (de interbeoordelaarsbetrouwbaarheid bij de leerlingen en de correlatie tussen de profielen volgens de docent en volgens zijn leerlingen) hingen niet systematisch samen.
Als indicatie van de mate waarin op basis van de gegevens van leerlingen onderscheid kan worden gemaakt tussen docenten is over de docenten waarbij voldoende leerlingen de vragenlijst hebben ingevuld op de data van de leerlingen over het huidig handelen de intraclassecorrelatie ICC(1) berekend. Hieruit bleek dat bij alle taakgebieden het instrument in ruim voldoende mate onderscheid kan maken tussen docenten. Het percentage verklaarde variantie varieerde daarbij van 31% tot 50%. De ICC(2), als indicator van de betrouwbaarheid van de geaggregeerde scores, bleek bij alle taakgebieden hoog te zijn.
Op basis van de hierboven beschreven uitkomsten kan worden gesteld dat de schalen in voldoende mate betrouwbaar zijn. De verschillen in gemiddelde scores tussen leerlingen en docenten op de taakgebieden en tussen huidig en gewenst handelen laten een gevarieerd beeld zien, afhankelijk van het gekozen taakgebied. Bij de interpretatie van de verschillen tussen twee taakgebieden is naar de scholen toe een maat aangehouden van 0,5 schaalpunt, met dien verstande dat aan verschillen < 0,5 inhoudelijk geen betekenis moet worden toegekend. Deze maat komt bij deze gegevens overeen met globaal 1,5 x de standaardmeetfout.
De scores op de taakgebieden over het huidige en het gewenste handelen bij de leerlingen en bij de docenten vormden in alle vier de gevallen een betrouwbare overallschaal. Ten aanzien van het huidige handelen bleek het instrument op basis van de overallscore bij de leerlingen in hoge mate te discrimineren tussen docenten (de ICC(1) is 0.50) en de betrouwbaarheid van de op docentniveau geaggregeerde scores van leerlingen (de ICC (2)) was navenant hoog. De ICC(1) en ICC(2) van de overallscore bij de leerlingen ten aanzien van het gewenste handelen van hun docenten waren onder de maat. Ten aanzien van het gewenste handelen maken leerlingen minder onderscheid tussen docenten.

Indicaties voor bruikbaarheid
Een belangrijk aspect van de bruikbaarheid van het instrument is de mate waarin het terugkoppelen van de met het instrument verkregen resultaten bijdraagt aan de dialoog tussen de schoolleiding en docenten, tussen docenten onderling, en tussen docenten en leerlingen. Steeds wanneer de uitkomsten per school aan schoolleiders, teamleiders en docenten werden gepresenteerd (door mondelinge presentaties en door rapportages op papier met tabellen en grafieken aangevuld met een interpretatie van de resultaten door de onderzoekers), leidde dit tot vragen, opmerkingen en discussies, soms met de onderzoeker maar vooral ook onderling.
Het reflecteren op de uitkomsten brengt duidelijk het gesprek op gang, maar voor het samen geven van betekenis aan de uitkomsten (via interpretaties en beslissingen over te ondernemen acties en te stellen prioriteiten) is een voortgezette en inhoudelijke dialoog

gewenst. Voor zover bekend, heeft een dergelijke inhoudelijke dialoog nog slechts beperkt plaatsgevonden (voornamelijk op twee van de zes scholen).
Een ander belangrijk aspect van de bruikbaarheid van het instrument en de daarmee te verzamelen gegevens betreft de organisatie van de afname. Voor het verkrijgen van valide en betrouwbare gegevens moet het afnemen in een school van het instrument bij leerlingen en docenten goed worden georganiseerd, en dit vergt uiteraard een zekere inspanning. Voor zover een school (vooralsnog) niet optimaal in staat is om zelf de afname van het instrument goed te organiseren en/of de verkregen resultaten te duiden en te benutten, vermindert dit de bruikbaarheid van het instrument in de situatie waarin de school het instrument zelfstandig en op eigen kracht zou willen inzetten.
De constructie van de verkorte versie, waarbij meer en minder vernieuwende subtaakgebieden zijn ondergebracht in afzonderlijke taken, maakt de interpretatie van de uitkomsten wellicht eenduidiger dan bij de oorspronkelijke versie in taakgebieden. Het aantal taken waarover wordt gerapporteerd wordt daardoor echter wel groter, met als nadeel dat het geheel lastiger is te overzien en daardoor wellicht ook moeilijker te interpreteren.

Discussie

Het is mogelijk gebleken een instrument te ontwikkelen om het pedagogisch-didactisch handelen van docenten in kaart te brengen door het begrip handelen te positioneren ten opzichte van verwante begrippen zoals vaardigheid, competentie en gedrag door dit begrip in te vullen als intentioneel gedrag dat door anderen als zodanig kan worden waargenomen en ervaren (vergelijk Reynolds, 1992; Roelofs et al., 2008; Schulman, 1987; Vermunt et al., 1999), door het pedagogisch-didactische op te vatten als een eigen dimensie in het handelen van docenten naast onder meer het interpersoonlijke, het organisatorische en het vakdidactische (vergelijk Bransford et al., 2005; Brekelmans et al., 2000), en door het te specificeren in termen van taken (in plaats van rollen, instructiestrategieën of leeromgevingen) (vergelijk Brown & Campione, 1996; Collins et al., 1989; De Corte, 2000).
Door de keuze en formulering van taken mede te baseren op de onderwijsconcepten van scholen en door bij de operationalisering van de taken in uitspraken over docenthandelen een aantal docenten te betrekken (en het instrument ook bij een aantal leerlingen uit te proberen) is bijgedragen aan de herkenbaarheid en ecologische validiteit van het instrument (vergelijk Beijaard, 1990; Carter, 1990) en aan de systemische validiteit (Frederiksen & Collins, 1989). Dit draagt bij aan de kans dat het instrument en de daarmee te verkrijgen data bij docenten de reflectie op en evaluatie van het eigen handelen stimuleren en ondersteunen zodat het instrument inderdaad, zoals beoogd, kan worden gebruikt voor leren en professionele ontwikkeling van docenten (vergelijk Calderhead & Gates, 1993; Grimmett & MacKinnon, 1992; Russell & Mundy, 1991).

Uit het feit dat de terugrapportage op de scholen naar de schoolleiders, teamleiders en docenten niet steeds en niet vanzelf leidde tot de gewenste kennisname, dialoog, interpretatie en besluitvorming, kan worden afgeleid dat het benutten van dergelijke datafeedback mogelijk zelf een leerproces vergt, waarbij voortgezette ondersteuning wenselijk is.
De uitkomsten die in dit onderzoek met het instrument zijn verkregen, laten enkele opvallende verschillen zien tussen de manieren waarop docenten en hun leerlingen dit handelen waarnemen. De mate waarin de leerlingen onderling overeenstemmen in hun beoordeling van het huidig handelen van hun docent verschilt tussen docenten onderling. Dit kan duiden op verschillen tussen de leerlingen, verschillen tussen leerlingen in hun relatie met de docent, en verschillen in de mate waarin de docent in zijn handelen differentieert tussen leerlingen. Tussen de docenten zijn opmerkelijke grote verschillen geconstateerd in de mate waarin het profiel volgens de docent van zijn huidig handelen en het gemiddelde profiel volgens de leerlingen bij die docent met elkaar overeenkomen. Dit kan erop wijzen dat niet elke docent goed kan inschatten hoe zijn of haar eigen handelen bij leerlingen overkomt.
Uit het feit dat de overallscore van de leerlingen ten aanzien van het *gewenste* handelen niet duidelijk onderscheid maakt tussen de docenten kan worden afgeleid dat bij het beantwoorden door de leerlingen per taakgebied per item van de vragen over het huidige en het gewenste handelen hun score ten aanzien van het gewenste handelen niet sterk is gebaseerd geweest op hun score op het huidige handelen, maar meer op een algemener beeld van wat zij als gewenst handelen van een docent zien (op de betreffende school).
De afname van het instrument is gebleken geen sinecure te zijn. Het invullen van de uitgebreide versie van het instrument bleek in de ervaring van een aantal leerlingen (en ook sommige docenten) veel tijd te kosten (voor sommigen te veel). De verkorte versie bleek duidelijk minder arbeidsintensief. Vmbo-t leerlingen en leerlingen met dyslexie hadden echter ook dan nog relatief veel tijd nodig (ongeveer drie kwartier). Voor hen bleek het moeilijk de concentratie op te brengen om de hele lijst achter elkaar in te vullen. Het verdelen van de invulling over meer momenten en/of het gericht selecteren van de taken waar het op dat moment om gaat, zijn suggesties om deze bezwaren weg te nemen.
Tijdens het ontwikkelproces van de uitgebreide versie is verschillende malen met contactpersonen op de scholen het belang besproken van een goede organisatie van de afname. Het blijkt in de dagelijkse praktijk voor scholen niet eenvoudig om volgens een strakke planning systematisch gegevens te verzamelen bij grote groepen docenten en leerlingen. Bij de analyse van de uit de oorspronkelijke vragenlijst gedestilleerde verkorte versie bleek dat bij slechts 35 van de 141 docenten die hadden deelgenomen acht of meer leerlingen de betreffende vragen hadden ingevuld. Bij de andere docenten waren de bij hun leerlingen verzamelde gegevens hierdoor minder bruikbaar op individueel niveau.

Een goede coördinatie, organisatie en timing van afname van het instrument is een belangrijke voorwaarde voor een betrouwbare en bruikbare gegevensverzameling. Om te beginnen moet de keuze van de docenten en leerlingen die meedoen weloverwogen plaatsvinden, in het licht van het beoogde gebruik van de uitkomsten, moet duidelijk zijn welke leerlingen de vragenlijst zullen invullen voor welke docent, en moet een geschikt tijdstip voor de afname worden vastgesteld. Ook is een zorgvuldige introductie nodig waarbij voor de deelnemende docenten en leerlingen een duidelijk beeld wordt geschetst van de bedoeling en de te verwachten opbrengst en wat daarmee zal worden gedaan. Er moeten heldere afspraken worden gemaakt over wie welke gegevens zal ontvangen, waarbij de privacy van de docenten en leerlingen (ook in hun ogen) voldoende moet worden gewaarborgd. De leerlingen moeten weten voor welke docent ze de vragen moeten beantwoorden en bij het invullen die docent ook voor ogen houden, en er moet voor worden gezorgd dat daadwerkelijk voldoende leerlingen per docent de vragenlijst invullen en dat ook volledig en zorgvuldig doen. Ten slotte moeten de verkregen gegevens zorgvuldig worden verzameld en bewaard.

De uitkomsten geven aanleiding tot vervolgonderzoek. Ter nadere validering van het instrument is het wenselijk de relatie te onderzoeken tussen de dimensie van het pedagogisch-didactisch handelen en andere dimensies van het handelen van docenten, zoals bijvoorbeeld het interpersoonlijk gedrag, en ook om te onderzoeken hoe de uitkomsten op het instrument al dan niet samenhangen met kenmerken van docenten, leerlingen en scholen. Hoofdstuk 4 beschrijft de resultaten van een onderzoek dat hierop was gericht.

Ook kan worden nagegaan in hoeverre de taakgebieden respectievelijk de taken een Raschschaal vormen (vergelijk wat Van de Grift, Van der Wal en Torenbeek (2011) hebben gedaan ten aanzien van het pedagogisch-didactisch handelen van leraren in het basisonderwijs).

Daarnaast kunnen de uitkomsten die met dit instrument kunnen worden verkregen, worden vergeleken met de uitkomsten bij gebruik van andere methoden om zicht te krijgen op het pedagogisch-didactisch handelen van docenten, zoals bijvoorbeeld observaties van lesssituaties, mondelinge bevraging van docenten over hun handelen, en meer gespecificeerde bevraging van leerlingen. De hoofdstukken 5 en 6 beschrijven twee onderzoeken waarin hieraan is gewerkt, toegespitst op twee van de onderscheiden taakgebieden.

Ook is het interessant te onderzoeken in welke mate de inhoud van het instrument (de taken) en de scores daarop in relatie kunnen worden gebracht met de vernieuwende onderwijsconcepten van de scholen. In hoofdstuk 7 wordt verslag gedaan van een dergelijk onderzoek.

Tot slot ligt vervolgonderzoek voor de hand naar de mate waarin het pedagogisch-didactisch handelen van docenten zoals dit kan worden vastgesteld met het ontwikkelde instrument samenhangt met de leerprestaties van de leerlingen.

4 Bijdragen aan validering van de verkorte vragenlijst (PDH)

4.1 Inleiding en doelstelling

Op verzoek van een aantal schoolleiders is een instrument ontwikkeld waarmee het pedagogisch-didactisch handelen van docenten in kaart kan worden gebracht (zie hoofdstuk 3). Van dat instrument is vervolgens een verkorte en web-based versie gemaakt. De ontwikkeling van de uitgebreide versie van het instrument is gebaseerd op een analyse van het begrip pedagogisch-didactisch handelen vanuit relevante wetenschappelijke literatuur, beleidsdocumenten, onderwijsconcepten van scholen en de inbreng van een aantal docenten, leerlingen en schoolleiders. Op die manier is de gewenste validiteit van het instrument (construct- en inhoudsvaliditeit) al zoveel mogelijk ingebouwd (zie Borsboom, Mellenbergh & Van Heerden, 2004; De Groot, 1971; Messick, 1995b). Doordat de inhoud van het instrument mede is gebaseerd op in scholen gewenst handelen van docenten zal het gebruik ervan ook kunnen bijdragen aan de discussie over en reflectie op dat handelen en daarmee kunnen bijdragen aan het leerproces van individuele docenten (*systemic validity*; Frederiksen & Collins, 1989).
In de uitgebreide versie is het pedagogisch-didactisch handelen uiteengelegd in 14 taakgebieden (naast een meer voorwaardelijk taakgebied met betrekking tot de eigen professionele ontwikkeling). Factoranalyses per taakgebied brachten aan het licht dat bij zes taakgebieden twee subtaakgebieden kunnen worden onderscheiden. In de verkorte versie zijn deze subtaakgebieden ondergebracht in aparte taken, waardoor de verkorte versie onderscheid maakt in 20 taken. De verkorting betreft het aantal items: het aantal items per taakgebied is gemiddeld ruim 10 (totaal 144), het aantal items per taak is gemiddeld 5 (totaal 99). De kwaliteit van deze meer zuinige operationalisering is op twee manieren gecontroleerd: door berekening van de correlaties tussen de nieuwe taken en de oude (sub)taakgebieden waaruit zij zijn ontstaan, en door berekening van de betrouwbaarheid per taak (itemrestcorrelaties en Cronbach's alpha). De resultaten hiervan waren positief (zie hoofdstuk 3).
Aangezien het om een nieuw instrument gaat en geen andere instrumenten bekend zijn die het brede repertoire van het pedagogisch-didactisch handelen van docenten in kaart kunnen brengen, is verdere validering wenselijk. Daarvoor is het nodig nieuwe data te verzamelen, en omdat de uitgebreide versie van het instrument een relatief groot beroep doet op de tijd van docenten en leerlingen, wat juist de reden was om een verkorte versie te maken, ligt het voor de hand hiervoor nu de verkorte versie te gebruiken. Het verzamelen van nieuwe data met de verkorte versie is ook wenselijk omdat de hiervoor genoemde controle op de kwaliteit ervan was gebaseerd op de reeds beschikbare data die met de uitgebreide versie was verzameld. In zo'n geval moeten de

resultaten worden getoetst aan de hand van nieuwe data, en dit is te meer van belang omdat de items die in de verkorte versie zijn opgenomen in de uitgebreide vragenlijst waren omgeven door andere items (in de taakgebieden waren immers meer items opgenomen), en strikt genomen het lezen van een uitspraak en het beantwoorden van een vraag daarover kan worden beïnvloed door uitspraken die daaraan voorafgaan of daarop volgen (respondenten percipiëren en interpreteren uitspraken vaak in het licht van hun context).
De onderhavige studie richt zich op verdere validering van de verkorte versie van het instrument aan de hand van deels oude en deels nieuw verzamelde data.

4.2 Kader en vraagstelling

In de literatuur over meten en validiteit zijn begrippen en grondslagen die aanvankelijk leken te zijn uitgekristalliseerd in de afgelopen decennia opnieuw bediscussieerd (Borsboom et al., 2004; Crooks, Kane & Cohen, 1996; Ellis, 1968; De Groot, 1971; Haertel, 1985; Kane, 1992; Messick, 1995b). Daarbij is het begrip constructvaliditeit centraal blijven staan (vaak omschreven met "meet ik wat ik wil meten?"). Andere al langer bekende begrippen zijn inhoudsvaliditeit (bij toetsen van kennis in het onderwijs), congruente validiteit (samenhang met een ander instrument dat hetzelfde construct meet) en predictieve validiteit (voorspellend vermogen voor de meting van een ander construct in een toegepaste context)(De Groot, 1971), confirmerende validiteit (voorspelde samenhang met een of meer andere constructen) en discriminante validiteit (ontbreken van (substantiële) samenhang met een of meer andere constructen) (Drenth, 1988).
In de recente discussie is daar door Messick het begrip consequentiële validiteit aan toegevoegd, gericht op het gebruik en de effecten van een meetinstrument (zie bijvoorbeeld Messick, 1995b). Het pleidooi van Messick om voortaan ook de consequentiële validiteit te zien als een belangrijk kwaliteitskenmerk van een meetinstrument en om het begrip constructvaliditeit te handhaven maar nu in een overkoepelende betekenis, is sindsdien door veel auteurs geaccepteerd maar heeft ook tot kritiek geleid (vergelijk Stokking, Van der Schaaf, Jaspers & Erkens, 2004). Borsboom et al. (2004) keerden zich expliciet tegen deze uitbreiding van de betekenis van het begrip validiteit en ook tegen de gedachte dat uitgebreid nagaan hoe een begrip zich blijkens de meting daarvan verhoudt tot andere constructen in een nomologisch netwerk bijdraagt aan validering. Naar hun mening moet worden vastgehouden aan de klassieke betekenis van het begrip (construct)validiteit. Het gaat erom of een instrument meet wat het beoogt, en dat kan en moet worden bevorderd en bewaakt tijdens de constructie van het instrument.
Een andere lijn in de discussie betreft de gedachte dat validering aan de hand van uitkomsten van onderzoek plaatsvindt aan de hand van argumenten, waarbij verschillende soorten argumenten kunnen worden aangevoerd en altijd nieuwe argumenten kunnen worden toegevoegd zodat validering nooit af is (Kane, 1992;

Messick, 1995b), net als in het algemeen bij interpretatie van uitkomsten van onderzoek (vergelijk De Groot, 1971; Stokking, 1984). De door Borsboom et al. (2004) bekritiseerde en ook al eerder door anderen gesignaleerde "wildgroei" (Drenth, 1988) van validiteitsbegrippen kan ook worden gezien als een scala van nuttige soorten argumenten zoals hier bedoeld. Door Crooks et al. (1996) is concreet uitgewerkt hoe metingen (in het onderwijs) onderdeel zijn van een keten van keuzen en activiteiten waarbij de kwaliteit van elke schakel in het geding is, en ook de kwaliteit van deze keten dient te kunnen worden beargumenteerd.

In deze studie beogen we bij te dragen aan de validering van de PDH door middel van empirische argumenten in relatie tot de verschillende hierboven genoemde validiteits-begrippen. Zoals in de inleiding al is opgemerkt, is tijdens de ontwikkeling van de uitgebreide versie gestreefd naar het maken van een valide instrument. Dit is gedaan door het te meten begrip duidelijk af te bakenen en de interne structuur te analyseren. Omdat het pedagogisch-didactisch handelen van docenten betrekking heeft op een breed repertoire van taken gaat het hierbij zowel om constructvaliditeit (het moet gaan om de dimensie pedagogisch-didactisch handelen) als inhoudsvaliditeit (alle relevante taken moeten zijn vertegenwoordigd). (Zie voor een ander voorbeeld van het verbinden van inhoudsvaliditeit en constructvaliditeit in onderwijsonderzoek Haertel, 1985).

Omdat betrouwbaarheid een voorwaarde is voor validiteit is het van belang aan de hand van de nieuwe data de interne consistentie van de taken te controleren. Deze schaalbetrouwbaarheid wordt overigens ook wel aangeduid als homogeniteit of eendimensionaliteit en kan in die zin ook worden gezien als een eigenstandige bijdrage aan de validering. Immers: als een taak geacht wordt een afgebakend en eenduidig construct te zijn, moeten de antwoorden op de uitspraken waarin die taak is geoperationaliseerd positief met elkaar samenhangen. Interne consistentie is zo bezien een niet alleen technische maar ook inhoudelijke aangelegenheid. Op een andere manier is controle op de betrouwbaarheid van belang waar het gaat om de beoordeling door een aantal leerlingen van het handelen van dezelfde docent. Dit betreft dan de interbeoordelaarsovereenstemming. Aanvullend is het ook relevant na te gaan in hoeverre het gemiddelde van de beoordelingen door de leerlingen samenhangt met de zelfbeoordeling door de docent.

Een ander aspect dat vaak wordt gezien als van meer technische aard maar in feite ook een meer inhoudelijke betekenis heeft, is het onderscheidingsvermogen. Het is praktisch van belang dat een instrument onderscheid kan maken (in dit geval tussen docenten), maar het is ook conceptueel relevant omdat een construct dat bij meting amper of niet blijkt te variëren kennelijk geen betekenisvolle variabele kwantiteit betreft (vergelijk Ellis, 1968).

Voor het onderzoeken van de discriminante validiteit van de PDH is in de literatuur een duidelijk aanknopingspunt beschikbaar. In hoofdstuk 3 konden in het handelen van docenten verschillende dimensies worden benoemd, waaronder naast de pedagogisch-didactische dimensie ook de interpersoonlijke dimensie. Naar het interpersoonlijk gedrag van docenten is veel onderzoek gedaan en de VIL (Wubbels et al., 1991) is

hierbij een veel gebruikt instrument. Het ligt dan ook voor de hand te onderzoeken of en hoe beide dimensies, c.q. resultaten van metingen bij dezelfde docenten met beide instrumenten, met elkaar samenhangen.
Bij de VIL worden twee achterliggende dimensies onderscheiden: een nabijheids- (ND, Samen-Tegen) en een invloedsdimensie (ID, Boven-Onder). In een circumplex model met beide dimensies loodrecht op elkaar worden acht sectoren onderscheiden, die elk staan voor bepaald docentgedrag: Leidend (BS), Helpend/Vriendelijk (SB), Begrijpend (SO), Ruimtegevend/Ruimtelatend (OS), Onzeker (OT), Ontevreden (TO), Corrigerend (TB), Streng (BT). Docentgedragingen waarbij relatief hoog wordt gescoord op de subschalen BS, SB en SO, wat minder hoog op OS en laag op OT, TO TB en BT, worden in verband gebracht met gedragingen die leiden tot het voeren van een adequaat klassenmanagement (Admiraal, 1994; Admiraal, Wubbels & Korthagen, 1996), het stimuleren van leeractiviteiten, en het bijdragen aan cognitieve en affectieve leerresultaten bij leerlingen (Brekelmans,1989; Den Brok, 2001).
Door vergelijking van de PDH met de VIL kan worden vastgesteld of met de PDH inderdaad een andere dimensie van docenthandelen wordt gemeten dan met de VIL. Ook kan worden nagegaan in hoeverre en op welke wijze het pedagogisch-didactisch handelen (taken en repertoire) en het interpersoonlijk handelen (sectoren en dimensies) samenhangen. Op grond van de aanname dat sprake is van twee te onderscheiden dimensies van docenthandelen (zie hoofdstuk 3) is de verwachting dat de PDH en de VIL slechts beperkt zullen samenhangen. Daarnaast is plausibel dat interpersoonlijk gedrag met pedagogisch-didactisch positief te waarderen effecten (zoals hierboven in termen van de sectoren van de VIL gespecificeerd) zal samengaan met hogere scores op (een aantal taken van) de PDH.
Naast onderzoek naar de discriminante validiteit is onderzoek naar de confirmerende validiteit zinvol, ook al delen we de mening van Borsboom et al. (2004) dat de betekenis van een instrument c.q. meting op zich duidelijk moet zijn. Indien immers van een variabele grootheid mag worden aangenomen en/of empirisch bekend is dat deze op bepaalde manieren samenhangt met een of meer andere grootheden, dan is het met een instrument niet vinden of niet kunnen aantonen van een dergelijke samenhang een mogelijke contra-indicatie voor de (confirmerende) validiteit ervan.
Er zijn geen duidelijke theorieën of onderzoeksresultaten bekend op basis waarvan de confirmerende validiteit van de PDH als instrument voor pedagogisch-didactisch handelen kan worden nagegaan. Wel kunnen we op basis van de beschikbare gegevens over kenmerken van docenten en leerlingen groepen vormen en die vergelijken voor wat betreft de resultaten op de PDH. Zowel indien er verschillen tussen de groepen blijken te zijn als indien die niet worden gevonden, kan dit bijdragen aan inzicht in de betekenis van dat wat met de PDH wordt gemeten. In het bijzonder kan indien er geen grond is om verschillen tussen bepaalde groepen te verwachten en zulke verschillen er ook inderdaad niet blijken te zijn, dit worden gezien als indicatie voor het ontbreken van ongewenste bias.

Bij de keuze van kenmerken van docenten op basis waarvan voor de hier bedoelde analyses zinvol groepen kunnen worden gevormd, kan wel bij verschillende literatuur worden aangesloten. Uit onderzoek naar het interpersoonlijk handelen van docenten is gebleken dat dit samenhangt met hun geslacht en mate van onderwijservaring (Levy, Den Brok, Wubbels & Brekelmans, 2002; Brekelmans, 2010). In studies naar het bestaan en de kenmerken van loopbaanfasen van docenten is het belang onderzocht van de leeftijd en de mate en aard van de onderwijservaring. Op basis van beschikbare literatuur hierover maakt Huberman (1990) onderscheid in vijf fasen: 1-3, 4-6, 7-18, 19-30 en 31-40 jaar ervaring. Docenten die in verschillende fasen van hun loopbaan verkeren, zouden in hun handelen kunnen verschillen en/of andere aspecten in hun handelen belangrijk kunnen vinden. De genoemde kenmerken (geslacht, leeftijd, mate van onderwijservaring) zullen, naast nog enkele andere, in de analyses van verschillen tussen groepen worden meegenomen.

Wat betreft het begrip consequentiële validiteit delen we de mening van Borsboom et al. (2004) dat de effecten van het gebruiken van een instrument en/of van de uitkomsten daarvan moeten worden losgezien van de constructvaliditeit. Gezien de aanleiding voor en bedoeling van de PDH is het echter wel van groot belang of men in de onderwijspraktijk met dit instrument uit de voeten kan en er ook echt iets aan heeft. Dit is door ons nog niet systematisch onderzocht, maar wel kan verslag worden gedaan van de ervaringen tot nu toe die zijn opgedaan bij het op verzoek (doen) afnemen van het instrument en het terugkoppelen van de uitkomsten daarvan aan schoolleiders, teamleiders en docenten op verschillende scholen. Deze ervaringen zijn opgedaan met de uitgebreide versie van het instrument (op de scholen die deelnamen aan het in hoofdstuk 3 beschreven onderzoek), met de verkorte versie PDH (op de scholen die deelnamen aan het onderzoek waarvan in dit hoofdstuk verslag wordt gedaan) en met twee onderzoeken die elk zijn toegespitst op één van de taakgebieden, namelijk feedback (zie hoofdstuk 5) en beoordeling (zie hoofdstuk 6).

Voor zover de confrontatie met een instrument als de PDH, en het invullen daarvan, op zich al bijdraagt aan vergroting van inzicht in en reflectie op het (potentieel relevante) repertoire van pedagogisch-didactisch handelen van een docent, is sprake van systemic validity, hetgeen ook kan worden beschouwd als specifieke vorm, of onderdeel, van consequentiële validiteit. Onderzoek naar de predictieve validiteit, ten slotte, is voor de PDH nog toekomstmuziek.

De vraagstelling van deze studie luidt: Wat zijn de betrouwbaarheid, het onderscheidingsvermogen en de validiteit van de PDH, op basis van de hiervoor genoemde procedures en criteria?

Deze vraagstelling wordt vooral beantwoord aan de hand van data over het huidig handelen van docenten. In 4.4.5 wordt ingegaan op het verschil tussen huidig en gewenst handelen.

4.3 Opzet en uitvoering van het onderzoek

Dataverzameling

In deze studie zijn drie datasets betrokken. Ten eerste zijn data gebruikt van docenten en leerlingen van zes scholen zoals verkregen in het onderzoek in 2008 waarover in het vorige hoofdstuk is gerapporteerd. Uit deze data zijn de data van de items van de PDH geselecteerd van docenten en leerlingen bij die docenten waarbij gegevens van ten minste acht leerlingen beschikbaar waren.
Ten tweede zijn op vier van deze zes scholen tussen november 2009 en januari 2011 nieuwe data verzameld, op drie scholen (brede scholengemeenschappen) met de web-based versie van de PDH en op één school (een vmbo school) met de PDH op papier. De betreffende schoolleiders waren geïnteresseerd in een nieuwe meting van het pedagogisch-didactisch handelen van docenten op hun school bij andere docenten dan de docenten die aan het onderzoek in hoofdstuk 3 hebben meegedaan. Per school heeft één team meegedaan en de dataverzameling is op elke school door de teamleider georganiseerd. De teamleider van de laatstgenoemde school gaf de voorkeur aan afname op papier.
Ten derde hebben de aan de PDH-afname deelnemende docenten op deze vier scholen die daartoe interesse en bereidheid toonden, meegewerkt aan afname van de VIL, bij henzelf en hun leerlingen, dat wil zeggen enkele dagen na afname van de PDH, bij dezelfde leerlingen. De VIL bestaat uit 77 items verdeeld over de acht eerder genoemde schalen met 9-11 items per schaal en kent een versie voor docenten en een versie voor leerlingen. Voor vmbo-leerlingen is een verkorte versie beschikbaar met 48 items verdeeld over dezelfde acht schalen met 5-7 items per schaal. Bij elke versie worden de twee eerder genoemde dimensies van interpersoonlijk docenthandelen, de nabijheidsdimensie en de invloedsdimensie, volgens een vaststaande berekening op basis van de schaalscores bepaald[1].

Onderzoeksgroep

PDH-data van 2008
De selectie van docenten waarbij van minimaal acht leerlingen data beschikbaar waren, resulteerde in een bestand met PDH-data van 21 docenten en 197 leerlingen. Deze selectie is op verschillende manieren gecontroleerd op eventuele selectiebias.
De 21 docenten verschillen op de beschikbare docentkenmerken niet significant van de overige 120 docenten die aan het onderzoek in 2008 hebben meegedaan. Dit is nagegaan voor de kenmerken leeftijd ($t(133) = -.125$, $p=.90$), sexe ($\chi^2(136) = 0{,}00$,

[1] De beide dimensies worden als volgt berekend (Mainhard, 2009): Invloed= (0.92*BS) + (0.38*SB) - (0.38*SO) - (0.92*OS) - (0.92*OT) - (0.38*TO) +(0.38*TB) + (0.92*BT); Nabijheid= (0.38*BS) + (0.92*SB) + (0.92*SO) + (0.38*OS)- (0.38*OT) - (0.92*TO) - (0.92*TB) - (0.38*BT).

p=.583), aantal jaren ervaring op deze school (χ^2(130)= 3,63, p= .606) en aantal jaren ervaring in het onderwijs (χ^2(130)= 6.43, p= .266) waarbij de laatste twee kenmerken waren geclassificeerd in zes klassen. De genoemde 21 docenten zijn op deze kenmerken dus vergelijkbaar met de overige docenten in de groep van 141 docenten.
De 197 leerlingen verschillen op de beschikbare kenmerken significant van de overige 703 leerlingen van 2008. Dit betreft de kenmerken sexe (χ^2(746) = 6.97, p<.001), leeftijd (t(806) = –0.14, p=.002) en schooltype (χ^2(804) = 56,565, p<.001). De groep van 197 bevat meer meisjes (68% tegenover 57% bij de overige leerlingen), minder 12- en 13-jarigen (23% versus 32%) en meer vmbo-havo-leerlingen (75% versus 55%).
Behalve op persoonskenmerken is de selectie ook gecontroleerd op de PDH-scores. Op drie van de 20 taken verschillen de 21 docenten significant (p<.05) van de overige 120. Dit betreft taak 4 (gemiddelden 4,1 en 3,6), taak 6 (3,8 en 3,6) en taak 11 (3,9 en 3,5). Op tien van de 20 taken is sprake van een significant verschil (p<.05) tussen de 197 leerlingen en de overige 703 leerlingen. Dit zijn de taken 4, 6 t/m 10, 13 en 16 t/m 18. De geselecteerde groep scoort hierbij meestal 0,2 schaalpunt hoger, behalve bij de taken 8 en 9 waarbij het verschil 0,3 schaalpunt is (3,7 om 3,4 en 3,9 om 3,6).

PDH-data van 2009-2011
Op vier scholen zijn nieuwe data verkregen van 28 docenten en 308 leerlingen (6 tot 14 leerlingen per docent, gemiddeld 11). De docenten waren 11 mannen en 17 vrouwen in de leeftijd van 22 tot 60 jaar (gemiddeld 36 jaar), met gemiddeld 4 jaar ervaring op de school en in totaal gemiddeld 6 jaar onderwijservaring. Van de leerlingen heeft 17% geen geslacht ingevuld, 34% betrof jongens en 49% meisjes. Ze waren 11 tot 18 jaar, en zaten in de leerjaren 1, 3 en 4, gespreid over de schooltypen van vmbo-basis tot en met vwo (c.q. gymnasium).

PDH-data samengevoegd
Hoewel de data van 2008 betrekking hebben op items waarvan een gedeelte in de verkorte versie qua formulering is aangepast, ging het daarbij slechts om een relatief beperkt aantal items, verspreid over de taken, en waren de aanpassingen niet ingrijpend. Dit in aanmerking genomen en mede gelet op de hoge interne consistenties, kan het verantwoord worden geacht om deze data te combineren met de nieuw verkregen data van 2009-2011.
Samengevoegd waren PDH-data beschikbaar van 49 docenten en 505 leerlingen. De docenten betreffen 20 mannen en 29 vrouwen, werkzaam op 6 verschillende scholen, variërend in de leeftijd van 22 tot 60 jaar (gemiddeld 37 jaar), met gemiddeld bijna 4 jaar ervaring op de betreffende school en gemiddeld bijna 9 jaar onderwijservaring in het totaal. Van de leerlingen zijn 33% jongens en 54% meisjes en van 13% is het geslacht niet bekend. Ze variëren in leeftijd van 11 tot 18 jaar en zitten gespreid over de leerjaren 1 tot en met 6 en de schooltypen vmbo-basis tot en met vwo/gymnasium.

VIL

De VIL is ingevuld door 17 docenten en 188 leerlingen (6 tot 16 leerlingen per docent, gemiddeld 11). De 17 docenten waren 10 mannen en 7 vrouwen in de leeftijd van 22 tot 60 jaar (gemiddeld 38 jaar), met 1 tot 19 jaar ervaring op de huidige school (gemiddeld 4 jaar) en 1 tot 24 jaar onderwijservaring in totaal (gemiddeld 8). De leerlingen variëren in leeftijd van 11 tot 17 jaar en zijn gespreid over de leerjaren 1 (15%), 3 (40%) en 4 (45%) en de schooltypen vmbo (5%), vmbo-t (32%), vmbo-t/havo (5%), havo (24% en havo/vwo (34%). In totaal heeft 76% van de leerlingen aangegeven of ze een jongen (33%) zijn of een meisje (43%).

Analyses

Controle op representativiteit van de bij deze studie betrokken data uit hoofdstuk 3

In dit hoofdstuk zijn de data betrokken van 21 docenten waarbij over het handelen van de docent niet alleen PDH-scores van de docent zelf maar ook van ten minste 8 leerlingen beschikbaar zijn gekomen, en de data van 197 leerlingen. Ter controle op selectiebias is nagegaan in hoeverre deze 21 docenten en 197 leerlingen op hun persoonskenmerken en PDH-scores verschillen van de overige 120 docenten respectievelijk 703 leerlingen die aan het onderzoek in hoofdstuk 3 hebben meegedaan. Hierbij is gebruik gemaakt van T-toetsen, χ^2 -toetsen en enkelvoudige variantie-analyses. De uitkomsten zijn hiervoor vermeld in 4.3.2.

Schaalbetrouwbaarheid en groepering van taken

De betrouwbaarheid van de schalen voor de PDH-taken is bij zowel de leerlingen als de docenten nagegaan met behulp van de itemrestcorrelaties (criterium: minimaal .20, liefst $\geq$ .35) en Cronbach's alpha (criterium: minimaal .60, liefst $\geq$ .70). Indien een schaal bij een van beide groepen onvoldoende betrouwbaar bleek (dit kwam alleen voor bij de docenten) zijn een of twee items verwijderd als die inhoudelijk niet essentieel werden geacht voor de betreffende taak en daardoor wel een betrouwbare schaal werd verkregen. Ter wille van de vergelijkbaarheid met de leerlingen is deze verwijdering ook daarbij doorgevoerd en is de schaalbaarheid in die groep opnieuw gecontroleerd.

Tevens is nagegaan of alle taken samen een betrouwbare overallschaal vormen waarmee het gehele repertoire van het pedagogisch-didactisch handelen in kaart kan worden gebracht.

In een onderzoek dat verder niet is opgenomen in dit proefschrift (Sol & Stokking, 2010) is aan 136 docenten een lijst voorgelegd van alle 22 taken (dus met als items de omschrijvingen van de taken) met daarbij de vraag om aan te geven hoe belangrijk zij deze taken achten voor hun werk. De respondenten waren 80 mannen en 56 vrouwen in leeftijd variërend van 22 tot 65 jaar (gemiddeld 44 jaar) en met 1 tot 40 jaar werkervaring (gemiddeld 17 jaar). Zij waren werkzaam op vijf scholen, variërend van vmbo tot gymnasium (andere scholen dan in onderhavig onderzoek zijn betrokken).

Factoranalyse op de set van 22 taken leidde tot vier goed interpreteerbare groepen van taken, met drie tot tien taken per groep, die blijkens schaalanalyses ook voldoende betrouwbaar bleken. In onderhavige studie is opnieuw nagegaan of deze groepen van taken betrouwbare schalen vormen. Dit is ook inhoudelijk relevant, omdat daarmee aanvullend inzicht wordt verkregen in de inhoud en betekenis van het domein dat door de PDH wordt bestreken. Daarnaast zijn deze betekenisvolle groepen van taken binnen deze studie ook gebruikt als criteriumvariabelen in enkele andere analyses, in aanvulling op de overallschaal voor het gehele repertoire.

Controle op uitbijters
Met behulp van boxplots is nagegaan in hoeverre in de data van de docenten respectievelijk de leerlingen sprake is van uitbijters in de schaalscores van de PDH-taken. Extreme uitbijters zijn verwijderd.

Beschrijvende analyses
Per groep (docenten, leerlingen) per taak zijn het gemiddelde en de standaardafwijking berekend. Daarnaast zijn van de persoonskenmerken percentages respectievelijk gemiddelden en standaarddeviaties berekend. Ook zijn per taak de gemiddelde verschilscores tussen de docenten en hun leerlingen berekend.

Overeenstemming tussen leerlingen per docent en tussen leerlingen en docent
Op de data van de leerlingen is per docent over de 20 taken de interbeoordelaars-betrouwbaarheid berekend, met de jury-alpha. Deze geeft aan in hoeverre de scores van de leerlingen die de lijst hebben ingevuld voor dezelfde docent onderling overeenstemmen. Aanvullend is per docent de correlatie berekend tussen het profiel (de reeks scores op de 20 taken) van het huidig handelen volgens de docent zelf en het profiel volgens de gemiddelde scores van zijn/haar leerlingen.

Onderscheidingsvermogen
In de data van de leerlingen is per taak van de PDH nagegaan in hoeverre daarmee onderscheid kan worden gemaakt tussen docenten, door berekening van de intraclassecorrelatie, de ICC(1), aangevuld met een berekening van de betrouwbaarheid van de op docentniveau geaggregeerde leerlingscores ICC(2) (Lűdke et al., 2009). Volgens Mainhard (2009) is een ICC(1) van .25 in onderwijsonderzoek te beschouwen als hoog en volgens Lüdke et al. komt een ICC(1) van > .30 daarbij zelden voor. Een ICC(2) score van ≥ .85 wordt gezien als indicatief voor een goede betrouwbaarheid (Mainhard, 2009; Lűdke et al., 2009).
Om na te gaan in hoeverre de 49 docenten in deze studie op basis van de PDH-data kunnen worden gegroepeerd in een beperkt aantal clusters, is daarnaast een hiërarchische clusteranalyse uitgevoerd (volgens de *average linkage* methode) op de data van de docenten en van de leerlingen. Ook dit kan inzicht bieden in het onderscheidingsvermogen van de PDH, en de verkregen clusters kunnen worden

getypeerd aan de hand van kenmerken van docenten en/of hun leerlingen. Bij sterk ongelijke aantallen docenten per cluster is bij het toetsen van verschillen tussen clusters de Kruskal-Wallis toets gebruikt.

Discriminante validiteit
De relaties tussen de PDH en de VIL zijn onderzocht met correlaties en multiple regressie-analyses, zowel op het niveau van de afzonderlijke schalen (taken c.q. sectoren) als op dat van de overallschaal van PDH en de twee dimensies van de VIL. Voorafgaand aan de multiple regressie-analyses zijn de daarbij relevante assumpties gecontroleeerd.

Verschillen tussen groepen en confirmerende validiteit
Verschillen en verbanden zijn op de data van de leerlingen en van de docenten onderzocht met behulp van t-toetsen en enkelvoudige variantieanalyses en op de data van de docenten en leerlingen gezamenlijk met behulp van multilevelanalyses op drie niveaus: school, docent, leerling.

Verschillen tussen huidig en gewenst handelen
Verschillen tussen huidig en gewenst handelen bij leerlingen en bij docenten zijn onderzocht met behulp van t-toetsen.

Bij alle toetsingen op statistische significantie is een significantieniveau aangehouden van .05.

4.4 Resultaten

4.4.1 Indicaties voor de betrouwbaarheid

De betrouwbaarheden van de schalen van de PDH bij docenten en leerlingen
De betrouwbaarheid van de schalen bij de leerlingen is goed tot zeer goed (alle Cronbach's alpha's $\geq .70$ en itemrestcorrelaties $\geq .35$). Bij de docenten bleek dit bij een aantal schalen niet het geval en zijn daarbij één item (bij 12 taken) of twee items (bij twee taken) verwijderd. De betreffende schalen bij de leerlingen zijn overeenkomstig aangepast. De Cronbach's alpha`s van de resulterende schalen bij de docenten variëren tussen .60 en .94. Bij de leerlingen is de Cronbach's alpha alleen bij taak 2 net onder de .70. De gegevens in tabel 4.1 betreffen 502 leerlingen, na verwijdering van drie uitbijters (zie hierna). De gemiddelde scores per taak bij de docenten variëren van 2,9 (taak 5) tot 4,4 (taak 20) en bij de leerlingen van 3,0 (taak 5) tot 3,9 (taken 1, 3, 15, 17). Op individueel niveau bestrijkt de range bij de meeste taken vrijwel de hele schaal, en op itemniveau worden steeds alle scores van 1 t/m 5 benut.

Controle op uitbijters op de schalen

De scores op de gemaakte schalen (per PDH-taak) zijn gecontroleerd op uitbijters. Bij de docenten is bij negen taken sprake van een of meer uitbijters (per taak variërend van 2 tot 6%, met een uitschieter tot 10% bij taak 12). Deze uitbijters bleken niet extreem te zijn (minder dan 3 maal de interkwartielafstand) en betreffen ook steeds wisselende docenten. Daarom is besloten alle gegevens van de docenten bij de verdere analyses te betrekken. Bij de uitkomsten van de leerlingen bleek bij in totaal 10 taken sprake van uitbijters (per taak variërend van 2-3% van het totaal aantal leerlingen), maar ook deze bleken niet extreem te zijn en ook steeds wisselende leerlingen te betreffen. Apart is ook nog gecontroleerd op eventuele extreme uitbijters bij de leerlingen per docent. Dit was het geval bij docent 14 en bij docent 15 (respectievelijk een leerling en twee leerlingen). De gegevens van deze leerlingen zijn alsnog verwijderd, waardoor het totale aantal leerlingen is teruggebracht tot 502.

Tabel 4.1: Schalen en beschrijvende uitkomsten PDH bij 49 docenten en 502 leerlingen

Taak	Aantal items	Docenten				Leerlingen			
		α	r_{ir}	*M*	*SD*	*α*	r_{ir}	*M*	*SD*
1	6	.80	≥.42	4.1	.543	.82	≥.54	3.9	.735
2	3	.69^	≥.48	3.7	.698	.67^	≥.39	3.7	.787
3	5	.69^	≥.34*	3.8	.554	.81^	≥.44	3.9	.731
4	3	.60^	≥.33*	3.7	.724	.76^	≥.59	3.3	1.044
5	3	.78^	≥.57	2.9	.764	.85^	≥.67	3.0	1.109
6	6	.71	≥.33*	3.6	.606	.87	≥.62	3.6	.879
7	4	.75^	≥.47	3.7	.640	.83^	≥.63	3.6	.935
8	4	.77^	≥.43	3.2	.873	.85^	≥.64	3.4	1.040
9	4	.65	≥.33	4.1	.585	.83	≥.61	3.7	.999
10	3	.63^	≥.40	3.6	.786	.73^	≥.51	3.3	1.041
11	4	.77	≥.50	3.6	.872	.82	≥.47	3.6	.984
12	3	.64^	≥.30*	3.6	.757	.72^	≥.44	3.8	.878
13	4	.86^	≥.60	3.2	.896	.88^	≥.74	3.3	1.055
14	5	.87	≥.69	3.4	.762	.86	≥.53	3.5	.938
15	4	.62^	≥.34*	3.8	.622	.78^	≥.56	3.9	.790
16	4	.75^	≥.63	4.2	.574	.86^	≥.62	3.8	.955
17	6	.85	≥.44	4.0	.584	.91	≥.71	3.9	.888
18	3	.65^	≥.38	3.0	.891	.85^	≥.70	3.3	1.125
19	5	.65^	≥.31*	3.5	.581	.86^	≥.59	3.6	.942
20	4	.80	≥.65	4.4	.557	.85	≥.65	3.8	.959
Overall (20 taken)		.94	≥.34	3.7	.473	.96	≥.56	3.6	.719

^: op basis van lage itemrestcorrelaties bij de docenten zijn een of twee items verwijderd.

* bij $r_{ir} \leq .35$ betrof het steeds slechts één item in de schaal; bij alle andere items was de $r_{ir} \geq .35$.

Overallschaal en indeling in groepen taken

Vanwege het grote aantal taken wordt in een aantal verdere analyses gewerkt met de overallschaal als criteriumvariabele. Aangezien uit de in hoofdstuk 3 gerapporteerde factoranalyse op de data van de leerlingen bleek dat sprake is van één dominerende (eerste) factor, representeert deze overallschaal een betekenisvolle maat voor het repertoire als geheel. Toch wilden we graag naast de overallschaal kunnen beschikken over enkele criteriumvariabelen die betekenisvolle groepen van taken zouden vertegenwoordigen. Daartoe maken we gebruik van de indeling van taken in vier groepen uit een eerdergenoemd onderzoek dat heeft plaatsgevonden in 2010 (zie Stokking et al., 2010).

Een factoranalyse op de data van 136 docenten leidde tot drie factoren (eigenwaarden: 6,7, 2,4 en 1,3, cumulatief percentage verklaarde variantie 52%). De factorladingen zijn opgenomen in tabel 4.2, waarbij ladingen ≥ .45 vet zijn gezet. Gelet op de verhouding tussen het aantal variabelen en het aantal cases moet de interpretatie van deze uitkomsten worden beschouwd als tentatief.

De eerste factor omvat taken die te maken hebben met het bijdragen aan de ontwikkeling van de leerlingen (zoals: persoonlijke vorming, samenwerken, maken van keuzen, zelfstandig leren). De tweede factor omvat taken die zijn gericht op het bevorderen van het (vakinhoudelijke) leerproces (zoals: zorgen voor een goed leerklimaat, geven van les en instructie, laten maken van opdrachten). De derde factor lijkt taken te omvatten die te maken hebben met de beoordeling en verbetering en het voldoen aan bepaalde normen: bij de taken 17-19 in relatie tot de leerlingen (feedback en beoordeling; de factorlading van taak 19 ligt net onder de grens van .45), bij de taken 20-22 in relatie tot de docenten (samenwerken, leren, professionele ontwikkeling). Dat taak 21 ook laadt op de eerste factor is niet vreemd (veel taken die op de eerste factor laden hebben een begeleidingscomponent; zie ook kolom C in tabel 3.2). Het laden van taak 12 op de derde factor is minder gemakkelijk te duiden.

Omdat het instrument primair is gericht op het handelen van docenten in interactie met leerlingen en het domein van feedback en beoordeling (formatieve en summatieve assessment) in de literatuur over het handelen van docenten een eigen plaats inneemt (zie de hoofdstukken 3, 5 en 6), ligt het voor de hand om de taken 17-19 en de taken 20-22 van elkaar te onderscheiden. Een bijkomend empirisch argument hiervoor is dat extractie van een extra, vierde factor, met een eigenwaarde van net boven de 1, ertoe leidt dat de taken 17-19 een eigen factor gaan vormen en de factorlading van taak 12 op de derde factor daalt (tot .40) en die op de eerste factor stijgt. Beide sets van drie taken vormen in de data van het betreffende onderzoek bovendien een schaal met een acceptabele betrouwbaarheid. De betrouwbaarheden (Cronbach's alpha) van de genoemde schalen in dit onderzoek zijn:

Groep 1: tien taken die zijn gericht op de *ontwikkeling* van leerlingen: .90.

Groep 2: zes taken gericht op het (vakinhoudelijke) *leerproces* van leerlingen: .81.

Groep 3: drie taken over geven van *feedback en beoordelen*: .69.

Groep 4: drie taken over *samenwerking en professionele ontwikkeling*: .69.

Ter controle is nagegaan of de eerste drie groepen taken ook in onderhavige studie bij 49 docenten en 502 leerlingen betrouwbare schalen vormen. Dit blijkt zo te zijn, met voor de takengroepen 1, 2 en 3 Cronbach's alpha's bij de leerlingen van respectievelijk .92 , .89 en .84 en bij de docenten van .87, .86 en .74. De vierde takengroep (de taken 20-22) is in deze studie alleen met taak 20 vertegenwoordigd.
De gemiddelden van de respondenten op deze groepen taken variëren sterk: bij de docenten varieert de range van 2,1 bij groep 2 (2,6-4,7) tot 2,8 bij groep 3 (2,0-4,8), bij de leerlingen varieert de range van 3,7 bij groep 2 tot 4,0 bij de groepen 3 en 4.

*Tabel 4.2: Uitkomsten factoranalyse op data van 136 docenten over 22 taken (Stokking & Sol, 2010)***

	Taken PDH	Factorladingen bij 136 docenten			Indeling van de taken in vier groepen			
		Factor 1	*Factor 2*	*Factor 3*	*1*	*2*	*3*	*4 **
1	Zorgen voor een goed leerklimaat	.36	**.51**	-.03		X		
2	Bijdragen aan de persoonlijke vorming	**.72**	.14	-.10	X			
3	Geven van les, uitleg en instructie	-.13	**.75**	.23		X		
4	Leerlingen het nut van de leerstof laten zien	.05	**.55**	.32		X		
5	Leerlingen wijzen op verbanden tussen vakken	.18	**.47**	.24		X		
6	Leerlingen leren werken aan opdrachten	.13	**.67**	.08		X		
7	Leerlingen stimuleren en activeren	.38	**.63**	.12		X		
8	Leerlingen vaardigheden aanleren	**.70**	.16	-.01	X			
9	Inspelen op verschillen tussen leerlingen	**.58**	.28	.15	X			
10	Inspelen op verschillen door middel van opdrachten	**.62**	.29	.23	X			
11	Lln. laten samenwerken en daarbij ondersteunen	**.51**	.26	.44	X			
12	Leerlingen eigen keuzes laten maken	**.61**	.00	**.51**	X			
13	Lln. ondersteunen bij het zelf maken van keuzes	**.67**	-.02	.44	X			
14	Leerlingen hun eigen leerproces laten sturen	**.74**	.04	.27	X			
15	Lln. ondersteunen bij het sturen van hun leerproces	**.80**	.09	.24	X			
16	Volgen wat lln. leren en hoe zij zich ontwikkelen	**.63**	.09	.28	X			
17	Geven van feedback aan leerlingen	.19	.36	**.64**			X	
18	Bevorderen dat feedback wordt gebruikt	.25	.34	**.59**			X	
19	Beoordelen van het werk/ prestaties van leerlingen	-.11	.42	.44			X	
20	Open staan voor leren en samenwerken	.19	.17	**.66**				X
21	Samen met andere leraren een groep begeleiden	**.48**	-.15	**.50**				X
22	Professionele ontwikkeling	.18	.25	**.57**				X

** Factorladingen ≥ .45 zijn vet gezet.

* In het genoemde onderzoek in 2010 waren alle 22 taken van de PDH betrokken. De onderhavige studie betreft alleen de taken 1-20, en daarbij is groep 4 dus alleen vertegenwoordigd door taak 20.

De betrouwbaarheid van de schalen van de PDH bij de groep van 197 leerlingen (bij 17 docenten)

Door 17 docenten en 197 leerlingen is naast de PDH ook de VIL ingevuld. Omdat de PDH en VIL worden vergeleken (zie 4.4.4) is apart bij deze 197 leerlingen de betrouwbaarheid van de PDH-schalen gecontroleerd (het aantal docenten was daarvoor te beperkt). Alle schalen bleken ruim voldoende tot (zeer) goed betrouwbaar, zie tabel 4.3. (Zie voor de betrouwbaarheid van de VIL par. 4.4.3).

Tabel 4.3: Schalen en beschrijvende uitkomsten PDH, 197 leerlingen (17 docenten)

Taak	*k*	*α*	r_{ir}	*M*	*SD*	*Taak*	*k*	*α*	r_{ir}	*M*	*SD*
1	6	.78	≥.44	3.9	.672	11	4	.81	≥.44	3.6	.986
2	5	.80	≥.46	3.6	.776	12	5	.80	≥.56	3.9	.812
3	6	.79	≥.47	4.1	.669	13	5	.88	≥.59	3.1	1.034
4	4	.77	≥.65	3.3	.968	14	5	.84	≥.45	3.4	.946
5	4	.85	≥.65	2.9	1.047	15	5	.80	≥.46	3.9	.967
6	6	.84	≥.56	3.6	.856	16	5	.88	≥.54	3.7	.997
7	5	.85	≥.62	3.7	.930	17	6	.89	≥.66	3.9	.881
8	5	.84	≥.79	3.2	1.023	18	4	.85	≥.65	3.0	1.067
9	4	.81	≥.46	3.7	.987	19	6	.85	≥.52	3.6	.912
10	4	.76	≥.64	3.2	1.017	20	5	.86	≥.64	3.7	.938
Overall (schaal die kan worden gemaakt op basis van de 20 taken)								.96	≥.55	3.6	.690

Overeenstemming tussen de leerlingen per docent

Per docent is de interbeoordelaarsbetrouwbaarheid van de leerlingen bij die docent berekend (de jury-alpha), zie tabel 4.4. De jury alpha is bij 23 docenten (47%) ≥ .70 en bij 35 docenten (72%) ≥ .60. Bij 11 docenten (22%) is er weinig overeenstemming tussen de scores van de betreffende leerlingen (jury alpha < .50).

Overeenstemming tussen de docenten en hun leerlingen, gemiddeld en per docent

Het verschil tussen de gemiddelde schaalscore van de docenten en de gemiddelde schaalscore van hun leerlingen (zie tabel 4.1) is bij zes van de 20 taken significant: taak 1 (t(48) = 2,613, p=.012), taak 4 (t(48) = 4,086, p<.001), taak 9 (t(48) = 4,604, p<.001), taak 16 (t(48) = 4,779, p<.001), taak 18 (t(48) = -2,497, p=.016) en taak 20 (t(48) = 6,357, p ≤.001). De gemiddelde verschilscore *per docent* tussen de docent en zijn leerlingen varieert tussen de docenten per taak, van −0,2 tot 0,6 (gemiddeld 0,4), waarbij een negatieve verschilscore betekent dat het gemiddelde van de docent hoger is dan het gemiddelde van de leerlingen. De verschilscore op de overallschaal (over alle 20 taken) varieert tussen de docenten van −1,5 tot 1,2 met een gemiddelde van 0,06 en een standaarddeviatie van .525.

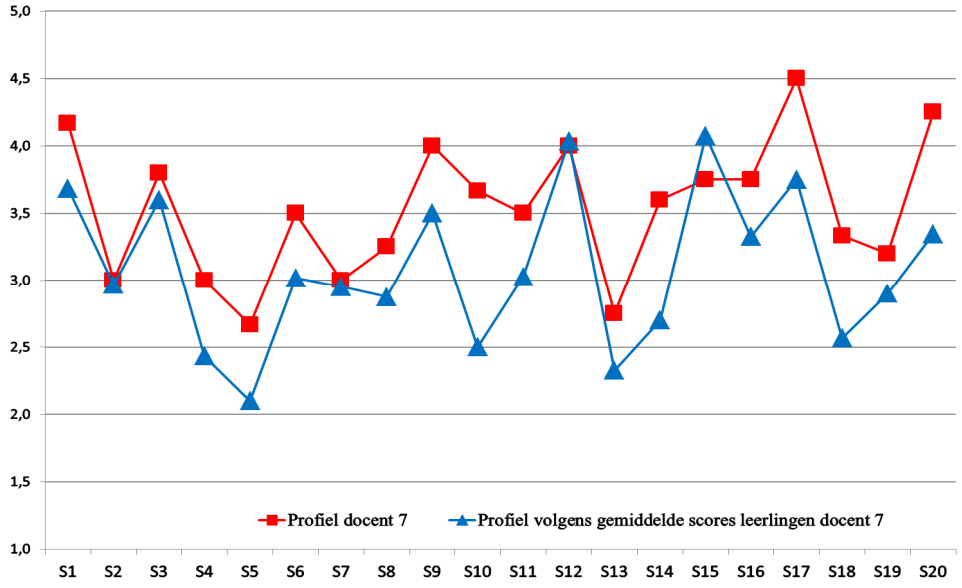
Tabel 4.4 bevat de correlatie tussen het docentprofiel en het gemiddelde leerlingen-profiel per docent. Deze correlatie varieert sterk tussen de docenten. Bij 31 docenten (63%) is deze significant positief bij p≤ .05 en bij 34 docenten (69%) bij p ≤ .10. Bij 26 docenten is de correlatie sterk (≥ .50, tot .79), bij 5 docenten matig tot sterk (.40 tot .50), bij 6 docenten matig (.30 tot .40) en bij 9 docenten zwak tot matig (< .30). Bij 3 docenten is de correlatie niet significant negatief. Gemiddeld is de correlatie .46.

Tabel 4.4: Interbeoordelaarsbetrouwbaarheid (jury alpha) bij de leerlingen per docent en correlaties tussen het profiel volgens de docent en dat volgens zijn/haar leerlingen

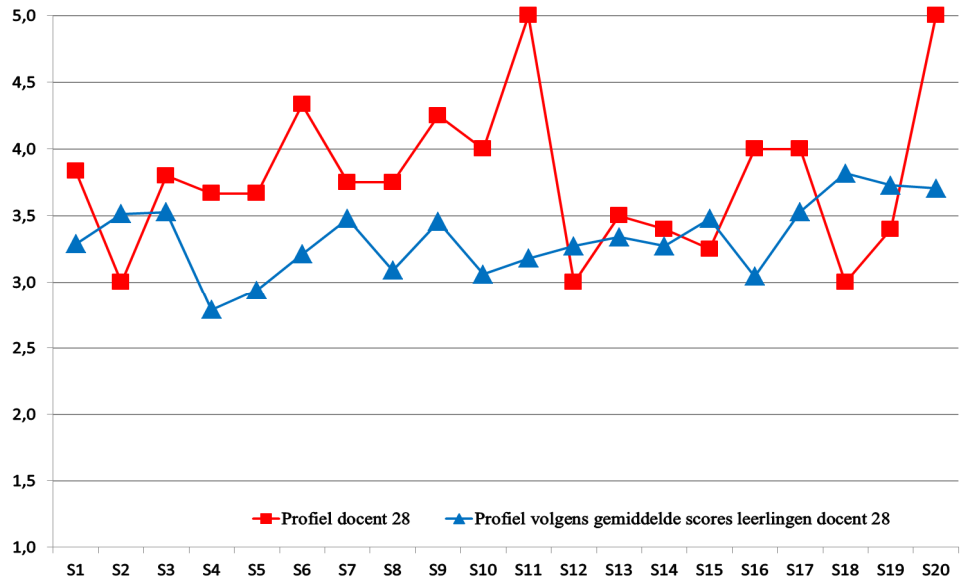
Docent nummer	*Aantal leerlingen*	*Jury-Alpha*	*Correlatie (p-waarde)*	*Docent nummer*	*Aantal leerlingen*	*Jury-Alpha*	*Correlatie (p-waarde)*
1	8	.78	.14 (.56)	26	12	.69	.65 (.00)*
2	10	.50	.44 (.05)*	27	10	.75	.52 (.02)*
3	9	.28	.56(.01)*	28	11	.63	-.13 (.58)
4	14	.87	.66 (.00)*	29	9	.77	.22 (.35)
5	14	.92	.64 (.00)*	30	10	.69	-.09 (.72)
6	10	.56	.68 (.00)*	31	12	.82	.66 (.00)*
7	10	.88	.79 (.00)*	32	12	.84	.70 (.00)*
8	12	.15	.17 (.46)	33	12	.60	.56 (.01)*
9	11	.24	.16 (.50)	34	10	.84	.68 (.00)*
10	11	.67	.26 (.26)	35	13	.69	.05 (.85)
11	12	.46	.39 (.09)	36	8	.77	.37 (.11)
12	11	.86	.47 (.04)*	37	8	.88	.58(.01)*
13	8	.62	.60 (.01)*	38	8	.86	.57 (.01)*
14	8	.43	-.00 (.98)	39	8	.79	.75 (.00)*
15	10	.20	.62 (.00)*	40	8	.93	.70 (.00)*
16	15	.67	.69 (.00)*	41	7	.60	.66 (.00)*
17	14	.71	.29 (.22)	42	7	.73	.53 (.02)*
18	11	.32	.55 (.01)*	43	6	.65	.27 (.25)
19	10	.72	.38 (.10)	44	6	.25	.57 (.01)*
20	14	.86	.57 (.01)*	45	13	.65	.49 (.03)*
21	10	.64	.39 (.09)	46	6	.83	.69 (.00)*
22	13	.87	.75 (.00)*	47	7	.20	.46 (.04)*
23	11	.04	.33 (.15)	48	9	.59	.45 (.05)*
24	16	.89	.37 (.11)	49	13	.85	.69 (.00)*
25	10	.49	.08 (.72)				

*: significant bij $p \leq 0.05$

Van twee docenten zijn het docentprofiel en het gemiddelde leerlingprofiel visueel weergegeven: in figuur 4.1 van docent 7, met de hoogste correlatie tussen beide profielen, en in figuur 4.2 van docent 28, met de laagste correlatie tussen beide profielen.



Figuur 4.1: Profiel huidig handelen docent 7 volgens docent en leerlingen (hoogste correlatie)



Figuur 4.2: Profiel huidig handelen docent 28 volgens docent en leerlingen (laagste correlatie)

4.4.2 Indicaties voor het onderscheidingsvermogen

Intraclassecorrelaties

Tabel 4.5 bevat de waarden van de intraclassecorrelatie ICC(1) en de betrouwbaarheid van de op docentniveau geaggregeerde leerlingscores ICC(2) berekend op de data van de leerlingen (n=502). Bij 13 taken is de ICC(1) ≥ .25. bij de overige taken ≥ .20, behalve bij taak 10 (.14) De ICC(2) varieert tussen .79 en .93 met uitzondering van taak 10 (.69). Bij de overallschaal is de ICC(1) .35 en de ICC(2) .93. Dit betekent dat in deze groep leerlingen 35% van de variantie in de scores samenhangt met het feit dat de leerlingen per groep de vragenlijst hebben ingevuld voor één bepaalde docent. Bij de vier takengroepen zijn de intraclassecorrelaties .36, .32, .27 en .31 en de ICC(2)'s rond de .90.

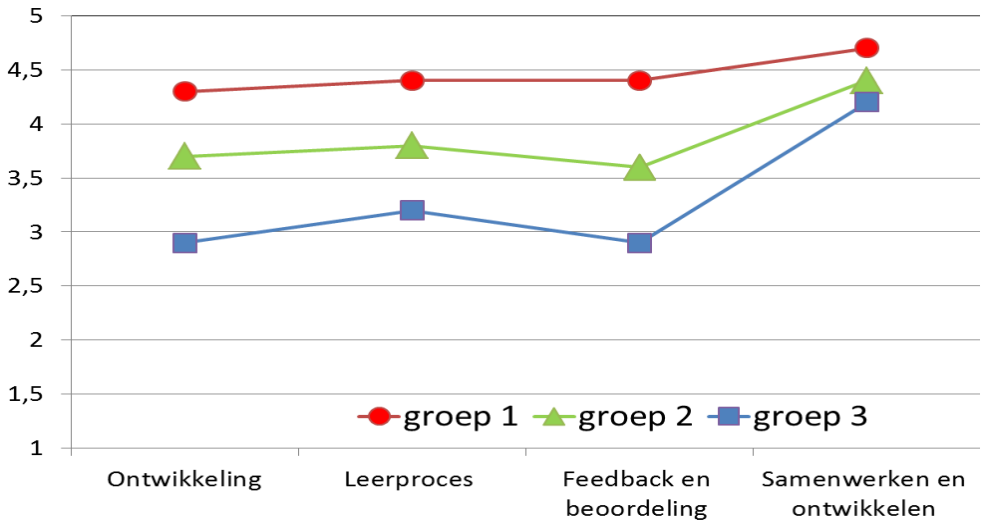
Tabel 4.5: Intraclassecorrelatie (ICC(1)) en betrouwbaarheid (ICC(2)) per taak (over 49 docenten)

	Taken PDH	*ICC(1)*	*ICC(2)*
1	Zorgen voor een goed leerklimaat	.36	.93
2	Bijdragen aan de persoonlijke vorming van leerlingen	.27	.87
3	Geven van les, uitleg en instructie	.26	.86
4	Leerlingen het nut van de leerstof laten zien	.25	.85
5	Leerlingen wijzen op verbanden tussen vakken	.21	.80
6	Leerlingen leren werken aan opdrachten	.30	.89
7	Leerlingen stimuleren en activeren	.31	.90
8	Leerlingen vaardigheden aanleren	.25	.84
9	Inspelen op verschillen tussen leerlingen	.27	.86
10	Inspelen op verschillen door middel van opdrachten	.14	.69
11	Leerlingen laten samenwerken en daarbij ondersteunen	.24	.83
12	Leerlingen eigen keuzes laten maken	.25	.84
13	Leerlingen ondersteunen bij het zelf maken van keuzes	.24	.84
14	Leerlingen hun eigen leerproces laten sturen	.27	.87
15	Leerlingen ondersteunen bij het sturen van hun leerproces	.20	.79
16	Volgen wat leerlingen leren en hoe zij zich ontwikkelen	.27	.87
17	Geven van feedback aan leerlingen	.24	.84
18	Bevorderen dat feedback wordt gebruikt	.23	.82
19	Beoordelen van het werk en de prestaties van leerlingen	.22	.81
20	Open staan voor leren en samenwerken	.31	.90
PDH	Overall (gehele repertoire, schaal gebaseerd op 20 taken)	.35	.93

Onderscheiden van clusters van docenten

Om na te gaan in hoeverre de 49 docenten kunnen worden geclusterd in een beperkt aantal groepen, zijn twee hiërarchische clusteranalyses uitgevoerd: op de scores op de PDH-taken van de docenten zelf en op de gemiddelde scores per docent van hun leerlingen. Gestreefd werd naar een indeling in drie tot vier groepen en het bij een groep ingedeeld zijn van alle 49 docenten, visueel na te gaan met behulp van een dendogram. Bij de data van de leerlingen bleek aan deze criteria te worden voldaan met een indeling in vier groepen bij een rescaled distance van 9 en op de data van de docenten met een indeling in drie groepen bij een rescaled distance van 15 (dus met

een relatief grotere afstand tussen de groepen dan bij de leerlingen). De vier groepen op basis van de data van de leerlingen hebben een omvang van 30, 8, 8 respectievelijk 3 docenten en verschillen weliswaar significant maar in absolute zin slechts weinig in gemiddelde op de overallschaal (respectievelijk 3,5, 3,6, 3,8 en 3,9). De drie groepen op basis van de data van de docenten hebben een omvang van 8, 31 en 10 docenten en verschillen significant en duidelijk in gemiddelde op de overallschaal: (4,4 (groep 1), 3,7 (groep 2) en 3,0 (groep 3)), en ook op de eerste drie onderscheiden groepen van taken en op alle afzonderlijke taken (met uitzondering van taak 20). De gemiddelde scores van deze drie groepen docenten op de vier groepen taken zijn weergegeven in figuur 4.3. De indeling in drie groepen hangt niet samen met de school waar de docenten werken en ook niet met hun geslacht, leeftijd en ervaring.



Figuur 4.3: Verschillen tussen drie groepen docenten op groepen van taken (data docenten)

4.4.3 Indicaties voor discriminante validiteit

Betrouwbaarheid en beschrijvende gegevens VIL

Ter vergelijking met de PDH is door 17 docenten en 188 leerlingen (55 vmbo en 133 havo/vwo) ook de VIL ingevuld. Eerst is de betrouwbaarheid van de VIL-schalen gecontroleerd, en gelet op het kleine aantal docenten is dit alleen gedaan bij de leerlingen. De uitkomsten staan in tabel 4.6. De betrouwbaarheid van de schaal OS is bij beide groepen leerlingen, vmbo en havo/vwo, lager dan wenselijk. Besloten is deze schaal toch te benutten, om ook scores op de twee dimensies te kunnen bepalen. Tabel 4.6 laat zien dat de leerlingen hun docenten gemiddeld het hoogste scoren op de eerste vier sectoren en dat zij hun docenten hoger scoren op de dimensie Nabijheid dan op de dimensie Invloed. Het verschil tussen de scores voor Nabijheid en Invloed is

significant voor beide groepen (vmbo-leerlingen: t(54)= 12.561, p≤.001 en havo/vwo-leerlingen: t(132)= 15.208, p≤.001). De gemiddelden op de schalen en die op beide dimensies verschillen tussen beide groepen leerlingen significant van elkaar (p ≤.001), behalve bij BS en TB.

Tabel 4.6: Schalen en beschrijvende uitkomsten VIL, leerlingen vmbo en havo/vwo

Schaal	Aantal items	Leerlingen vmbo (n=55)				Aantal items	Leerlingen havo/vwo (n=133)			
		α	r_{ir}	*M*	*SD*		*α*	r_{ir}	*M*	*SD*
BS	4	.71	≥.07	4.0	.502	9	.85	≥.19*	3.8	.650
SB	3	.84	≥.66	4.4	.635	9	.77	≥.29*	3.6	.557
SO	3	.78	≥.48	4.4	.632	9	.85	≥.36	3.7	.633
OS	3	.57^	≥.31	3.1	.603	4	.55^	≥.13	3.6	.550
OT	6	.73	≥.30	1.8	.578	8	.75	≥.17	2.5	.592
TO	6	.74	≥.19	1.7	.600	6	.83	≥.33*	1.9	.578
TB	5	.67	≥.11	2.3	.669	4	.62	≥.06	2.5	.546
BT	3	.63^	≥.30	2.4	.506	8	.60	≥.18	2.7	.469
Dimensie Invloed				1,5	1.142	Dimensie Invloed			0.5	1.342
Dimensie Nabijheid				5,4	2.356	Dimensie Nabijheid			3.5	2.161

^ één of meer items uit de schaal verwijderd vanwege te lage betrouwbaarheid en/of negatieve itemrestcorrelatie

* bij $r_{ir} \leq .35$ betrof het steeds slechts één item in de schaal; bij alle andere items was de $r_{ir} \geq .35$.

Relaties tussen PDH en VIL: uitkomsten van correlaties en multiple regressieanalyses

Vrijwel alle taken van de PDH en de PDH-overallschaal hangen significant positief samen met de VIL-sectoren BS (leidend), SB (helpend/vriendelijk) en SO (begrijpend). De hoogte van de significante correlaties varieert tussen .17 en .55 en het percentage verklaarde variantie varieert van 3% tot 30%. Zie tabel A in de Appendix bij dit hoofdstuk. Een aantal taken van de PDH en ook de PDH-overallschaal hangen significant negatief samen met de VIL-sectoren OT (onzeker) en TO (Ontevreden). Bij OT varieert de hoogte van deze correlaties van -.36 tot -.16 en de verklaarde variantie van 3 tot 13%. Bij TO varieert de hoogte van de correlaties van -.28 tot -.16 en de verklaarde variantie van 3 tot 8%.

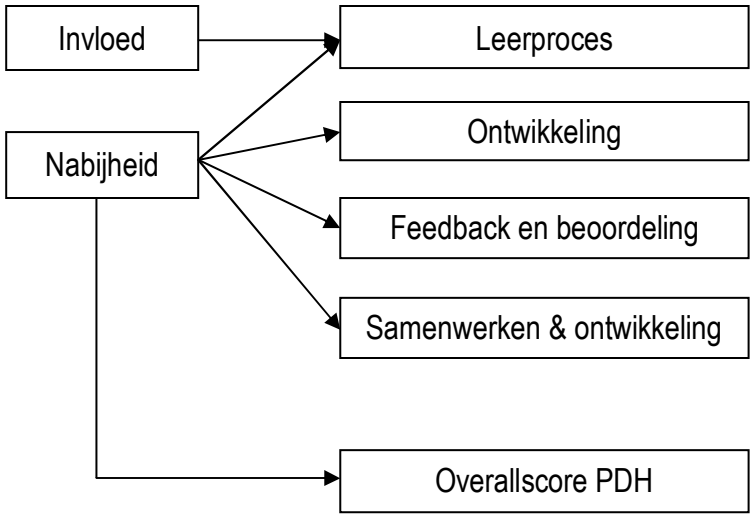
Dertien taken van de PDH hangen significant positief samen met de dimensie Invloed, waarbij de correlaties variëren van .15 tot .34. Hiermee wordt 2 tot 12% van de variantie verklaard. Alle taken van de PDH hangen significant positief samen met de dimensie Nabijheid waarbij de correlaties variëren van .15 tot .46. Hiermee wordt 2 tot 21% van de variantie verklaard.

Uit deze resultaten blijkt dat de PDH en de VIL niet sterk samenhangen. De PDH representeert kennelijk een andere dimensie van docenthandelen dan de dimensie waar de VIL zich op richt.

Aanvullend zijn multiple regressieanalyses uitgevoerd van de PDH-taken en de PDH-overallschaal op de beide dimensies van de VIL. Voorafgaand zijn diverse assumpties gecontroleerd. Beide predictoren hangen matig positief samen (.35) en de VIF (Variance Inflation Factor) is 1.14 zodat geen sprake is van multicollineariteit (Field, 2005). De variantie van de residuen voldoet aan de eis van homoscedasticiteit en de gestandaardiseerde residuen zijn normaal verdeeld. Er zijn geen outliers en het spreidingsdiagram wijst op lineariteit.

De uitkomsten van de regressieanalyses staan in tabel B van de Appendix. Beide VIL-dimensies zijn significant positieve voorspellers ($p \leq .05$) van zes taken van de PDH: 1 (zorgen voor een goed leerklimaat), 2 (bijdragen aan de persoonlijke vorming van leerlingen), 3 (geven van les, uitleg en instructie), 4 (leerlingen het nut van de leerstof laten zien), 6 (leerlingen laten werken aan opdrachten) en 9 (inspelen op verschillen tussen leerlingen). De dimensie Nabijheid is ook een significant positieve voorspeller voor de meeste andere taken, met uitzondering van 5 (leerlingen wijzen op verbanden tussen vakken), 8 (leerlingen vaardigheden aanleren) en 18 (bevorderen dat feedback wordt gebruikt), en is ook een significante voorspeller voor de overallschaal van de PDH (met een bèta van .35).

In de regressieanalyses zijn ook de in 4.4.2 beschreven vier takengroepen van de PDH meegenomen (niet in tabel B opgenomen). De dimensie Invloed is een significante voorspeller van de takengroep Leerproces (bèta .22) en de dimensie Nabijheid is dat van alle vier takengroepen (bèta's tussen .27 en .37). Figuur 4.4 brengt de significante relaties tussen de VIL-dimensies en de PDH in beeld.



Figuur 4.4: De VIL-dimensies als voorspellers voor de PDH (overall en vier groepen van taken)

4.4.4 Verschillen tussen groepen en indicaties voor confirmerende validiteit

De PDH bij verschillende groepen leerlingen

Op de data van 502 leerlingen zijn de relaties nagegaan tussen de beschikbare kenmerken van de leerlingen en hun scores op de PDH. Voor de exacte uitkomsten: zie tabel C van de Appendix.

Verschillen tussen jongens en meisjes
De jongens scoren het handelen van hun docenten op zeven taken (4, 7, 8, 10, 11, 13, 18) en op de overallschaal significant hoger dan de meisjes. Het verschil bij deze taken is steeds 0,2 of 0,3 schaalpunt en het verschil bij de overallschaal is 0,1 schaalpunt.

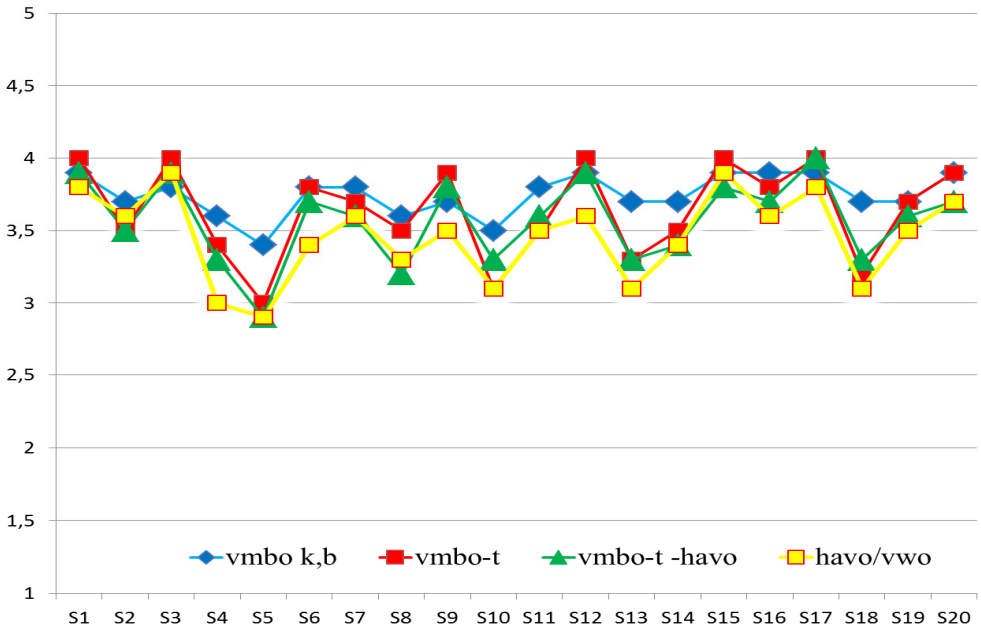
Verschillen tussen leeftijdsgroepen
Gezien de relatief beperkte aantallen leerlingen van 11-12 jaar en van 17-18 jaar zijn deze samengenomen met respectievelijk de 13-jarigen en de 16-jarigen. Dit resulteerde in de volgende indeling in vier leeftijdsgroepen: 11-13-jarigen (n=88), 14-jarigen (n=130), 15 jarigen (n=155) en 16-18-jarigen (n=127). Op 11 van de 20 taken is er een significant verschil tussen deze vier groepen: de taken 2, 4, 6, 8 en 11-17. Bij 10 daarvan zijn de gemiddelden in de twee middelste leeftijdsgroepen het laagst. Als proces in de tijd bezien kan worden gesteld dat de jongere leerlingen (11-13 jaar) hun docenten op deze taken relatief hoog scoren maar dat in de daarop volgende twee jaar (14 en 15 jaar) leerlingen hun docenten lager scoren, waarna de scores daarna (16-18 jaar) weer omhoog gaan. Deze afname bij 14 jaar is het grootst bij de taken 8 (van 3,7 naar 3,3, vaardigheden aanleren), 11 (van 4,1 naar 3,5, laten samenwerken en daarbij ondersteunen) en 14 (van 3,8 naar 3,3, eigen leerproces laten sturen). De vier groepen verschillen ook significant op de overallschaal (met als gemiddelden: 3,7, 3,5, 3,5, 3,7).

Verschillen tussen leerjaren
Deze analyse is gelet op de aantallen leerlingen per leerjaar beperkt tot de leerjaren 1 t/m 4. Deze verschillen significant op 10 van de 20 taken en de overallschaal. Het betreft de taken 2, 8, 9, 11, 12, 14 t/m 17 en 20. Bij de meeste van deze taken is een vergelijkbaar patroon zichtbaar als hierboven voor leeftijd is beschreven, met een meer of minder duidelijke afname in leerjaar 3. Bij drie taken is het gemiddelde in leerjaar 2 iets hoger dan in leerjaar 1, namelijk de taken 9 (inspelen op verschillen), 16 (volgen leerproces) en 20 (open staan voor leren en samenwerken). Mogelijk nemen leerlingen het docenthandelen op deze taken in het tweede leerjaar meer (of beter) waar dan in het eerste leerjaar, of betreft het (deels) andere docenten.

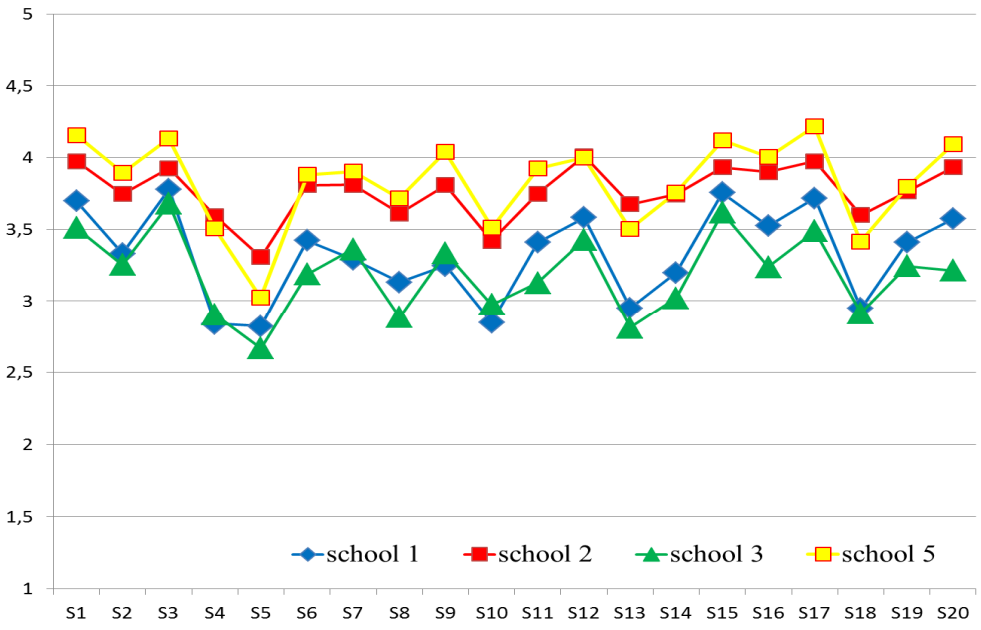
Verschillen tussen schooltypen
Deze analyse is gelet op de beschikbare aantallen leerlingen uitgevoerd met de volgende indeling in vier groepen: vmbo-beroeps en -kader (n=135), vmbo-t (n=92),

vmbo-t/havo (n=118) en havo/vwo (n=157). Deze groepen verschillen significant bij 11 taken (de taken 4, 5, 6, 8, 9, 10, 12, 13, 14, 16 en 18) en bij de overallschaal. Op deze taken scoren de vmbo-b/k leerlingen (en op enkele taken ook de vmbo-t leerlingen) systematisch hoger dan de vmbo-t/havo en havo/vwo leerlingen. Deze laatste groep scoort op 10 van de 11 taken waarin de groepen leerlingen significant van elkaar verschillen het laagst (taak 9 uitgezonderd). De verschillen tussen de schooltypen zijn zichtbaar in figuur 4.5.
Vervolgens is de variabele schooltype opgevat als ordinale variabele, en gecorreleerd (Spearman's rangcorrelatie) met de taken van de PDH en de overallschaal. Om een zo duidelijk mogelijk profiel te verkrijgen zijn de schooltypen hierbij niet geclassificeerd (alleen de zeven gymnasiumleerlingen zijn ingedeeld bij de 22 VWO-leerlingen, zodat zes waarden resteren: vmbo-beroeps en –kader, vmbo-t, vmbo-t/havo, havo, havo/vwo en vwo/gymnasium). De meeste taken correleren significant bij $p < .05$ met schooltype (van -0.09 (taak 20) tot -0.24 en dus alle negatief) met uitzondering van de taken 1, 2, 3, 7, 9, 11, 15 en 17. De overallscore correleert significant negatief met schooltype (r_s= -.14, p=.002). Uit een variantie-analyse met als factoren schooltype en school blijken bij de overallschaal de verschillen tussen de schooltypen niet significant maar die tussen de scholen wel ($F(5,483) = 10{,}066$, $p<.001$). Het interactie-effect is ook significant ($F(8,483) = 3{,}574$, $p < .001$. Eenzelfde beeld wordt zichtbaar in variantie-analyses per taak: de verschillen tussen de schooltypen zijn niet significant maar die tussen de scholen wel. Uitzonderingen zijn taak 3 (geven van les, uitleg en instructie) en taak 9 (inspelen op verschillen tussen leerlingen): bij deze taken zijn de verschillen tussen de schooltypen ook significant ($F(5,483) = 4{,}195$, p=.001, respectievelijk $F(5,483) = 3{,}237$, p=.007).

Verschillen tussen scholen
Gelet op het relatief gering aantal leerlingen van de scholen 4 (n=25) en 6 (n=11) is deze analyse beperkt tot de scholen 1 (n=75), 2 (n=178), 3 (n=108) en 5 (n=105). De overallscore verschilt significant tussen de scholen ($F(93,462) = 24{,}661$ bij $p<.001$) en ook op alle afzonderlijke taken (gemiddelde leerlingscores) verschillen de scholen significant ($p<.001$). Figuur 4.6 laat zien dat vooral sprake is van verschillen tussen de scholen 1 en 3 enerzijds en de scholen 2 en 5 anderzijds.



Figuur 4.5: De gemiddelde scores van de leerlingen op de PDH-taken per schooltype

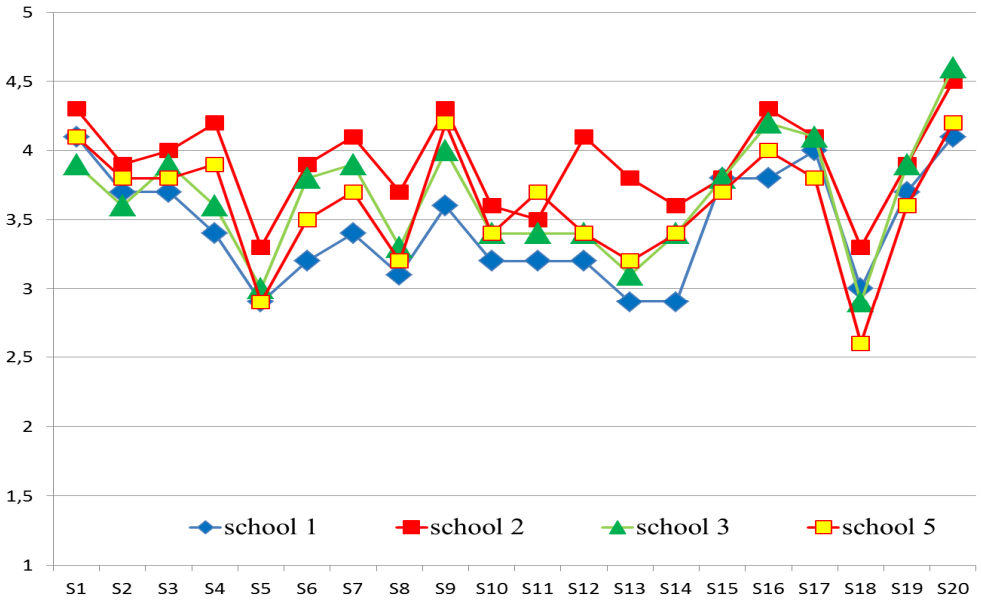


Figuur 4.6: De gemiddelde scores van de leerlingen op de PDH-taken per school

De PDH bij verschillende groepen docenten

Verschillen in aantal jaren onderwijservaring en onderwijservaring op deze school
Conform de indeling van Huberman (1990) (zie 4.2) zijn de docenten gelet op hun onderwijservaring verdeeld in vijf groepen: groep 1: 18 docenten (1-3 jaar ervaring); groep 2: 12 docenten (4-6 jaar ervaring); groep 3: 11 docenten (7-18 jaar ervaring), groep 4: 6 docenten (19-30 jaar ervaring) en groep 5: 2 docenten (31-40 jaar ervaring). De onderzoeksgroep is wat betreft onderwijservaring dus enigszins scheef verdeeld (afgerond 37%, 25%, 22%, 12% respectievelijk 4%). Ook de verdeling op grond van het totaal aantal jaren ervaring op de betreffende school is scheef: 34 (69%) docenten behoren tot groep 1 (1-3 jaar ervaring op deze school), 9 (18%) tot groep 2 (3-6 jaar ervaring), 4 (8%) tot groep 3 (7-18 jaar ervaring) en 2 (4%) tot groep 4 (19-30 jaar ervaring). Vanwege het kleine aantal docenten in de groepen 4 en 5 zijn de verschillen in beide gevallen (ervaring totaal en ervaring op deze school) geanalyseerd tussen de eerste drie groepen docenten. In beide gevallen zijn er tussen de drie groepen geen significante verschillen, noch op de overallschaal noch op de afzonderlijke taken.

Verschillen in geslacht en leeftijd
De PDH-scores verschillen niet significant tussen mannelijke en vrouwelijke docenten, noch op de overallschaal, noch op de taken, behalve op taak 9 (inspelen op verschillen tussen leerlingen). Hierbij scoren de mannen gemiddeld 3,7 en de vrouwen 4,3 ($t(47) = -3{,}687$, $p=.05$). De scores hangen ook niet significant samen met de leeftijd.



Figuur 4.7: De gemiddelde scores van de docenten op de PDH taken per school

Verschillen tussen scholen

Figuur 4.7 toont de verschillen in scores van de docenten op de taken tussen scholen. Het handelen van de docenten op de scholen 1, 2, 3 en 5 (respectievelijk 7, 16, 10 en 13 docenten) verschilt op vier taken significant:

Taak 4: leerlingen het nut van de leerstof laten zien ($F(3,42) = 3,493$, $p= .02$)

Taak 7: leerlingen stimuleren en activeren ($F(3,42) = 3.411$ $p=.03$)

Taak 12: leerlingen eigen keuzes laten maken ($F(3,42) = 4.423$, $p< .001$)

Taak 13: leerlingen steunen bij het sturen van hun leerproces ($F(3,42) = 3,240$, $p= .03$).

De docenten van school 1 scoren gemiddeld op deze vier taken steeds het laagst (3,4, 3,4, 3,2 en 2,9) en de docenten van school 2 steeds het hoogst (4,2, 4,1, 4,1 en 3,8). De verschillen tussen de docenten van de scholen 3 en 5 zijn slechts gering.

Verschillen in overeenstemming tussen docentprofiel en gemiddeld leerlingenprofiel

De sexe, leeftijd, totale onderwijservaring en ervaring op de betreffende school van de docenten staan niet in verband met de mate waarin het docentprofiel en het gemiddelde leerlingenprofiel correleren.

De PDH en kenmerken van leerlingen en docenten (analyse op drie niveaus)

Met multilevelanalyse (met het programma MLWin) is nagegaan in welke mate de scores van de leerlingen op de PDH variëren tussen scholen, docenten en leerlingen, en in hoeverre deze scores kunnen worden verklaard uit de beschikbare gegevens over kenmerken van de leerlingen en van hun docenten. Deze analyses zijn, net als de analyses in 4.4.5 van de verschillen tussen de scholen, beperkt tot de vier scholen met voldoende aantallen leerlingen. Deze 466 leerlingen hebben gegevens ingevuld over 46 docenten, met 6 tot 15 leerlingen per docent. De verdeling over de scholen is: 75, 178, 108 en 105 leerlingen respectievelijk 7, 16, 10 en 13 docenten. De analyses zijn uitgevoerd met de PDH-overallscore als criteriumvariabele op leerlingniveau en vervolgens met de gemiddelden op de vier in par. 4.4.2 onderscheiden groepen van taken (zie tabel 4.2). Op leerlingniveau zijn meegenomen: sexe, leeftijd, leerjaar en schooltype, en op docentniveau: geslacht, leeftijd, ervaring op deze school, totaal aantal jaren onderwijservaring en de correlatie tussen het profiel op de 20 PDH-taken volgens de docent en het profiel volgens zijn/haar leerlingen (zie 4.4.1).

Tabel 4.7 geeft de uitkomsten van de multilevelanalyses met als criteriumvariabele de PDH-overallscore op het huidig handelen van de docenten volgens hun leerlingen. In het nulmodel, dus zonder mogelijk verklarende predictoren, bevindt 64,7% van de variantie in de criteriumvariabele zich op leerlingniveau (significant bij $p<.05$), 23,9% op docentniveau (significant bij $p<.05$) en 11,4% op schoolniveau (niet significant).

In het leerlingmodel, met de leerlingkenmerken als predictoren, verandert de verdeling van de residuele variantie over de drie niveaus met 62,9%, 23,4% en 13,6% slechts weinig. Van de predictoren is alleen de sexe significant, met een bèta van -0.191 ($p=.002$), hetgeen betekent dat de jongens (op de gehanteerde schaal van 1 tot 5)

gemiddeld afgerond 0,2 schaalpunt hoger scoren dan de meisjes. Analyse met weglating van de niet-significante predictoren doet de deviance toenemen van 885.254 naar 888.344, maar dit verschil is niet significant ($\chi^2(2) = 3.09$, p=.213), zodat kan worden gesteld dat het model met alleen sexe als predictor even goed past.
Toevoeging aan dit model van alle beschikbare docentkenmerken blijkt de deviance amper te doen afnemen ($\chi^2(7) = 3.5$, p=.885), terwijl geen van de docentkenmerken een significante predictor blijkt te zijn. De verdeling van de residuele variantie over de niveaus is in dit model 64,6 %, 20,9 % en 14,5 %. Aanvullend is nog een model geanalyseerd met alleen de predictoren op docentniveau, dus zonder het leerlingkenmerk sexe; dit model past niet beter dan het nulmodel.

Tabel 4.7: Multilevelanalyses met de PDH-overallschaal als criteriumvariabele (data leerlingen)

Criterium variabele: huidig handelen (overallschaal) volgens de leerlingen	Nulmodel		Leerlingmodel met alle predictoren op leerlingniveau		Docentmodel met daarbij predictor sexe op leerlingniveau	
	Par.	*SE*	*Par.*	*SE*	*Par.*	*SE*
Fixed effects						
Intercept	3.520	0.137	3.510	0.148	3.514	0.149
Leerlingvariabelen						
sexe			-0.191	0.063	0.204	0.063
leeftijd			0.067	0.043		
leerjaar			0.092	0.061		
schooltype			0.030	0.051		
Docentvariabelen						
Leeftijd					0.009	0.007
geslacht					-0.113	0.124
Ervaring school					0.010	0.015
Ervaring onderwijs					0.000	0.008
Correlatie profielen					-0.048	0.263
Random effects						
Variantie op schoolniveau	0.060	0.053	0.072	0.061	0.075	0.062
Variantie op docentniveau	0.126	0.035	0.123	0.034	0.108	0.031
Variantie op leerlingniveau	0.341	0.024	0.331	0.023	0.334	0.023
Deviance	898.240		885.254		884.857	

Alle analyses zijn herhaald met achtereenvolgens de gemiddelden op de vier afzonderlijke groepen van taken als criteriumvariabele. Dit gaf enige verschillen te zien in de verdeling van de variantie over de drie niveaus in vergelijking met het nulmodel met de overall score als criteriumvariabele.
Tabel 4.8 geeft een overzicht van de verdeling van de variantie over de drie niveaus volgens het nulmodel en het eindmodel met alleen de sexe van de leerling als predictor en de vier groepen van taken als afzonderlijke criteriumvariabelen.

Tabel 4.8: Verdeling van de residuele variantie (in %) van de vier groepen taken over de drie niveaus

Groep van taken	Nulmodel			Eindmodel		
	School	*Docent*	*Leerling*	*School*	*Docent*	*Leerling*
Groep 1 (ontwikkeling)	12,5	19,9	67,5	16,2	16,9	66,9
Groep 2 (leerproces)	8,0	28,6	63,2	9,7	25,4	64,7
Groep 3 (feedback en beoordeling)	6,7	18,3	74,9	9,5	15,6	74,9
Groep 4 (samenwerken en ontwikkeling)	1,0	18,5	71,4	13,4	14,9	71,7

Vergeleken met het nulmodel bij de PDH-overallschaal (zie pagina 83) trekken de taken in groep 2 enige variantie weg van schoolniveau naar docentniveau, terwijl de scores op de taken in de groepen 3 (drie taken) en 4 (bij de leerlingen één taak) meer variëren tussen leerlingen en amper meer tussen scholen. Groep 2 verschilt het meest tussen docenten (het percentage verklaarde variantie is relatief gezien het hoogst op dat niveau); bij groep1 is de variantie op schoolniveau relatief het hoogst en bij groep 3 is dat het geval op leerlingniveau.
De algemene conclusie is dat bij deze leerlingen en docenten van de beschikbare persoonlijke kenmerken alleen de sexe van de leerlingen een significante (maar bescheiden) voorspeller is voor de scoring door de leerlingen van het pedagogisch-didactisch handelen van hun docenten.

4.4.5 Verschilscores tussen huidig en gewenst handelen vergeleken [2]

Hiervoor stond steeds het huidig handelen centraal. In deze paragraaf worden de resultaten weergegeven van het verschil tussen het huidig en het gewenst handelen volgens de data van docenten respectievelijk leerlingen en worden hun onderlinge verschillen getoetst en gerelateerd aan bepaalde kenmerken.

Verschillen tussen huidig en gewenst handelen bij docenten
De betrouwbaarheid van de schalen per taak voor het gewenst handelen bij de docenten varieert tussen .62 en .86, waarbij de Cronbach's alpha slechts bij vier taken kleiner is dan .70. De itemrestcorrelaties zijn vrijwel allemaal > .35 en in ieder geval > .20. Tabel 4.9 toont de gemiddelde verschilscore per taak tussen huidig en gewenst handelen volgens de docenten, de standaarddeviatie en de range.

[2] Vanwege de leesbaarheid is ervoor gekozen te spreken over het verschil tussen huidig en gewenst handelen. Bij de analyses zijn de verschilscores berekend op basis van het verschil tussen gewenst en huidig handelen. Een positieve score betekent hierdoor dat de respondenten meer willen, een negatieve score dat de respondenten minder willen dan nu.

Tabel 4.9: Uitkomsten docenten en leerlingen: gemiddelde verschilscores tussen huidig en gewenst

Taak	Docenten				Leerlingen				Verschil D-L	
	min	*max*	*gem*	*sd*	*min*	*max*	*gem*	*sd*	*t (513)*	*p*
1	-0.3	2.3	0.6	.458	-1.3	3.8	0.5	.717	-2.090	.835
2	.0	1.7	0.7	.516	-1.7	3.7	0.3	.698	2.498	.013*
3	.0	1.8	0.6	.445	-1.4	3.6	0.5	.700	1.784	.078
4	-1.0	2.0	0.5	.601	-2.0	4.0	0.5	.949	-1.557	.124
5	-0.3	2.7	0.9	.694	-2.3	4.0	0.5	.913	2.748	.006*
6	-0.3	2.2	0.5	.511	-2.2	3.7	0.5	.862	-0.297	.767
7	-0.3	2.0	0.6	.505	-1.8	4.0	0.6	.871	0.341	.734
8	.0	3.5	0.8	.702	-1.0	4.0	0.7	.939	1.427	.158
9	-0.8	1.5	0.3	.408	-1.5	4.0	0.7	.944	-5.431	≤.001*
10	-0.8	2.3	0.3	.738	-2.7	4.0	0.7	1.054	-4.271	≤.001*
11	-1.0	2.0	0.6	.605	-3.0	4.0	0.3	.903	2.059	.040*
12	-1.3	1.7	0.1	.571	-4.0	3.3	0.4	.829	-2.179	.030*
13	-0.5	4.0	0.7	.740	-3.3	4.0	0.6	.963	0.757	.452
14	-0.2	2.0	0.5	.462	-2.0	4.0	0.5	.859	-0.050	.960
15	-.2	1.5	0.3	.614	-3.0	4.0	0.2	.816	0.508	.611
16	.0	1.0	0.3	.315	-3.0	4.0	0.5	.918	-3.004	.003*
17	-0.3	2.3	0.5	.459	-2.0	4.0	0.4	.826	0.591	.556
18	-0.3	3.7	1.0	.723	-2.0	4.0	0.5	1.002	3.298	.002*
19	-0.2	2.4	0.5	.489	-2.0	4.0	0.5	.894	-0.154	.878
20	-0.3	2.0	0.2	.378	-2.3	4.0	0.4	.849	-3.165	.002*

**: significant bij p≤05*

De scholen (1, 2, 3 en 5) verschillen op een aantal taken significant van elkaar in gemiddelde verschilscore tussen huidig en gewenst handelen volgens de docenten. Dit doet zich voor bij taak 4 (leerlingen het nut en de samenhang laten zien) met een range van 0,1 tot 0,9 schaalpunt ($F(3,42) = 5,143$, $p=.004$), taak 7 (leerlingen stimuleren en activeren) met een range van 0,3 tot 0,9 schaalpunt ($F(3,42) = 2,948$, $p=.044$), taak 10 (inspelen op verschillen door middel van opdrachten) met een range van – 0,3 en 0,6 schaalpunt ($F(3,42) = 4,277$, $p=.01$) en taak 17 (geven van feedback aan leerlingen) met een range van 0,1 en 0,7 schaalpunt ($F(3,42) = 3,100$, $p=.040$).
Het verschil tussen huidig en gewenst handelen per taak volgens de docenten vormt over alle 20 taken een betrouwbare schaal (Cronbach´s alpha .86) met een gemiddelde van 0,5 schaalpunt (standaard-deviatie .299, range 0,0-1,6). Op deze overallschaal van verschilscores tussen huidig en gewenst handelen volgens de docenten verschillen de scholen (1, 2, 3 en 5) niet significant van elkaar.
De mate waarin de docenten meer of minder willen dan zij nu doen, hangt niet significant samen met hun geslacht, leeftijd, totale onderwijservaring en ervaring op deze school.

Verschillen tussen huidig en gewenst handelen bij leerlingen
De betrouwbaarheid van de schalen bij het gewenst handelen bij de leerlingen varieert tussen .64 en .88, en is slechts bij twee taken kleiner dan .70. De itemrestcorrelaties zijn >.46. Tabel 4.9 toont de gemiddelde verschilscores per taak tussen huidig en gewenst handelen volgens de leerlingen, alsmede de standaarddeviatie en de range.

De mate waarin leerlingen willen dat docenten meer doen op bepaalde taken dan nu verschilt per taak (zie tabel 4.9), maar op 15 van de 20 taken ook tussen jongens en meisjes (zie tabel 4.10). De verschillen zijn niet heel groot, maar wel significant. De meisjes blijken op alle taken iets meer van hun docenten te willen dan de jongens (maximaal 0,3 schaalpunt meer).

Tabel 4.10: Gemiddelde verschilscores tussen huidig en gewenst bij leerlingen uitgesplitst naar sexe

	Jongens (n=163)				Meisjes (n=273)				Verschil J-M	
taak	*min*	*max*	*gem*	*sd*	*min*	*max*	*gem*	*sd*	*t(434)*	*p*
1	-1.3	3.2	0.4	.656	-0.8	3.3	0.6	.704	-3.881	≤.001*
2	-1.8	2.4	0.3	.645	-1.0	2.6	0.5	.625	-3.115	.442
3	-1.3	2.5	0.4	.601	-1.0	3.5	0.5	.687	-2.630	.005*
4	-1.8	3.0	0.4	.798	-1.5	4.0	0.7	.980	-2.750	.004*
5	-2.0	4.0	0.4	.794	-2.0	4.0	0.6	.929	-1.253	.017*
6	-2.0	3.7	0.4	.764	-2.2	3.3	0.6	.888	-2.290	.067
7	-1.6	3.4	0.4	.696	-1.6	4.0	0.7	.845	-3.562	.002*
8	-0.8	3.8	0.5	.805	-1.0	3.8	0.7	.909	-2.810	.036*
9	-1.5	4.0	0.6	.899	-1.3	4.0	0.7	.944	-1.604	.372
10	-2.5	4.0	0.6	.948	-1.5	4.0	0.9	1.029	-3.049	.144*
11	-3.0	3.3	0.2	.837	-2.8	4.0	0.4	.957	-1.961	.023*
12	-1.4	3.0	0.3	.760	-1.4	3.0	0.4	.707	-1.704	.957
13	-2.0	4.0	0.4	.874	-2.0	4.0	0.6	.919	-2.904	.168
14	-1.4	2.6	0.4	.708	-2.0	4.0	0.6	.885	-2.437	.007*
15	-1.8	3.4	0.2	.709	-3.0	4.0	0.2	.849	-0.195	.139
16	-1.6	3.2	0.3	.782	-3.0	4.0	0.6	.982	-2.885	.013*
17	-2.0	4.0	0.3	.767	-2.0	4.0	0.5	.823	-3.420	.046*
18	-1.5	4.0	0.4	.854	-2.0	4.0	0.7	1.055	-2.430	.004*
19	-1.3	3.7	0.3	.690	-2.0	4.0	0.6	.909	-3.289	.001*
20	-1.8	2.4	0.3	.686	-2.0	4.0	0.5	.884	-2.311	.014*
Over all	-0.8	2.3	0.4	.526	-0.9	3.2	0.6	.667	-3.390	.001*

**: significant bij p≤.05*

De scholen (1, 2, 3 en 5) verschillen op een aantal taken significant van elkaar in gemiddelde verschilscore tussen huidig en gewenst handelen van de docenten volgens hun leerlingen. Dit doet zich voor bij taak 1 (zorgen voor een goed leerklimaat) met een range van 0,4 tot 0,8 ($F(3,462) = 8,257$, $p<.001$), taak 4 (leerlingen het nut van de leerstof laten zien) met een range van 0,5 tot 0,8 schaalpunt ($F(3,462) = 3,246$, $p=.022$, taak 7 (leerlingen stimuleren en activeren) met een range van 0,5 tot 0,8 schaalpunt ($F(3,462) = 3,119$, $p=.026$) en taak 15 (leerlingen ondersteunen bij het sturen van hun leerproces) met een range van 0 tot 0,3 ($F(3,462) = 2,723$, $p=.044$).
Het verschil tussen huidig en gewenst handelen per taak volgens de leerlingen vormt over alle 20 taken een betrouwbare schaal (Cronbach´s alpha .96) met een gemiddelde van 0,5 schaalpunt en een standaarddeviatie van .635 (minimumscore -0,9, maximumscore 3,3). Op deze overallschaal van verschilscores tussen huidig en gewenst handelen volgens de leerlingen verschillen de scholen (1, 2, 3 en 5) niet significant van elkaar.

De schooltypen (vmbo-beroeps en -kader (n=135), vmbo-t (n=92), vmbo-t/havo (n=118) en havo/vwo (n=157)) verschillen op negen van de 20 taken en op de overallscore significant van elkaar in hoeveel de leerlingen meer willen. Bij vijf taken scoren de havo/vwo-leerlingen hoger dan de andere drie groepen: taak 1 (zorgen voor een goed leerklimaat), taak 4 (leerlingen het nut van de leerstof laten zien), taak 6 (leerlingen leren werken aan opdrachten), taak 9 (inspelen op verschillen) en taak 12 (leerlingen eigen keuzes laten maken). Leerlingen van vmbo-beroeps en -kader scoren het hoogst op taak 5 (leerlingen wijzen op verbanden tussen vakken), taak 11 (leerlingen laten samenwerken en hun daarbij ondersteunen), taak 15 (leerlingen ondersteunen bij het sturen van hun leerproces) en taak 19 (beoordelen van het werk en de prestaties van leerlingen).
Op de overallschaal verschillen de schooltypen significant ($F(3, 498) = 4{,}519$, $p<.001$). Het verschil tussen vmbo-beroeps- en kader (kleinste verschil: 0,3) en vmbo-t (grootste verschil: 0,6) is significant (Scheffe's post hoc-test: $p=.021$), evenals het verschil tussen vmbo-t en havo/vwo ($p=.025$).
Uit een variantie-analyse met als factoren school en schooltype (de 7 gymnasium-leerlingen gevoegd bij de 22 vwo-leerlingen) en als afhankelijke variabele de overall-verschilscores tussen huidig en gewenst handelen van docenten volgens de leerlingen, blijkt dat zowel beide hoofdeffecten als het interactie-effect significant zijn: school: $F=(3,453) = 4{,}704$, $p=.003$; schooltype: $F(5,453) = 2{,}413$, $p=.036$, en het interactie-effect: $(F(4,453) = 4{,}424$, $p=.002$.
De overall-verschilscore verschilt niet significant tussen de docenten en de leerlingen. Bij een aantal afzonderlijke taken verschillen de leerlingen en docenten wel significant van elkaar in het verschil tussen huidig en gewenst handelen: taak 2 (bijdragen aan de persoonlijke vorming van leerlingen), taak 5 (leerlingen wijzen op verbanden tussen vakken), taak 9 (inspelen op verschillen tussen leerlingen), taak 10 (inspelen op verschillen door middel van opdrachten), taak 11 (leerlingen laten samenwerken en hun daarbij ondersteunen), taak 12 (leerlingen eigen keuzes laten maken), taak 16 (volgen wat leerlingen leren en hoe ze zich ontwikkelen), taak 18 (bevorderen dat feedback wordt gebruikt) en taak 20 (open staan voor leren en samenwerken). Het verschil tussen huidig en gewenst handelen is bij de docenten bij bijna alle taken gemiddeld net iets groter dan bij de leerlingen, behalve bij de taken 9, 10 en 12; bij deze taken is het verschil bij de leerlingen groter (zie ook tabel 4.9).

4.4.6 Aanwijzingen voor consequentiële validiteit

Tot slot behandelen we de verkregen indicaties voor wat wordt genoemd de consequentiële validiteit. Gezien de aanleiding voor en bedoeling van de PDH is het van belang na te gaan of de onderwijspraktijk met dit instrument uit de voeten kan en er ook echt iets aan heeft. Dit is door ons nog niet systematisch onderzocht, maar er kan wel verslag worden gedaan van de ervaringen tot nu toe die zijn opgedaan bij het op verzoek (doen) afnemen van het instrument en het terugkoppelen van de uitkomsten

daarvan aan schoolleiders, teamleiders en docenten op verschillende scholen. Hieronder wordt verslag gedaan van de ervaringen die zijn opgedaan met de uitgebreide versie van het instrument (op de scholen die deelnamen aan het in hoofdstuk 3 beschreven onderzoek) en met de verkorte versie PDH (op de scholen die deelnamen aan het onderzoek waarvan in dit hoofdstuk verslag wordt gedaan). Daarnaast is inmiddels ook ervaring opgedaan met het terugkoppelen van de gegevens van de twee onderzoeken die elk zijn toegespitst op één van de taakgebieden, namelijk feedback (zie hoofdstuk 5) en beoordeling (zie hoofdstuk 6).

Na de ontwikkeling van de uitgebreide versie van het instrument hebben de scholen die hebben meegedaan een schriftelijke rapportage toegestuurd gekregen, waarin de gegevens waren uitgewerkt op school- en teamniveau (waarbij ook de uitkomsten van de verschillende docenten binnen het team, weliswaar geanonimiseerd, zichtbaar waren gemaakt). De rapportage is mondeling besproken met de schoolleiders die daar prijs op stelden. Daarbij werd ook de relatie met het onderwijsconcept van de school gelegd. De gegeven toelichting werd door de schoolleiders gewaardeerd. Het hielp hen de gegevens te interpreteren en betekenis te geven in relatie tot hun onderwijsconcept, speerpunten te formuleren en prioriteiten te stellen voor de toekomst. Zij gaven aan dat uitkomsten een reëel beeld gaven van de huidige situatie op hun school.

Op drie scholen zijn de uitkomsten in aanwezigheid van de onderzoeker besproken met de teamleiders van de betreffende school. Tijdens deze bijeenkomsten bleek dat de school- en teamleiders zich konden vinden in de uitkomsten, zoals gepresenteerd door de onderzoekers, in die zin dat ze deze herkenden, konden duiden en konden koppelen aan de prioriteiten die waren gelegd binnen de school en/of het team of aan de ontwikkelingsfase waarin de school of het team zich bevond. Met name het zichtbaar maken van de verschillen tussen docenten in huidig en gewenst handelen en de relatie met de gegevens van de leerlingen, uitgesplitst naar door de schoolleiders op voorhand gekozen kenmerken, leverde veel herkenning, en ook onderlinge discussie op. Naar aanleiding van de uitkomsten ontstond tijdens deze bijeenkomsten vaak tegelijkertijd een inhoudelijk discussie over het belang van bepaalde taken en hun onderlinge relaties. Hierdoor kon het handelen van de docenten ook worden gerelateerd aan wat bekend is uit de wetenschappelijke literatuur en/of het belang van bepaalde taken van docenten binnen het vernieuwende onderwijsconcept van de betreffende school.

Apart zijn de schoolleiders die bij dit onderzoek waren betrokken, uitgenodigd voor een gezamenlijke bijeenkomst, waarbij de uitkomsten op schoolniveau tussen de scholen werden vergeleken. Hierbij stond, naast het kennis nemen van elkaars uitkomsten, het duiden en betekenis geven van de uitkomsten centraal, werd opnieuw de relatie gelegd met wat bekend is uit wetenschappelijk onderzoek en werd gebrainstormd over mogelijke verbeterpunten.

De terugkoppeling over dit onderzoek gebeurde met tabellen en grafieken, waarbij deze laatste de verschillen tussen taken binnen het profiel van een individuele docent en tussen de profielen van verschillende docenten, maar ook tussen teams beter zichtbaar maakten, waardoor de uitkomsten meer gingen spreken en meer betekenis kregen. De

belangrijkste conclusies die op grond van de uitkomsten konden worden getrokken werden door de onderzoekers in een aparte paragraaf beschreven.
Een aantal schoolleiders en teamleiders benoemden dat hulp van de onderzoekers bij het interpreteren van de gegevens voor hen essentieel was, omdat zij niet vanzelfsprekend gewend zijn dergelijke gegevens te lezen en betekenis te geven. Deze ervaring komt overeen met uitkomsten uit eerder onderzoek waaruit blijkt dat zogenoemde "datageletterdheidcompetenties" bij schoolleiders niet vanzelfsprekend aanwezig zijn, waardoor zij soms moeite hebben feedbackgegevens te duiden en te relateren aan concrete doelstellingen in het kader van interne kwaliteitszorg en externe verantwoording (zie o.a Geijsel & Krüger, 2005; Luo, 2008; Verhaeghe, Vanhoof, Valcke & Van Petegem, 2011). Dit pleit voor een aanbod waarbij niet alleen wordt volstaan met een schriftelijke rapportage, maar tevens de dialoog wordt aangegaan door de betreffende onderzoekers met de betrokkenen op de school.
De ervaringen die zijn opgedaan met de rapportages en de terugkoppelingen naar aanleiding van het eerste onderzoek zijn benut bij het terugkoppelen van de uitkomsten van het aanvullende onderzoek met de verkorte versie van de vragenlijst. Afhankelijk van het aantal docenten dat mee had gedaan in dit vervolgtraject, werden na het terugkoppelen van gegevens gesprekken gevoerd met de afzonderlijke teamleiders of een plenaire bijeenkomst met de schoolleiding en alle teamleiders belegd.
In de schriftelijke terugrapportage die voorafgaand aan de gesprekken werd toegestuurd, werden de uitkomsten nog meer gevisualiseerd aan de hand van grafieken en werd een stappenplan toegevoegd waarmee de betrokkene de gegevens zelf systematisch kon nalopen, interpreteren en betekenis geven. In de gesprekken met de afzonderlijke teamleiders bleek opnieuw de variatie in datageletterdheid en in bekwaamheid om gegevens te relateren aan concrete doelstellingen. De ene teamleider was hier zelfstandig toe in staat en al een eind mee gevorderd voordat het gesprek plaatsvond, terwijl de ander extra uitleg en toelichting nodig had om de gegevens te kunnen interpreteren. In beide gevallen leverde het gesprek (alsnog) inhoudelijk een zinvolle bijdrage aan het duiden van de uitkomsten.
Tijdens een plenaire bijeenkomst op één van de scholen is, aan het eind van een bijeenkomst waarin de uitkomsten werden besproken, onderling mondeling en daarna door de onderzoeker schriftelijk geïnventariseerd hoe duidelijk en begrijpelijk de betrokkenen de terugrapportage vonden. Op deze school waren de volgende rapportages verzorgd: een schoolrapport (met daarin ook zichtbaar gemaakte niet geanonimiseerde verschillen tussen teams), een teamrapport (met geanonimiseerde verschillen tussen individuele docenten) en docentrapportages (uitsluitend gegevens van de docent zelf; deze werden per post toegestuurd aan de betreffende docent). De betrokkenen tijdens de bijeenkomst (één schoolleider en negen teamleiders) gaven aan dat de rapportages duidelijk waren en de gegevens op een begrijpelijk wijze werden teruggekoppeld. Ook de terugrapportages voor de individuele docenten werden door hen duidelijk en begrijpelijk gevonden.

Op de vraag wat de belangrijkste opbrengst was van dit traject, antwoordde men steevast in termen van het verkrijgen van meer inzicht (in het verschil tussen huidig en gewenst handelen, in wat de leerlingen aangeven en willen, in de verschillen tussen docenten, in wat men al goed doet, in waar concrete punten liggen ter verbetering). Daarnaast gaven betrokkenen aan dat de rapportages en de bijeenkomsten daadwerkelijk de dialoog op gang brachten, waarbij werd benadrukt dat het erg behulpzaam was dat deze dialoog nu ook op basis van concreet materiaal gevoerd kon worden.
Hierboven is vooral verslag gedaan van de terugkoppeling van gegevens naar schoolleiders en teamleiders. In de onderzoeken die worden beschreven in de hoofdstukken 5 en 6 hebben terugkoppelingen van gegevens ook naar individuele docenten plaatsgevonden in de vorm van schriftelijke rapportages met een toegevoegde interpretatie, al dan niet aangevuld met persoonlijke gesprekken. Ook hierbij bleek dat niet iedereen vanzelfsprekend gemakkelijk kan omgaan met zulke gegevens en dat een persoonlijk gesprek naar aanleiding van de rapportage heel verhelderend kan zijn en kan bijdragen aan het kunnen relateren van uitkomsten aan concrete doelstellingen en handelingen.

4.5 Conclusies en discussie

We geven hier eerst antwoorden op de onderzoeksvragen.

Betrouwbaarheid
Bij de leerlingen bleek de betrouwbaarheid van de schalen van het huidig handelen bij alle 20 taken goed tot zeer goed. Bij de docenten was bij een aantal schalen een aanpassing nodig (verwijdering van een of twee items met te lage itemrestcorrelatie) waardoor de betrouwbaarheid van de schalen voldoende verbeterde (Cronbach's alpha minimaal $\geq .60$).
Groepering van de taken op basis van een eerder onderzoek leidde bij de data in onderhavig onderzoek tot vier betrouwbare schalen, met tien taken gericht op de ontwikkeling van leerlingen, zes taken gericht op het leerproces van leerlingen, drie taken gericht op het geven van feedback en het beoordelen, en in de vierde groep taak 20 (gericht op samenwerken en ontwikkelen).
De interbeoordelaarsbetrouwbaarheid van de leerlingen per docent (jury alpha) is bij ruim 70% van de docenten voldoende tot goed. De correlatie tussen het scoreprofiel op de 20 taken volgens de docent zelf en het gemiddelde profiel volgens de leerlingen is bij rond 70% significant positief.

Onderscheidingsvermogen
De gemiddelde scores per taak bij de docenten variëren van 2,9 tot 4,4 en bij de leerlingen van 3,0 tot 3,9. Ook de gemiddelden op de onderscheiden groepen van taken variëren sterk, zowel bij de docenten (range 2,1 tot 3,7) als bij de leerlingen (range 2,8

tot 4,0). Op individueel niveau bestrijkt de range bij de meeste taken vrijwel de hele schaal, en op itemniveau worden steeds alle scores van 1 t/m 5 benut.
De intraclassecorrelatie ICC(1) is bij de leerlingen voor de afzonderlijke taken tussen 0.20 en 0.36 (alleen bij taak 10 lager: 0.14), bij de vier groepen van taken tussen 0.27 en 0.36 en bij de overallschaal 0.35. Dit zijn voor onderwijsonderzoek relatief hoge waarden, docenten worden goed onderscheiden. De bijbehorende betrouwbaarheid van de op docentniveau geaggregeerde leerlingscores ICC(2) per taak varieert tussen .79 en .93 (alleen bij taak 10 weer lager: .69) en is voor de overallschaal .93.
Via clusteranalyse kunnen drie groepen docenten worden onderscheiden die duidelijk verschillen in de hoogte van de scores op de groepen taken. Deze groepen verschillen niet in het geslacht, de leeftijd of de ervaring van de docenten en hangen ook niet samen met de school waar ze werken. Docenten kunnen ook worden onderscheiden in de mate waarin hun eigen beoordeling van hun huidig handelen overeenkomt met de gemiddelde beoordeling van hun leerlingen (deze correlatie varieert van rond 0 tot .80).

Discriminante validiteit
De taken, groepen taken en overallschaal van de PDH hangen enigszins maar niet sterk samen met de beide dimensies Invloed en Nabijheid van de VIL. De PDH representeert kennelijk inderdaad een andere dimensie van het handelen van docenten dan de VIL. De dimensie Invloed is wel een significant positieve voorspeller voor zes taken van de PDH en voor de groep taken over het leerproces en de dimensie Nabijheid is een significant positieve voorspeller voor de meeste andere taken, alle vier groepen taken, en de overallschaal van de PDH.

Verschillen tussen groepen en confirmerende validiteit
Bij de leerlingen scoren de jongens op een aantal taken gemiddeld iets hoger dan de meisjes. Leerlingen van 14-15 jaar en leerlingen in leerjaar 3 scoren op een aantal taken gemiddeld lager dan de overige leerlingen. De leerlingen die een vmbo-basis of –kader opleiding volgen scoren op de meeste taken gemiddeld hoger dan de overige leerlingen. De scores van de leerlingen verschillen op alle taken, de groepen taken en de overallschaal significant tussen de scholen.
Bij de docenten is tussen mannen en vrouwen alleen een verschil gevonden in hun scores op schaal 9 (inspelen op verschillen tussen leerlingen); vrouwen scoren hierop gemiddeld duidelijk hoger dan mannen. Er zijn geen significante relaties tussen de scores op de PDH en de mate van ervaring of de leeftijd. Het handelen van de docenten verschilt op vier van de 20 taken significant tussen de scholen.
Van de scores van de leerlingen op de overallschaal bevindt 65% van de variantie zich op leerlingniveau, 24% op docentniveau en 11% op schoolniveau. Dit betekent dat 24% van de variantie kan worden toegeschreven aan het feit dat de leerlingen de vragenlijst hebben ingevuld voor verschillende docenten en 65% te maken heeft met individuele verschillen tussen leerlingen. Van alle bij de multilevelanalyse betrokken persoonskenmerken is alleen het geslacht van de leerlingen een significante predictor.

Consequentiële validiteit

Er is nog geen systematisch onderzoek gedaan naar de consequentiële validiteit maar er zijn wel aanwijzingen dat schoolleiders, teamleiders en docenten de uitkomsten van de PDH kunnen duiden in relatie tot de dagelijkse schoolpraktijk, de fase waarin de school of het team zich bevindt en het eigen handelen in relatie tot de leerlingen. Op grond van terugrapportages en presentaties bleek men in staat (al dan niet met ondersteuning van een onderzoeker) het handelen van docenten te relateren aan het onderwijsconcept van de school, speerpunten te formuleren en prioriteiten te stellen voor de toekomst.

Het gebruik van de PDH wordt door de betrokkenen gezien als een systematische en bruikbare manier om gegevens te verzamelen en feedback te genereren op een concreet niveau. Dat daarbij verschillen tussen huidig en gewenst handelen zichtbaar worden gemaakt, zowel vanuit het perspectief van de docenten zelf als vanuit dat van hun leerlingen, wordt als waardevol gezien. De uitkomsten blijken op elke school een aanzet te geven tot dialoog tussen schoolleiding en teamleiders en docenten onderling. Een enkele docent gebruikt de uitkomsten uit zichzelf ook om met de leerlingen in gesprek te gaan.

Discussie

Van de in 4.2 beschreven en in deze studie gehanteerde criteria voor bepaling van de kenmerken van een instrument is gebleken dat ze kunnen leiden tot elkaar aanvullende en gedifferentieerde inzichten in de kwaliteit ervan. Dit betreft de constructvaliditeit, inhoudsvaliditeit, ecologische en systemische validiteit, schaalbetrouwbaarheid, interbeoordelaarsbetrouwbaarheid (tussen de leerlingen onderling per docent, tussen het gemiddelde van de leerlingen en de docent), onderscheidingsvermogen (tussen docenten), discriminante validiteit en consequentiële validiteit.

Door het gevolgde ontwikkelproces van de uitgebreide versie (zie hoofdstuk 3) zijn waarborgen ingebouwd voor de constructvaliditeit (gerichtheid op het construct ‘pedagogisch-didactisch handelen’) en de inhoudsvaliditeit (dekking van een breed repertoire van taakgebieden), alsmede voor de ecologische validiteit (herkenbaarheid voor docenten en leerlingen; zie Beijaard, 1990) en de *systemic validity* (congruentie met en bijdragend aan de beoogde professionele ontwikkeling in de context van het onderwijsconcept van de scholen; zie Frederikson & Collins, 1989). Voor de constructvaliditeit, inhoudsvaliditeit, ecologische validiteit en systemic validity van de verkorte versie (die in het onderhavige hoofdstuk is onderzocht) kan worden verwezen naar de uitgebreide versie waarop die is gebaseerd, met dien verstande dat de *systemic validity* door het schrappen van een aantal items mogelijk enigszins is verminderd.

Ook het feit dat bij een aantal taken de schaalbetrouwbaarheid en de intraclasse-correlatie iets lager zijn dan bij de overeenkomstige taakgebieden in de uitgebreide versie moet mogelijk in verband worden gebracht met de beperktere operationalisering door de inperking van het aantal items. De interbeoordelaarsovereenstemming lijkt hier echter niet onder te hebben geleden: ook bij de verkorte versie is deze maat bij de

leerlingen per docent bij bijna driekwart van de docenten voldoende tot goed, terwijl de correlatie tussen het profiel volgens de docent en het gemiddelde profiel volgens zijn leerlingen bij bijna driekwart van de docenten significant positief is.
Indicaties voor discriminante validiteit van de PDH zijn verkregen door middel van correlaties van de taken, de onderscheiden groepen taken en de overallscore van de PDH met de dimensies Invloed en Nabijheid van de VIL. De PDH hangt significant maar gematigd positief samen met beide dimensies van de VIL. Daaruit kan worden geconcludeerd dat deze instrumenten twee te onderscheiden constructen meten.
De gevonden relaties van de VIL-dimensie Invloed met de PDH-takengroep leerproces en de VIL-dimensie Nabijheid met alle vier PDH-takengroepen passen goed bij de aard van de taken die in deze groepen aan de orde zijn. De PDH bestrijkt naast meer traditionele taken die zijn gericht op het cognitieve leerproces een aantal min of meer vernieuwende taken die de afgelopen 15 jaar in het voortgezet onderwijs meer aandacht hebben gekregen en een toenemende gerichtheid op de leerlingen inhouden en die minder met de dimensie Invloed en meer met de dimensie Nabijheid samenhangen.
Gelet op de positieve relaties van de (taken van de) PDH met de VIL-sectoren BS, SB en SO en de negatieve relaties met de sectoren OT en TO, en de in onderzoek vastgestelde relaties van deze sectoren met het leerproces van leerlingen, kan worden verondersteld dat zulke relaties ook bij (taken van) de PDH kunnen worden gevonden. Een mogelijke interpretatie van de bescheiden maar positieve en interpreteerbare relatie tussen VIL en PDH is dat adequaat interpersoonlijk handelen een voorwaarde is om het pedagogisch-didactisch handelen tot zijn recht te kunnen laten komen.
Indicaties voor confirmerende validiteit zijn verkregen door middel van analyses van de verschillen tussen groepen die kunnen worden gevormd met behulp van een aantal kenmerken van de leerlingen (geslacht, leeftijd, leerjaar, schooltype, school) en de docenten (geslacht, leeftijd, school, ervaring op deze school, ervaring totaal).
Het pedagogisch-didactisch handelen van docenten, ook op vernieuwende scholen, blijkt individueel in hoge mate te variëren. Deze grote verschillen tussen docenten, ook op dezelfde school, betekent dat zij vrij sterk verschillen in hun handelen naar de leerlingen, mogelijk op bepaalde taken meer dan wenselijk wordt geacht. Dit kan betekenen dat er voor scholen respectievelijk docenten aanleiding is om daarover het gesprek aan te gaan. Tussen de scores op de PDH en de genoemde persoonskenmerken van docenten werden geen relaties verwacht, en ook niet gevonden. De variatie in de scores hangt niet sterk samen met kenmerken zoals geslacht, leeftijd en ervaring.
Dat geen duidelijke relatie is gevonden tussen de scores op de PDH en de leeftijd en ervaring van de docenten kan komen doordat in dit onderzoek relatief weinig oudere en meer ervaren docenten waren betrokken. Het kan ook komen doordat de docenten in dit onderzoek werkzaam zijn in een context van verandering en vernieuwing. Ook docenten die al de nodige ervaring hebben, kunnen daardoor opnieuw in een leerproces betrokken raken, met gevolgen voor hun pedagogisch-didactisch handelen. Aangezien een deel van de taken van de PDH betrekking heeft op relatief vernieuwend (en niet gemakkelijk) handelen, is deze mogelijkheid niet denkbeeldig.

De *gemiddelde* scores op de taken liggen tussen 2,9 en 4,4 (docenten) respectievelijk 3,0 en 3,9 (leerlingen) en dus ongeveer in het derde kwartiel van de schaal van 1-5. Dit kan worden gezien als indicatie dat veel docenten in behoorlijke mate de verschillende taken vervullen. Hierbij kan echter een rol spelen dat de docenten in dit onderzoek werkzaam zijn op scholen met een vernieuwend onderwijsconcept. Mogelijk scoren docenten op meer traditionele scholen op een aantal taken gemiddeld minder hoog.
Bij de leerlingen werden wel enkele relaties met onderscheiden persoonskenmerken gevonden. Dat bij de leerlingen de jongens op een aantal taken gemiddeld iets hoger scoren dan de meisjes kan betekenen dat meisjes gemiddeld iets kritischer aankijken tegen het pedagogisch-didactisch handelen van hun docenten. Dat leerlingen van 14-15 jaar en leerlingen in leerjaar 3 op een aantal taken lager scoren dan jongere en oudere leerlingen past wellicht bij de kritische houding die pubers kunnen hebben.
Dat de leerlingen die een vmbo-basis of –kader opleiding volgen op de meeste taken gemiddeld hoger scoren dan de overige leerlingen kan betekenen dat het pedagogisch-didactisch repertoire van docenten op zulke opleidingen gemiddeld meer is ontwikkeld en/of meer wordt aangesproken.
Zowel de scores van de leerlingen op het huidig handelen als de verschilscores tussen huidig en gewenst handelen bij docenten en leerlingen verschillen tussen taken en tussen scholen. Deze verschillen duiden mogelijk op reële verschillen tussen scholen in de mate waarin docenten aandacht geven aan de taken, en verdienen nader onderzoek.
Beperkingen van dit onderzoek waren het deels beperkte aantal leerlingen per docent, het relatief beperkte aantal oudere en meer ervaren docenten, en het gebruik van alleen schriftelijke c.q. digitale voorgestructureerde vragenlijsten als instrument.
In dit onderzoek bleek dat de operationalisering van taak 10 aandacht behoeft, evenals de interne consistentie bij verschillende taken bij de docenten. Ook zou het de bruikbaarheid van de PDH kunnen vergroten als deze nog iets zou kunnen worden ingekort, bijvoorbeeld door het schrappen van enkele minder belangrijke taken en/of van een of twee taken die moeilijk betrouwbaar te krijgen zijn.
Verder onderzoek is interessant en relevant, onder meer naar de onderscheiden groepen taken. Zo is het bijvoorbeeld intrigerend dat, vergeleken met de verdeling van de variantie in de overallschaal bij de leerlingen over de drie niveaus, er variantie bij de takengroep *leerproces* verschuift van school- naar docentniveau, bij de takengroep *ontwikkeling* van docent- naar leerlingniveau, en bij de takengroep *feedback en beoordeling* van school- en docentniveau naar leerlingniveau. Ook is verder onderzoek wenselijk naar de consequentiële validiteit van de PDH, in relatie tot het gebruik van het instrument voor professionele ontwikkeling van docenten en schoolontwikkeling. Tot slot is onderzoek gewenst naar de predictieve validiteit, zowel per taak (in aansluiting op de onderzoekstradities die op het terrein van een aantal in de PDH onderscheiden taken al bestaan), als in relatie tot het repertoire als geheel van het pedagogisch-didactisch handelen van docenten.

Tabel A: Correlaties tussen de PDH-taken en de VIL-sectoren en -dimensies in de data van de leerlingen

PDH	*BS*	*SB*	*SO*	*OS*	*OT*	*TO*	*TB*	*BT*	*Invloed*	*Nabijheid*
1	.51**	.55**	.42**	.23**	-.36**	-.28**	-.11	-.02	.32**	.46**
2	.42**	.46**	.28**	.14	-.33**	-.16*	-.05	.04	.33**	.33**
3	.44*	.41**	.31**	.08	-.27**	-.20**	-.01	.05	.33**	.31**
4	.35**	.44**	.33**	-.00	-.33**	-.18*	-.06	.03	.34**	.32**
5	.19*	.30**	.16*	.07	-.09	-.02	.02	.05	.14	.15*
6	.33**	.45**	.35**	.07	-.26**	-.21**	-.11	.00	.26**	.35**
7	.45**	.48**	.39**	.17*	-.32**	-.30**	-.18*	-.06	.27**	.44**
8	.20**	.18**	.17*	.14	-.09	.03	.04	.04	.12	.17*
9	.43**	.41**	.41**	. 12	-.31**	-.19*	-.12	.02	.30**	.34**
10	.21**	.24**	.24**	.12	-.07	-.12	-.14	-.06	.05	.24**
11	.20**	.25**	.21**	.16*	-.07	-0.6	-.02	-.08	.05	.21**
12	.27**	.52**	.37**	.15	-.20**	-.20**	-.18*	-.23**	.10	.40**
13	.18*	.35**	.22**	-.04	-.12	-.06	-.07	-.03	.16*	.21**
14	.22**	.27**	.27**	.02	-.19*	-.15	-.13	-.15*	.14	.29**
15	.14	.18**	.18*	.08	-.12	-.02	-.09	-.07	.07	.19*
16	.26**	.31**	.31**	.15	-.17*	-.07*	-.02	-.06	.13	.27**
17	.30**	.42**	.33**	.09	-.23**	-.17*	-.16*	-.10	.17*	.34**
18	.18*	.29**	.20**	.00	-.23**	.02	-.06	-.01	.18*	.18*
19	.28**	.42**	.30**	.09	-.16*	-.05	-.08	-.04	.15*	.28**
20	.32**	.50**	.37**	.11	-.25**	-.16*	-.16*	-.14	.18*	.38**
Overall (O)	.38**	.54**	.38**	.13	-.27**	-.16*	-.11	-.05	.25**	.39**

* p≤ .05; ** p≤.001

Tabel B: Uitkomsten multiple regressieanalyses van de twee dimensies van de VIL op de taken en de overallschaal van de PDH

	Model			Dimensie Invloed (VIL)					Dimensie Nabijheid (VIL)			
PDH	R^2	*F (df=2)*	*p*	*C*	*B*	*β*	*t*	*p*	*B*	*β*	*t*	*p*
1	.24	27.232	≤.001**	3.955	.091	.185	2.614	.010*	.111	.394	5.547	≤.001**
2	.16	16.350	≤.001**	3.156	.134	.237	3.181	.002*	.081	.249	3.334	.001**
3	.15	15.581	≤.001**	3.713	.122	.252	3.362	.001**	.062	.224	2.995	.003*
4	.17	16.968	≤.001**	2.704	.181	.261	3.505	.001**	.092	.233	3.128	.002*
5	.03	2.897	.058	2.586	.076	.100	1.246	.215	.052	.119	1.492	.138
6	.14	14.403	≤.001**	3.123	.095	.153	2.035	.043*	.106	.297	3.937	≤.001**
7	.21	22.393	≤.001**	2.986	.092	.137	1.885	.061	.150	.388	5.353	≤.001**
8	.03	2.808	.063	2.895	.048	.065	.816	.415	.061	.144	1.797	.074
9	.18	19.419	≤.001**	3.054	.134	.185	2.518	.013*	.136	.328	4.456	≤.001**
10	.06	5.337	.006*	2.738	-.025	-.034	-.431	.667	.106	.252	3.184	.002*
11	.04	3.883	.022*	3.234	-.019	-.026	-.329	.743	.089	.216	2.707	.007*
12	.16	16.892	≤.001**	3.304	-.031	-.052	-.696	.487	.142	.420	5.648	≤.001**
13	.05	4.598	.011*	2.739	.070	.094	1.181	.239	0.74	.175	2.202	.029*
14	.09	8.228	≤.001**	2.896	.031	.046	.586	.558	.108	.276	3.554	≤.001**
15	.04	3.365	.037*	3.594	.000	.001	.009	.993	.064	.194	2.427	.016*
16	.08	7.136	.001**	3.199	.027	.039	.497	.620	.105	.261	3.333	.001**
17	.12	11.783	≤.001**	3.387	.036	.055	.725	.470	.120	.324	4.241	≤.001**
18	.05	4.289	.015*	2.672	.097	.128	1.616	.108	.059	.137	1.721	.087
19	.08	7.674	.001**	3.166	.041	.062	.795	.428	.098	.259	3.314	.001**
20	.15	15.149	≤.001**	3.078	.036	.052	.696	.487	.143	.366	4.869	≤.001**
Overall	.17	16.943	≤.001**	3.081	.062	.125	1.677	.095	.098	.345	4.633	≤.001**

* p≤ .05; ** p≤ .001

Tabel C: Verschillen tussen jongens en meisjes en tussen de leerjaren op de PDH-taken en de PDH-overallschaal

	M		V		t-toets		Leerjaar 1		Leerjaar 2		Leerjaar 3		Leerjaar 4		F-toets	
T	*M*	*SD*	*M*	*SD*	*t (434)*	*p*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F (4, 497)*	*p*
1	4.0	.689	3.9	.749	1.628	.104	4.1	.551	4.0	.769	3.8	.768	3.8	.757	2.364	.052
2	3.7	.737	3.6	.754	1.443	.150	3.9	.651	3.8	.725	3.5	.812	3.5	.756	5.734	≤ .001*
3	4.0	.698	3.9	.727	1.237	.217	4.1	.535	3.8	.774	3.8	.798	3.9	.740	1.690	.151
4	3.5	.892	3.3	1.013	2.092	.037*	3.3	.940	3.3	1.009	3.2	1.054	3.4	.991	1.299	.269
5	3.2	1.076	3.0	1.029	1.740	.082	3.4	.943	2.8	1.127	3.1	1.068	2.9	1.072	3.015	.080
6	3.7	.798	3.6	.903	.799	.425	3.8	.760	3.7	.885	3.5	.942	3.7	.852	2.011	.092
7	3.9	.800	3.6	.917	2.535	.012*	3.8	.763	3.7	.881	3.6	1.002	3.7	.882	.976	.420
8	3.6	.946	3.4	.984	2.477	.014*	3.8	.898	3.7	.790	3.3	1.096	3.3	.952	5.050	.001*
9	3.7	.988	3.7	.997	.260	.795	3.6	.923	3.9	1.071	3.5	1.027	3.8	.976	2.753	.028*
10	3.4	.984	3.2	1.013	2.153	.032*	3.4	1.023	3.3	.885	3.2	1.078	3.2	1.004	.649	.628
11	3.8	.892	3.5	1.011	2.684	.008*	4.0	.819	4.0	.756	3.4	1.044	3.6	.959	8.115	≤ .001*
12	3.9	.789	3.8	.782	.519	.604	3.9	.663	3.6	.766	3.7	.868	4.0	.723	5.083	.001*
13	3.6	.919	3.3	1.018	2.705	.007*	3.5	.877	3.4	.827	3.2	1.099	3.4	.989	1.851	.118
14	3.6	.924	3.5	.922	1.449	.148	3.8	.775	3.7	.736	3.3	1.017	3.5	.935	4.015	.003*
15	4.0	.773	3.9	.795	1.191	.234	4.1	.714	4.1	.709	3.8	.898	3.9	.745	3.108	.015*
16	3.9	.847	3.7	.998	1.628	.104	3.8	.907	4.0	.736	3.5	1.034	3.8	.922	3.583	.007*
17	4.0	.791	3.9	.915	1.865	.063	4.0	.727	4.0	.812	3.8	.995	4.0	.825	2.642	.033*
18	3.6	1.003	3.3	1.112	3.150	.002*	3.5	.937	3.6	1.011	3.2	1.197	3.3	1.087	2326	.055
19	3.8	.786	3.6	.952	1.912	.057	3.7	.849	3.6	.789	3.5	.991	3.6	.899	.684	.603
20	3.9	.814	3.8	.977	1.351	.178	3.9	.611	4.1	.845	3.6	1.047	3.8	.927	3.760	.005*
O#	3.7	.648	3.6	.732	2.352	.0.23*	3.8	.559	3.7	.671	3.5	.799	3.6	.679	3.005	.018*

*significant bij tweezijdig toetsen bij p≤ .05; O#: Overallscore PDH

Tabel C (vervolg): Verschillen tussen de schooltypen en tussen de leeftijdsgroepen op de PDH-taken en de PDH-overallschaal

	vmbo-b,k		vmbo-t		vmbo-t/havo		havo/vwo		F-toets		11-13 jaar		14 jaar		15 jaar		16-18 jaar		F-toets	
T	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F (3,498)*	*p*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F(3,496)*	*p*
1	3.9	.708	4.0	.592	3.9	.835	3.8	.738	2.565	.054	4.0	.631	3.9	.760	3.8	.821	3.9	.649	1.640	.179
2	3.7	.738	3.5	.696	3.5	.926	3.6	.699	2.347	.072	3.8	.690	3.6	.816	3.5	.835	3.6	.656	3.988	.008*
3	3.8	.773	4.0	.602	3.9	.856	3.9	.669	13.934	.123	4.0	.620	3.8	.816	3.8	.789	4.0	.645	1.742	.157
4	3.6	1.014	3.4	.826	3.3	1.040	3.0	.992	8.096	≤.001*	3.2	.974	3.2	1.058	3.3	1.003	3.5	.928	3.473	.016*
5	3.4	.957	3.0	.906	2.9	1.199	2.9	1.049	9.468	≤.001*	3.2	1.053	2.9	1.097	3.0	1.096	3.1	.983	2.125	.096
6	3.8	.860	3.8	.705	3.7	.951	3.4	.875	4.969	.002*	3.7	.808	3.5	.955	3.6	.905	3.8	.756	3.240	.022*
7	3.8	.887	3.7	.745	3.6	1.030	3.6	.903	1.591	.191	3.8	.797	3.6	1.018	3.6	.936	3.8	.801	1.617	.184
8	3.6	.996	3.5	.824	3.2	1.003	3.3	1.049	4.475	.004*	3.7	.865	3.3	1.065	3.3	1.014	3.5	.935	5.057	.002*
9	3.7	.997	3.9	.827	3.8	1.027	3.5	1.032	3.955	.008*	3.6	.967	3.6	1.096	3.7	.973	3.9	.921	1.977	.117
10	3.5	1.060	3.1	.911	3.3	1.072	3.1	.975	3.790	.010*	3.4	.980	3.1	1.089	3.2	.968	3.4	1.023	1.956	.120
11	3.8	.933	3.5	.873	3.6	.937	3.5	1.090	2.253	.081	4.1	.825	3.5	1.018	3.4	.977	3.7	.938	10.710	≤.001*
12	3.9	.725	4.0	.733	3.9	.781	3.6	.821	6.451	≤.001*	3.8	.732	3.6	.893	3.9	.765	4.0	.642	8.024	≤.001*
13	3.7	.883	3.3	.960	3.3	1.074	3.1	.966	12.093	≤.001*	3.5	.873	3.1	1.102	3.3	.991	3.6	.937	5.451	≤.001*
14	3.7	.879	3.5	.826	3.4	1.084	3.4	.876	4.182	.006*	3.8	.757	3.3	.968	3.4	1.000	3.7	.843	6.952	≤.001*
15	3.9	.834	4.0	.663	3.8	.839	3.9	.807	.629	.597	4.1	.700	3.8	.950	3.8	.786	3.9	.678	3.045	.029*
16	3.9	.982	3.8	.759	3.7	1.009	3.6	.968	2.703	.045*	3.9	.862	3.5	1.027	3.7	.981	3.9	.870	3.633	.013*
17	3.9	.891	4.0	.698	4.0	.940	3.8	.909	1.368	.252	4.1	.729	3.8	.981	3.8	.937	4.0	.754	3.278	.021*
18	3.7	1.010	3.2	1.012	3.3	1.183	3.1	1.069	9.108	≤.001*	3.5	.985	3.2	1.183	3.2	1.165	3.4	.983	1.925	.125
19	3.7	.942	3.7	.773	3.6	1.000	3.5	.875	2.395	.068	3.7	.791	3.5	.976	3.6	.990	3.7	.807	1.891	.130
20	3.9	.893	3.9	.720	3.7	1.090	3.7	.934	2.170	.091	3.9	.672	3.7	1.047	3.7	1.002	3.9	.854	2.416	.066
O#	3.8	.731	3.6	.550	3.6	.790	3.5	.689	4.495	.004*	3.7	.592	3.5	.800	3.5	.743	3.7	.618	4.210	.006*

* significant bij tweezijdig toetsen bij p≤ .05; O#: Overallscore PDH

5 Nader onderzoek naar geven van feedback

5.1 Inleiding

De operationalisering van de pedagogisch-didactische taakgebieden van docenten zoals verwerkt in het in hoofdstuk 3 beschreven instrument, kan worden aangevuld door er per taakgebied meer specifieke wetenschappelijke kennis uit theorie en onderzoek bij te betrekken. Dit maakt het mogelijk het handelen van docenten meer gedifferentieerd te onderzoeken. In plaats van één instrument voor het hele repertoire dat geschikt is voor grootschalige afname wordt dan gekozen voor meer diepgaand en kleinschalig onderzoek waarin het mogelijk is het handelen van docenten op meer manieren (zoals observeren en bevragen op cognities en directe vergelijking van docent- en leerling-percepties) en op meer momenten in kaart te brengen. Dit is uiteraard alleen uitvoerbaar door in te zoomen op een bepaald taakgebied.
Naar aanleiding van de toenemende aandacht, zowel in het onderwijs als in theorie en onderzoek, voor het belang van feedback in onderwijsleerprocessen, is nader onderzoek gericht op dat taakgebied. Voor het onderwijs is dit taakgebied relevant omdat het geven van feedback regelmatig wordt beschreven als een potentieel zeer effectieve didactische interventie. Vanuit wetenschappelijk oogpunt is interessant na te gaan op welke manieren docenten feedback geven, in hoeverre zelfbeschrijvingen door docenten van hun feedbackgedrag, rapportage daarover door hun leerlingen, en observaties door een onderzoeker overeenkomen, wat de opvattingen, intenties en percepties van docenten zijn bij het geven van feedback en hoe de relatie is tussen hun gedrag en hun cognities, en in welke mate gedrag en cognities van docenten op dit taakgebied variëren en waarmee die variatie samenhangt.
De vraagstelling voor het onderzoek luidt: Hoe geven docenten mondeling feedback als leerlingen werken aan een taak en hoe verhoudt dit zich tot wat volgens de literatuur bijdraagt aan motivatie en leren van de leerlingen?
De onderzoeksvragen zijn:

1. Welke aanwijzingen bevat de wetenschappelijke literatuur voor het geven van feedback?
2. Wat zijn de opvattingen en intenties van docenten met betrekking tot het geven van feedback?
3. Hoe geven docenten feitelijk feedback, volgens henzelf, volgens hun leerlingen, en volgens observaties door onderzoekers, en hoe hangt dit samen met hun opvattingen en intenties?
4. Hoe verhoudt zich het handelen van docenten op dit taakgebied in het algemeen tot hun handelen in een specifieke situatie (tijdens een lesuur in een klas)?

5. In hoeverre hangt de motivatie van de leerlingen voor het maken van taken voor het vak bij de betreffende docent samen met hun inschatting en waardering voor zijn handelen op dit taakgebied?
6. In hoeverre hangen de verschillen tussen docenten in hun opvattingen en handelen op dit taakgebied samen met de scholen waar zij werken?

5.2 Feedback in de wetenschappelijke literatuur

Vormen en omschrijvingen van feedback in het onderwijs

Vormen van feedback

Leerlingen die een toets maken of aan een werkstuk werken, kunnen op drie manieren feedback krijgen. Zij kunnen tijdens het werken aan de taak merken of het goed gaat (feedback vanuit de taak) of direct na afronding (het doel is wel of niet bereikt). Een tweede manier is dat de leerling actief nadenkt over de uitvoering en het resultaat. In dit geval dringt de feedback zich niet vanuit de taak op, maar is reflectie nodig om iets te signaleren en te concluderen (zelf gegenereerde feedback). De derde manier is dat iemand anders ziet of heeft gezien hoe het gaat (vordert, de goede kant op gaat) en wat je hebt gemaakt, en dat aan je meedeelt (feedback van een ander).

In het onderwijs is de feedbackgever vaak een docent, maar het kan ook een andere leerling zijn (peer-feedback), of, in de praktijkcomponent van beroepsopleidingen, een patiënt of klant of opdrachtgever. De rollen kunnen ook worden omgedraaid: docenten (al dan niet in opleiding) kunnen ook de ontvangers van feedback zijn, waaronder feedback van hun leerlingen.

Behalve dat de bron van de feedback kan verschillen (taak, zelf, ander), kan ook het medium of kanaal variëren: mondeling, schriftelijk (op papier) of digitaal, en verbaal of non-verbaal (in face-to-face situaties met intonatie, mimiek en gebaren, en schriftelijk en digitaal met afbeeldingen, grafieken en symbolen). Feedback kan zowel individueel worden gegeven als aan een groep, en zowel op initiatief van de feedbackgever als in reactie op vragen of problemen van leerlingen.

Objecten van feedback

We richten ons hier verder op feedback van een docent aan een leerling (of een groepje leerlingen), op eigen initiatief of naar aanleiding van een vraag of probleem van de leerling, bij een bepaalde taak.

De leerling werkt aan een taak of heeft daaraan gewerkt, en de feedback kan zich dan richten op een of meer objecten: de aanpak die de leerling volgt (of heeft gevolgd), het proces of verloop, en het tussentijdse of uiteindelijke resultaat: de (deels) gemaakte opgave of het (deels) gemaakte werkstuk. De feedback kan behalve op de taak (aanpak, uitvoering, resultaat) ook worden gericht op de leerling zelf en zijn of haar houding, motivatie of gedrag.

De taak waar het om gaat kan veel uitmaken, zowel in het soort problemen waar een leerling tegenaan kan lopen als in de feedback die kan of zou moeten worden gegeven. Taken kunnen verschillen in de hoeveelheid kennis die nodig is om de taak uit te voeren, de mate waarin de aanpak of procedure die moet of kan worden gevolgd vastligt, en de vraag of slechts één antwoord of oplossing goed is of dat meerdere goed kunnen zijn (Baxter & Glaser, 1998; Pellegrino, Baxter & Glaser, 1999). Hierbij zullen zich ook verschillen kunnen voordoen tussen schoolvakken: bij het ene vak zijn taken gemiddeld kennisintensiever dan bij het andere, en bij het ene vak kunnen vaker verschillende aanpakken worden gevolgd en meerdere uitkomsten goed zijn dan bij het andere. Docenten zullen mede daardoor kunnen verschillen in hun feedbackrepertoire.

Omschrijvingen van feedback

De literatuur bevat uiteenlopende omschrijvingen van feedback. Askew en Lodge (2000) maken onderscheid tussen feedback als "gift", als "ping-pong" en als "dialoog". Feedback als gift is eenrichtingverkeer: de feedbackgever reageert op of geeft informatie aan de feedbackontvanger, in de vorm van een beschrijving en/of beoordeling van diens prestatie op een taak. Zulke feedback kan bijdragen aan het leerproces maar kan ook geen effect hebben of juist het leren blokkeren. Het wordt aan de ontvanger overgelaten de feedback te verwerken en er al dan niet iets mee te doen.

Feedback als ping-pong doet zich voor als de feedback uitnodigt tot een reactie van de ontvanger. De feedback zal hier meer dan in de vorige situatie zijn gericht op het bevorderen van reflectie op de taak en de eigen prestatie en/of het bevorderen van de motivatie om te werken aan de taak.

Bij feedback als dialoog staat het beoordelende aspect minder centraal en zijn de betrokkenen vooral gericht op het uitwisselen van persoonlijke betekenistoekenningen en het gezamenlijke leerproces. Dit leerproces betreft dan zowel de taak en de uitvoering daarvan als de manier waarop beide partners deze waarnemen en waarderen en wat beiden in hun interactie met hun inbreng bedoelen.

Van de Ridder, McGaghie, Stokking en Ten Cate (2008) maken een vergelijkbaar onderscheid, tussen feedback als informatie (eenrichtingverkeer), feedback als reactie (met interactie tussen gever en ontvanger), en feedback als cyclus (waarbij de ontvanger de feedback gebruikt ter verbetering van de prestatie en daarop opnieuw feedback krijgt). Bij taken in onderwijs en opleidingen is altijd sprake van bepaalde eisen of standaarden waaraan moet worden voldaan en heeft feedback altijd de functie om bij te dragen aan leren en verbetering.

Een definitie die al deze ingrediënten bevat is die van Ramaprasad (1983, p. 4): *"Feedback is information about the gap between the actual level and the reference level of a parameter which is used to alter the gap in some way"*. Deze formele definitie is algemeen toepasbaar, in veel contexten. Een specifieke omschrijving voor onderwijs die dezelfde elementen bevat, is die van Sadler (1989, p. 121): *"the learner has to (a) possess a concept of the standard (or goal, or reference level) being aimed*

for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap."
Iemand die feedback ontvangt, zal daaraan zelf betekenis geven. Feedback hoeft niet te worden gegeven met de intentie bij te dragen aan leren en verbetering om door de ontvanger wel zo te worden waargenomen en gebruikt. Bij feedback in onderwijssituaties, door een docent (of een medeleerling) aan een leerling, mogen we er vanuit gaan dat degene die feedback geeft dat steeds wel doet met die intentie. In navolging van Van de Ridder et al. (2008) voegen we daarom deze intentie toe. Al met al ontstaat dan de volgende omschrijving: *Feedback is een reactie op een prestatie op een taak met een standaard, gericht op leren en verbetering.*

Feedback als interventie: veelbelovend maar ongewis

Over feedback als didactische interventie in onderwijs, opleidingen en HRD wordt vaak opvallend positief geschreven (zie bijvoorbeeld Hattie en Timperley (2007): "The power of feedback"). Van tijd tot tijd verschijnen overzichtsstudies en meta-analyses waaruit feedback soms inderdaad naar voren komt als een krachtig middel voor het bevorderen van leerresultaten, maar soms ook niet. In zulke publicaties worden grote aantallen onderzoeken vergeleken die verschillen in context, doelgroep, taak en feedbackvorm, en in de kwaliteit van het onderzoek zelf. Het effect van feedback waarnaar men zoekt, is dan een gemiddelde over in feite onvergelijkbare grootheden. We noemen drie van zulke studies, waarin steeds werd gezocht naar factoren die samenhangen met cognitieve leerresultaten.
Fraser, Walberg, Welch en Hattie (1987) gaven een synthese van enkele duizenden onderzoeken waarin 26 zeer verschillende aspecten van onderwijs waren onderzocht. Feedback behoorde, samen met enkele geheel andere activiteiten, tot de top-4 van meest effectieve interventies.
Wang, Haertel en Walberg (1993) analyseerden ruim 100 overzichtsartikelen en meta-analyses. Zij vergeleken 28 categorieën van factoren, variërend van "*classroom management*" tot demografische kenmerken van de schoolomgeving. Feedback scoorde in het midden, samen met "student and teacher academic interactions" (kennelijk werd feedback niet gezien als 'academic interaction').
Scheerens (2007) vergeleek 12 brede categorieën van schoolfactoren en 15 onderwijsvariabelen op klasniveau. Bij de "schoolfactoren" eindigde "*reïnforcement en feedback*" op de laatste plaats, bij de variabelen op klasniveau kwam "feedback/ controle" uit op de tiende plek.
Uit deze voorbeelden blijkt dat feedback er soms wel maar soms ook niet positief uitkomt. In zulke studies worden de onderzochte factoren, waaronder feedback, nogal verschillend en vaak slechts globaal aangeduid. Dat maakt hun waarde beperkt.
Er zijn ook studies verricht waarin de onderzochte factoren nauwkeuriger zijn omschreven. Twee daarvan zijn Bangert-Drowns, Kulik, Kulik en Morgan (1991) en Kluger en DeNisi (1996). In beide studies bleek dat feedback weliswaar gemiddeld een

duidelijk positief effect heeft, maar in een derde deel van de gevallen geen of zelfs een negatief effect. Hierbij moet wel worden aangetekend dat de meeste onderzoeken die in deze studies zijn vergeleken, plaatsvonden in laboratoriumsituaties en/of betrekking hadden op kleine afgebakende taken waarvoor kennis en/of vaardigheden nodig zijn die in vrij korte tijd kunnen worden geleerd, waardoor het effect van feedback daarop gemakkelijk experimenteel kan worden onderzocht. De generaliseerbaarheid daarvan naar de vele verschillende leertaken en situaties in de onderwijspraktijk is twijfelachtig, maar zelfs indien beperkt tot de genoemde specifieke situaties en taken is het effect van feedback kennelijk ongewis.

Uitkomsten van onderzoeken naar effecten van feedback in het onderwijs kunnen variëren door verschillen tussen onderzoeken in opzet en kwaliteit, maar ook door verschillen in relatie tot de taak, de docent, het communicatieproces, de leerling (de ontvanger van de feedback), en de context.

Bij veel hedendaagse taken is geen sprake meer van eenduidig goede of foute antwoorden of oplossingen. Werken aan een taak is vaak een proces van enige duur waarbij een leerling keuzes moet maken en op verschillende momenten en manieren een vergissing kan begaan of kan vastlopen. De “prestaties” van leerlingen waarop feedback zou kunnen worden gegeven, zijn daardoor vaak niet meer zo eenduidig. Dit is ook het geval als een leerling wel tot een goede oplossing komt, omdat er vaak meer wegen zijn waarlangs deze kan zijn bereikt en een docent in zijn of haar feedback vaak ook het oplossingsproces zal willen betrekken. Ook zijn bij vrij veel taken meer goede oplossingen mogelijk, of kunnen oplossingen verschillen in kwaliteit.

De beoordeling van de prestatie, die volgens de eerder gegeven omschrijving van feedback aan het geven daarvan ten grondslag zou moeten liggen, is hierdoor ook minder eenvoudig geworden. De docent zal vaak het proces moeten reconstrueren en de kwaliteit van de aanpak, de gemaakte keuzes en het resultaat moeten afwegen. Om deze reconstructie en beoordeling te kunnen uitvoeren en om de boodschap duidelijk over te kunnen brengen zullen docent en leerling vaak moeten samenwerken, tijdens een proces van communicatie en interactie.

De deskundigheid en bekwaamheid van de docent als beoordelaar en feedbackgever speelt ook een grote rol. In de beschreven studies waarin feedback vaak betrekking had op eenmalige reacties op prestaties op taken met eenduidig goede of foute antwoorden of oplossingen, bleek de geloofwaardigheid van de feedbackgever een van de factoren te zijn die van invloed was op het effect van de feedback. Des te meer zal dit het geval zijn in de situatie waarin taken, prestaties en beoordelingen minder eenduidig zijn. Een docent zal zijn of haar vak goed moeten kennen, de kennis en vaardigheden die daarin centraal staan, en de manieren waarop leerlingen zich deze kunnen eigen maken en de problemen die daarbij kunnen optreden. Een docent zal goed moeten kunnen observeren en analyseren om zicht te krijgen op wat een leerling heeft gedaan en gedacht en hoe de kwaliteit van het (voorlopige) resultaat kan worden ingeschat. Een docent zal, ten derde, in staat moeten zijn de boodschap die hij of zij wil overbrengen zorgvuldig, duidelijk en precies te formuleren.

Het communicatieproces tussen docent en leerling over de taak en de uitvoering en het (voorlopige) resultaat is intrinsiek complex en ambigu. Dit proces bevat meerdere lagen van betekenis: die van de taal, die van de onderlinge sociale relatie, en die van de constructie van kennis (Sadler, 1998; Torrance & Pryor, 1998). De kwaliteit van de interactie is hierdoor een belangrijke factor in het succes van de feedback. Docenten en leerlingen kunnen elkaar verkeerd begrijpen, en er kunnen zich discrepanties voordoen tussen wat een docent met de feedback bedoelt en hoe een leerling deze interpreteert.
Bij leerlingen doen zich individueel verschillende cognitieve, affectieve en motivationele processen voor die van invloed zijn op het kennisnemen, begrijpen, accepteren en willen en kunnen benutten van de feedback. Deze hebben onder meer betrekking op de verwerking van informatie (cognitive load), de toeschrijving van succes en falen (attributie), de oriëntatie op een doel (leren, presteren, gunstig afsteken bij anderen), het zelfvertrouwen voor de taak (self-efficacy), en behoeften aan relatie, competentie en autonomie (Deci & Ryan, 2000; Ilgen & Davis, 2000; Shute, 2008).
In de studies die wisselende resultaten van feedback te zien gaven, ging het steeds om feedback als eenmalige reactie of informatie. Zodra feedback onderdeel is van een proces van interactie zou de kans dat feedback een negatief effect heeft, moeten kunnen afnemen, doordat misverstanden tussen docent en leerling of affectieve processen die de feedback bij de leerling teweegbrengt, kunnen worden gesignaleerd en bijgestuurd. Anderzijds wordt het ook lastiger om de effecten van feedback op zich te onderzoeken. Anders gezegd: waar feedback wordt gegeven in de context van een proces, zal het ook moeten worden onderzocht en beoordeeld in die context.

Feedback bezien vanuit verschillende invalshoeken

De diversiteit in publicaties over feedback hangt ook samen met het feit dat feedback wordt benaderd vanuit verschillende invalshoeken en met het oog op verschillende contexten. Men kan feedback zien als formatieve beoordeling, als reactie op of informatie over een prestatie, als proces van communicatie en interactie, en als onderdeel van begeleiding of coaching. Deze opties representeren min of meer de ontwikkeling die zich de laatste 40 jaar heeft voorgedaan, waarbij men het beoordelen van werk van leerlingen steeds meer is gaan zien als integraal onderdeel van het onderwijsleerproces.

Feedback als formatieve beoordeling
Kenmerkend aan deze benadering is dat feedback geacht wordt te zijn gebaseerd op beoordeling. Eind 80-er jaren hebben drie studies de aftrap gegeven voor meer aandacht voor het geven van zulke feedback door docenten aan hun leerlingen: Natriello (1987), Crooks (1988) en Sadler (1989).
Natriello (1987) plaatste feedback in de volgende reeks stappen. 1. Bepalen van de doelen van de beoordeling (zoals: prestaties vaststellen, beslissingen nemen, motivatie stimuleren, leren bevorderen). 2. Ontwerpen van geschikte taken (wie wil kunnen

beoordelen wat leerlingen weten en kunnen, moet hen taken geven waaraan dat kan worden afgelezen). 3. Vaststellen van de beoordelingscriteria (zoals: wordt bij een werkstuk voor geschiedenis ook gelet op de kwaliteit van het schrijven? gaat het ook om de manier waarop eraan is gewerkt?). 4. Vaststellen van standaarden (inhoudelijk en/of vergelijkend). 5. Verzamelen van gegevens over de prestaties van de leerlingen (objectief toetsbaar en/of kwalitatief beoordeelbaar; meer of minder frequent en over een kortere of langere periode). 6. Beoordelen van de prestaties (waarbij menselijke beoordelaars fouten kunnen maken en beoordelingen door leerlingen kunnen worden ervaren als niet terecht of onbegrijpelijk en dan in mindere mate zullen bijdragen aan het verdere leerproces). 7. Geven van feedback (een aparte stap, omdat de docent altijd moet kiezen waarop hij of zij zal reageren en op welke manier en met welke boodschap). 8. Monitoren van de resultaten (nagaan of de met de beoordeling gestelde doelen ook zijn of worden gerealiseerd).

Crooks (1988) ging nader in op formatieve beoordeling door docenten als integraal onderdeel van het onderwijsleerproces. Hij inventariseerde de concrete vormen daarvan: met door de docent zelf gemaakte toetsen, toetsen als onderdeel van het lesmateriaal, schriftelijke vragen en oefeningen, mondelinge vragen, "and a wide variety of other performance activities" (p. 439). Hij benoemde ook de verschillende niveaus waarop zich effecten daarvan kunnen voordoen: het in een les of lessenserie of rond een taak activeren van voorkennis, het richten van de aandacht, het benadrukken van leerdoelen en standaarden, het bevorderen van een actieve leerstrategie, het bevorderen van motivatie, het terugkoppelen van het resultaat en geven van correctieve feedback, en de leerling helpen de voortgang zelf te monitoren. Crooks kwam tot de conclusie dat formatieve beoordeling door docenten krachtige directe en indirecte impact kan hebben, maar dat die impact zowel positief als negatief kan zijn, en dat formatieve beoordeling dus heel doordacht moet worden ingezet.

Sadler (1989) zoomde verder in op het belang van de aard van de taak die de leerling moet uitvoeren, de manier waarop docenten prestaties kunnen beoordelen, en het uiteindelijke doel dat formatieve beoordeling moet dienen. Hij constateerde dat ondanks regelmatige accurate feedback vaak geen verbetering optreedt van de kwaliteit van het werk van leerlingen. Volgens hem gaat het bij formatieve beoordeling over beoordeling van de kwaliteit van prestaties van leerlingen, met het oogmerk hun vaardigheden verder te ontwikkelen. Docenten hebben om te kunnen beoordelen een begrip van kwaliteit nodig dat past bij de taak. De leerling moet dit begrip van kwaliteit ook ontwikkelen, zodat hij of zij de eigen taakuitvoering al tijdens het proces kan beoordelen en over een repertoire beschikt van acties die kunnen worden ondernomen als er moet worden bijgestuurd. Onderwijs zou hierop moeten zijn gericht: dat leerlingen hun eigen werk op kwaliteit kunnen beoordelen.

Docenten vinden het volgens Sadler vaak lastig de standaard precies te omschrijven. Bij het beoordelen van de kwaliteit van complexe prestaties speelt altijd een aantal criteria een rol, die deels niet scherp zijn en waarvan per beoordeling slechts een deel manifest wordt. Dit laatste komt doordat criteria vaak pas naar boven komen als er niet

aan wordt voldaan. Een voorbeeld: tijdens het lezen van een werkstuk ben je er als docent niet steeds op gericht na te gaan of de zinnen grammaticaal juist zijn geformuleerd, maar zodra je een zin tegenkomt die dat niet is, wordt dit criterium geactiveerd. Het ontwikkelen van vaardigheden door leerlingen vergt oefening en feedback. In de kern betreft feedback het aangeven van de standaard, het terugkoppelen van de huidige prestatie, de vergelijking tussen die twee, en het suggereren van acties die de afstand tussen huidige prestatie en standaard kunnen terugbrengen. Hoe complexer echter de taak, hoe waarschijnlijker het is dat er meer manieren zijn om aan de standaard te gaan voldoen, en dan zal feedback meer moeten inhouden dan alleen het terugkoppelen van het verschil tussen prestatie en standaard en het suggereren van enkele acties.
Dat het belangrijk is, zoals Crooks en Sadler aangeven, dat leerlingen zelf de kwaliteit van hun werk leren beoordelen, is sindsdien regelmatig onderstreept. Leerlingen zouden hun houding moeten veranderen van "voldoen aan wat de docent wil zien" in "zelf nadenken over hoe het beter kan". Feedback zou daartoe moeten worden gevolgd door een dialoog (interactie) over de taak, de aanpak, het resultaat en de standaard, en de relaties daartussen (Smith & Ragan, 1993; Straub, 1996, 1997).
Sadler (1998) voegde hieraan toe dat leerlingen ook moeten leren hoe ze de feedback van de docent moeten interpreteren. De manier waarop de feedback wordt gecommuniceerd is volgens hem heel belangrijk, waarbij ook nog eens meespeelt dat het gaat om communicatie tussen twee ongelijke partijen (docent en leerling). In een communicatieproces kennen beide partijen aan elkaars uitingen betekenissen toe, waarbij zowel de inhoud als de onderlinge relatie een rol speelt en de deelnemers geneigd zijn alle onderdelen van de boodschap aan elkaar te verbinden.

Feedback als informatie
De al genoemde Bangert-Drowns et al. (1991) en Kluger en DeNisi (1996) zijn twee overzichtsstudies naar feedback vanuit de cognitieve benadering, en Shute (2008) past ook in deze traditie.
Bangert-Drowns et al. (1991) geven aan dat in onderwijssituaties feedback meestal doelbewust wordt gegeven en kan zijn gericht op de correctheid van de uitvoering en het resultaat van de taak, op de affectieve beleving en motivatie van de leerling, en op metacognitieve aspecten. Hun studie betrof kennistoetssituaties, met diverse taken (het goede antwoord kunnen geven, kennis kunnen toepassen) en vragen (multiple choice, aanvul- of korte antwoordvragen). Zij vonden zwakke positieve effecten van feedback en in een derde van de onderzoeken was het effect negatief. Feedback die louter bestond uit informatie over goed of fout had gemiddeld vrijwel geen effect, feedback waarbij het goede antwoord werd genoemd of begeleiding werd gegeven had wel een positief effect. De voornaamste functie van feedback bij dit type taken is het corrigeren van fouten, en de omvang van de feedback is daarbij niet van belang, aldus Bangert-Drowns et al. Volgens hen kan feedback het leren bevorderen als deze "*mindfully*" wordt ontvangen: de ontvanger moet willen weten *waarom* iets goed of fout is.

Kluger en DeNisi (1996) namen in hun studie voornamelijk onderzoeken op naar laboratoriumtaken (uitvoeren van een handeling, onthouden van informatie, reageren op stimuli) en voor ruim de helft taken die binnen een kort tijdsbestek moesten worden uitgevoerd. De meeste feedback betrof de prestatie op de taak (goed/fout of score), en niet het leren van de taak (bijvoorbeeld door feedback te geven op de aanpak). Kenmerken van feedback die samenhingen met positieve effecten waren bijvoorbeeld het geven van de juiste oplossing, het nalaten van prijzen, en het stimuleren van het stellen van doelen. Kenmerken die samenhingen met negatieve effecten hadden vooral te maken met bedreiging van de zelfachting (*self-esteem*) en ontmoediging.
Kluger en DeNisi presenteren een Feedback Intervention Theory (FIT) met als centrale hypothesen dat mensen meerdere doelen nastreven die liggen op verschillende hiërarchisch geordende niveaus, en dat feedback als belangrijk effect heeft dat het de aandacht richt op een bepaald niveau. Ze onderscheiden drie niveaus: het (leren) uitvoeren van de taak, de motivatie voor de taak, en de eigen persoon. Het is voor mensen belangrijk hun persoon te beschermen, en feedback wordt dan ook vaak goed opgemerkt, omdat deze daarvoor bedreigend kan zijn.
Kluger en DeNisi geven aan dat positieve effecten van specifieke en procesgerichte feedback kunnen worden begrepen doordat feedback de aandacht richt op de taak. Het is echter gemakkelijker om van een ander feedback te krijgen dan om deze zelf te genereren. Zulke feedback kan leiden tot meer leren, maar ook tot minder leren, namelijk als de lerende zich erop gaat richten om meer van zulke (leerzame) feedback van een ander te krijgen. Goede feedback kan zo ook leiden tot passiviteit. Hun theorie biedt ook een verklaring voor het feit dat geven van een cijfer vaak geen goede vorm van feedback is, want een cijfer appelleert aan vergelijking met andere leerlingen, richt de aandacht daardoor gemakkelijk op de eigen persoon en leidt daarmee af van de taak. Ook het negatieve effect van feedback die door de ontvanger wordt ervaren als sterk controlerend, is met behulp van de theorie te duiden, namelijk door aan te nemen dat zulke feedback de intrinsieke motivatie verlaagt.
Kluger en DeNisi gaan ook in op de manieren waarop ontvangers op feedback kunnen reageren. Zij kunnen hun gedrag veranderen (harder werken en/of hun aanpak aanpassen), hun persoonlijke standaard veranderen (de lat voor zichzelf lager, of juist hoger leggen), de feedback verwerpen (bijvoorbeeld als niet geloofwaardig of niet op hen van toepassing), en zich helemaal afkeren van de taak (hun committent staken).
Shute (2008) richtte haar studie op experimentele onderzoeken naar vormen van feedback, en daarmee op feedbackvormen die goed experimenteel manipuleerbaar zijn. Kenmerken van feedback die over het algemeen een negatief effect hebben, zijn: feedback die (te) kritisch of controlerend is, feedback in de vorm van cijfers, feedback die weinig specifiek is (vage feedback), feedback die (te) veel informatie bevat, en feedback die wordt gegeven terwijl de leerling bezig is om de taak uit te voeren (het probleem op te lossen) en daarbij door de feedback wordt gestoord. Shute voegt aan dit laatste toe dat het bij taken die bestaan uit het (leren) uitvoeren van bepaalde procedures op korte termijn effectief kan zijn om direct feedback te geven, maar op

langere termijn juist minder effectief, omdat de transfer (toepassing in andere situaties) dan geringer is. (Mogelijk draagt het niet meteen feedback geven ertoe bij dat lerenden leren om problemen die zich voordoen zelf op te lossen en dit daardoor ook in nieuwe situaties beter kunnen). Shute oppert dat het niet direct geven van feedback ook verschillende motivationele effecten kan hebben: het kan als frustrerend worden ervaren, maar ook als uiting van respect voor de eigen autonomie.
Vormen van feedback die blijkens de studie van Shute vaak een positief effect hebben, zijn: meer specifieke informatie geven dan alleen goed/fout, het goede antwoord geven, en aangeven *waarom* iets goed of fout is. Als verklaringen voor deze positieve effecten noemt Shute enkele cognitieve mechanismen. Door informatie te geven over de afstand tussen prestatie en standaard wordt de onzekerheid over hoe goed je het doet, verminderd, en daarmee ook de cognitieve belasting. Verder kan specifieke informatie nuttig zijn voor het corrigeren van misconcepties, procedurele fouten en ongeschikte aanpakken. De overige voorbeelden van interventies die volgens Shute vaak positieve effecten hebben, verbreden de scope van feedback in engere zin naar coaching en tutoring: de aandacht richten op essentiële aspecten van een taak; uitleggen van belangrijke begrippen; geven van hints, cues en prompts; het probleem vereenvoudigen; geven van aanpaksuggesties; voordoen (modelleren); en geven van aanvullende instructie.

Feedback als proces van communicatie en interactie
In de hiervoor beschreven behavioristische en cognitief-psychologische benaderingen werd feedback onderzocht als eenmalige actie en daarmee als eenrichtingverkeer. Door onderzoekers vanuit de sociale psychologie is feedback benaderd in termen van communicatie en interactie.
Ilgen, Fisher en Taylor (1979) stellen dat feedback over de effectiviteit van iemands gedrag al heel lang is onderkend als essentieel voor leren en motivatie in "*performance-oriented organizations*". Zij stellen dat al veel literatuur beschikbaar is, maar nog weinig algemene kennis is verkregen over de effecten van feedback. Zij schrijven dit toe aan het feit dat feedback zelf al een gecompliceerde stimulus is en daarnaast ook nog met veel andere factoren samenhangt, waardoor het bijna onmogelijk is de effecten van gegeven feedback op zich vast te stellen.
De auteurs stellen voor feedback te zien als onderdeel van de interpersoonlijke interacties die nodig zijn voor leren. Ze benaderen feedback als speciaal geval van het algemene proces van communicatie waarbij een zender een boodschap overbrengt aan een ontvanger. Het speciale van feedback is dat de boodschap informatie bevat over de ontvanger. De waarneming en acceptatie van de feedback door de ontvanger en de reactie erop worden beïnvloed door kenmerken van de zender, de boodschap en de ontvanger zelf.
Het belangrijkste kenmerk van de *zender* is diens geloofwaardigheid (in de ogen van de ontvanger, in termen van onder andere deskundigheid en betrouwbaarheid), met name voor de acceptatie van de boodschap.

Belangrijk bij de *boodschap* is dat deze voor de ontvanger betekenisvolle en nieuwe informatie bevat over eerder gedrag (presteren), die van belang is voor het leerproces (als de boodschap informatie bevat over de uitvoering van de taak) en voor de motivatie voor de taak (bij informatie over de inzet). Kenmerken van de boodschap die hieraan kunnen bijdragen zijn de timing, de mate waarin de boodschap positief of negatief is, de frequentie waarmee feedback wordt gegeven, de mate waarin de feedback wordt ervaren als controlerend, en de consistentie en specificiteit van de boodschap. Ook is van belang de mate waarin de boodschap bevordert dat de ontvanger verwacht de prestatie te kunnen verbeteren en dan positieve feedback te zullen krijgen. Ten aanzien van de eerste drie kenmerken (timing, balans positief/negatief, frequentie) zijn de onderzoeksresultaten niet eenduidig. Deze kenmerken zijn wel van belang, maar er kan niet in het algemeen worden aangegeven wat qua timing, balans tussen positief en negatief, en frequentie, optimaal is.

Belangrijke kenmerken van de *ontvanger* van feedback hebben betrekking op zijn motivatie voor het maken van de taak. Het begrip motivatie is in de literatuur op verschillende manieren uitgewerkt. Twee belangrijke theorieën over motivatie zijn die over *self-efficacy* (Bandura, 1986) en over motivatie in relatie tot de basisbehoeften relatie, competentie en autonomie (Deci & Ryan, 2000). Deci en Ryan onderscheiden verschillende typen motivatie (intrinsieke motivatie, geïdentificeerde regulatie, externe regulatie en amotivatie) op een continuüm van de mate waarin iemand het besef heeft het verloop van een actie of taak zelf te kunnen beïnvloeden (self-determination). Bij het ontvangen van feedback is van belang in welke mate waarin de ontvanger erop vertrouwt dat zijn of haar inzet tot resultaat zal leiden (de *self-efficacy*), intrinsiek gemotiveerd is (gevoed door ervaringen van competentie en autonomie), en de taak ervaart als specifiek, uitdagend, en niet opgedrongen. Ook hier zijn beide functies van feedback te herkennen: inzet bevorderen en richting geven.

Ook Larson (1984) benadert feedback als proces en hij voegt aan de door Ilgen et al. (1979) genoemde functies van feedback toe het geven van informatie over de mate waarin de eigen doelen worden bereikt en voor sociale vergelijking (waar sta ik in vergelijking met anderen). Volgens hem ontbreken bij veel taken objectieve prestatiecriteria en zijn anderen daardoor de voornaamste bron van feedback. Het geven van feedback is echter een complexe handeling. De prestatie waarop feedback moet worden gegeven, is niet altijd duidelijk. Verder vallen lage prestaties vaak meer op dan hoge, waardoor de neiging ontstaat om weinig positieve feedback te geven, maar zijn mensen ter wille van de onderlinge relatie juist terughoudend om negatieve feedback te geven. Ook speelt een rol hoe goed de feedbackgever de ontvanger kent en daardoor de gevolgen van de feedback kan inschatten.

Het gezichtspunt van de invloed van geven van feedback en de reacties daarop op de feedbackgever is nog weinig onderzocht. Reynolds, Martin en Groulx (1996) onderzochten de patronen van zichtbaar gedrag bij feedbackontvangers (basisschool-leerlingen) die voor de feedbackgevers (leerkrachten) aanleiding zijn om te concluderen dat zij iets hebben begrepen (en de interactie dus kan worden afgerond).

Zij vonden: veranderingen in gezichtsuitdrukking en houding, een begrip toepassen in een nieuwe situatie, spontaan zelf dingen aanpassen (en niet zomaar overnemen), een procedure toepassen in een nieuwe situatie, met vertrouwen verkortingen in procedures hanteren, iets aan een ander kunnen uitleggen, actief en gericht aan de slag gaan (de eigen aandacht richten).
In aanvulling op de studie van Crooks (1988) kwamen Black en Wiliam (1998) met een nieuw overzicht van het voorkomen en de effecten van processen van formatieve beoordeling door leraren. Black en Wiliam benadrukken het belang van open en gevarieerde taken, die bijdragen aan het gewenste leren en zicht geven op het presteren en de ontwikkeling van leerlingen. Waar het vervolgens om gaat is de kwaliteit van de interactie tussen leraar en leerlingen. De manier waarop de leerlingen de feedback ervaren, zal mede afhangen van hun doeloriëntatie (zijn ze gericht op leren, of meer op presteren), de mate waarin de feedback appelleert aan de taak dan wel de persoon van de leerling, de motivatie van de leerling (zijn *self-efficacy* en attributies), zijn opvattingen over leren, en zijn prestatieniveau, aldus Black en Wiliam.
Torrance en Pryor (1998) deden onderzoek naar feedback-interacties tussen leerkrachten en leerlingen in basisscholen. De leerling moet de feedback die hij krijgt, zien te duiden. Dat is volgens de auteurs niet eenvoudig, omdat leerkrachten zich vaak niet duidelijk uitdrukken. De interactie heeft bovendien meerdere lagen: die van de taal, de onderlinge relatie (inclusief de verdeling van macht), en de inhoud (de constructie van kennis). Bij de leerling is bovendien sprake van affectief-motivationele processen. Torrance en Pryor vonden de volgende docentgedragingen rond het geven van feedback: observeren van leerlingen die aan het werk zijn, bekijken van gemaakt werk, vragen stellen naar begrip, vragen wat een leerling doet of heeft gedaan, vragen waarom een leerling iets (op die manier) deed, aangeven van criteria waaraan de taak moet voldoen, aangeven van standaarden, kritiek uiten op een aspect van het werk, informatie geven, corrigeren, suggereren van een alternatief, evaluatieve feedback geven op de taak, de aanpak, de inzet etc., en met de leerling bespreken hoe deze verder kan of moet gaan.

Feedback als onderdeel van coaching
Volgens Torrance en Pryor (1998) kijkt feedback niet alleen terug, maar ook vooruit, namelijk naar wat de leerling nu kan gaan begrijpen. Daardoor wordt de leerling betrokken in het proces en komt een vorm van samenwerking op gang. Effectieve feedback legt volgens hen minder nadruk op de geleverde prestatie maar op wat de leerling heeft begrepen en waar kansen liggen voor verder leren. Zij pleiten voor het integreren van het individuele leerproces in de sociale situatie en het stimuleren van de leerling om te vertellen over zijn eigen denken.
De grens tussen louter feedback geven en coachen is niet scherp te trekken (Askew & Lodge, 2000; Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; Carnell, 2000; Graesser, Person & Magliano, 1995; King, 2002; Mory, 2004; Wittwer, Nickles & Renkl, 2006). Carnell (2000) stelt dat een echte “leerdialoog” vergt dat voldoende tijd

beschikbaar is en dat de neiging om leerlingen te controleren achterwege wordt gelaten. Mory (2004) noemt feedback een natuurlijk resultaat van interacties tussen de lerende, zijn denkconstructies en de leeromgeving, en feedback is ook volgens hem in deze visie een onderdeel van coaching. Deze constructivistische visie is door Collins, Brown en Newman (1989) uitgewerkt in hun model van *cognitive apprenticeship*. In dit model zijn belangrijke taken van docenten bij coaching: het voordoen (van het benodigde of gewenste proces van nadenken en probleemoplossen), het geven van hints, aanpaksuggesties, hulp en feedback (inclusief uitleg, verklaringen en voorbeelden), en het stellen van vragen (mede om leerlingen te ondersteunen hun denken te articuleren). Belangrijk is ook dat de docent zijn of haar hulp geleidelijk vermindert, want leerlingen moeten leren het zelf te doen.

Feedback als "boodschap" en de rol van betekenisgeving

Bij mondelinge feedback in een ruimte met leerlingen (een klas, domein of leerplein) zullen vooral beide laatste besproken benaderingen van toepassing zijn: feedback als onderdeel van communicatie en van coaching. Dit neemt niet weg dat elk moment dat een docent feedback geeft, de docent kan worden gezien als feedbackgever en de leerling als feedbackontvanger. De docent zal met de feedback een bepaalde bedoeling hebben en een boodschap met een bepaalde betekenis willen geven, en de leerling zal deze boodschap (moeten) interpreteren.
Bij het geven van betekenis door een leerling aan verkregen feedback kunnen ook kenmerken van de taak, de prestatie, de feedbackgever en de context invloed uitoefenen. De invloed daarvan kan op twee manieren worden geanalyseerd: aan de hand van kenmerken zoals die objectief kunnen worden vastgesteld, en met behulp van kenmerken zoals die (subjectief) worden waargenomen en ervaren door de ontvanger. Sommige kenmerken kennen beide gedaanten, objectief en subjectief. Dit is bijvoorbeeld het geval met kenmerken van de taak, bijvoorbeeld of slechts één antwoord of oplossing goed is of meerdere, en of vooraf wel of niet expliciete criteria en standaarden zijn gegeven. Zulke feitelijke kenmerken kunnen vrij eenvoudig worden bepaald, maar kunnen wel door leerlingen verschillend worden waargenomen en ervaren en ook op die manier worden onderzocht.
Een vergelijkbare variatie doet zich voor bij de prestatie (deels objectief: aantal goed beantwoorde vragen, omvang en volledigheid van een werkstuk, deels subjectief: het viel mee of niet, de leerling verwacht een goede beoordeling of niet), de feedbackgever (objectief: de docent heeft de leerling wel of niet aan het werk gezien; subjectief: de docent is in de ogen van de leerling als feedbackgever meer of minder geloofwaardig), en de context (objectief en subjectief: is er na ontvangst van de feedback wel of niet gelegenheid om er vragen over te stellen of om de feedback te benutten).
Bij het interpreteren van feedback zullen dergelijke kenmerken van de taak, de prestatie, de feedbackgever en de context in de ogen van de ontvanger vaak min of meer één geheel vormen met de feedback zelf. Dit zal vrijwel zeker het geval zijn met

kenmerken van de feedbackgever, omdat een feedbackontvanger bij het geven van betekenis aan de boodschap in ieder geval de ingeschatte geloofwaarheid en bedoelingen van de feedbackgever zal betrekken (Ilgen et al., 1979). Deze verwevenheid van de percepties van de feedbackgever en de feedbackboodschap doet zich bijvoorbeeld voor als de feedbackgever de indruk wekt controle te willen uitoefenen op de feedbackontvanger, in plaats van deze de ruimte te geven om zelf te bedenken hoe hij of zij de prestatie kan verbeteren (Straub, 1996, 1997).
De impact van ontvangen feedback zal ook afhangen van kenmerken van de ontvanger, vooral in relatie tot de eerder genoemde cognitieve en affectief-motivationele processen. Ilgen en Davis (2000) vestigen in dit verband de aandacht op de mogelijke effecten van feedback die wijst op een tekortschietende prestatie. Zij stellen dat een prestatie van een leerling op een taak, of het nu gaat om de aanpak en uitvoering of om het resultaat, nu eenmaal wel of niet aan bepaalde standaarden voldoet, en dat in het laatste geval, als de prestatie (nog) onvoldoende is, de leerling zich daarvan bewust moet worden om zich te kunnen verbeteren. Anders gezegd: als het niet goed is, moet je zeggen dat het niet goed is. Dat vinden mensen meestal niet leuk, en docenten maken daarop geen uitzondering. Zij kunnen daardoor geneigd zijn de negatieve boodschap af te zwakken, vaag te houden of in te kleden. De ontvanger, de leerling, vindt het ook niet prettig om een negatieve boodschap over zijn prestatie te krijgen, want die wordt gemakkelijk ervaren als pijnlijk en bedreigend voor het zelfconcept, de zelfachting en het zelfvertrouwen. Het is dus van belang hoe feedback wordt ingepast in de 'affectieve boekhouding' van de feedbackontvanger. Een feedbackgever zal, indien een prestatie niet aan de standaard voldoet, een balans moeten zien te vinden tussen enerzijds de ontvanger confronteren met dit negatieve oordeel en deze zelf verantwoordelijkheid laten nemen voor het nog onvoldoende presteren, en anderzijds de ontvanger behoeden voor te sterke aantasting van het zelfvertrouwen.

Mogelijk relevante kenmerken van feedback

In het voorgaande is, verspreid over de besproken benaderingen, een aantal kenmerken van feedback aan de orde geweest. Van een deel daarvan zijn uit onderzoeken, waarin feedback is opgevat als (eenmalige) informatie, (gemiddeld) positieve effecten bekend, meestal afgemeten aan de prestatie op een volgende, vergelijkbare taak. Daarbij gaat het echter grotendeels om kenmerken en taken die relatief gemakkelijk onderzoekbaar zijn: kenmerken die goed experimenteel manipuleerbaar zijn en taken die in korte tijd uitvoerbaar zijn en eenduidig met goed of fout beoordeelbaar zijn.
Over de effecten van een aantal andere kenmerken zijn nog weinig of geen resultaten uit onderzoek bekend, maar wordt in de literatuur wel aannemelijk gemaakt dat van feedback met zulke kenmerken positieve effecten kunnen worden verwacht. Daarbij gaat het dan niet alleen of niet zozeer om verwachte effecten op een volgende prestatie, maar vooral ook om effecten op denkstrategieën, motivatie en zelfregulatie. In zulke

publicaties wordt feedback meestal opgevat als proces, en gaat het minder om terugkoppeling van de prestatie als wel om feedback gericht op leren.
Bij een aantal kenmerken is bekend of lijkt aannemelijk dat zij niet eenduidig goed zijn, in de zin van: hoe meer, hoe beter, en zelfs niet in de zin van: baat het niet, schaadt het niet, maar dat het belangrijk is om ze te doseren. Goede feedback is dan mede een kwestie van maat en balans. Tot slot is een aantal kenmerken meer of minder vatbaar voor het ontstaan van misverstanden, waarbij de feedbackgever iets anders bedoelt of beoogt dan de ontvanger begrijpt. Dit kan zich voordoen doordat formuleringen voor meer interpretaties vatbaar zijn, maar bijvoorbeeld ook doordat een leerling niet beschikt over bepaalde voorkennis (waar de docent op dat moment wel vanuit ging) of affectief reageert op een manier die de docent niet heeft voorzien.
Figuur 5.1 bevat een overzicht van alle hiervoor genoemde kenmerken, inhouden en vormen van feedback, met daarbij aangegeven in hoeverre deze volgens de bestudeerde literatuur naar verwachting bijdragen aan het beoogde onderwijsleerproces.

Onderscheiden kenmerken, inhouden en vormen van feedback	Naar verwachting bijdragend aan het beoogde onderwijsleerproces
Context en timing	
a. Eenmalige reactie of onderdeel (mogelijkheid) van een interactieproces	a Interactie maakt afstemming en verduidelijking mogelijk
b. In aanwezigheid van anderen of individueel (persoonlijk)	b Bij voorkeur individueel
c. Timing	c Niet te snel, niet te laat
Aandacht voor de taak	
1. Geven van een cijfer en/of informatie over goed/fout en/of een beschrijving	1 Informatie/beschrijving zonder cijfer
2. Over de persoon en/of over de taak (uitvoering en resultaat)	2 Bij voorkeur over de taak
Algemene kenmerken van de informatie (cognitieve belasting)	
3. Hoeveelheid of omvang	3 Niet te veel of te omvangrijk
4. Duidelijkheid en consistentie (bij herhaalde feedback)	4 Deze kenmerken dragen positief bij
5. Specificiteit en concreetheid	5 Deze kenmerken dragen positief bij
Toonzetting (affectieve lading)	
6. Vriendelijkheid	6 Dit draagt positief bij
7. Respect in plaats van controle	7 Bij voorkeur niet controlerend
8. Balans tussen positief (ondersteunend) en negatief (kritiserend)	8 Voldoende positief maar ook benoemend wat beter kan/moet
Aandacht voor de standaard	
9. Geven van het goede antwoord c.q. benoemen van de standaard	9 Draagt meestal positief bij
10. Vergelijken van de prestatie met de standaard	10 Draagt meestal positief bij
11. Vergelijken van de prestatie met een vorige prestatie	11 Kan positief bijdragen indien sprake is van progressie
Feedback als mogelijke aanzet tot voortgezette interactie en eventuele zelfsturing	
12. Geven van nadere uitleg en voorbeelden	12-20 Deze vormen van interactie kunnen alle positief bijdragen, afhankelijk van de taak, de leerling, de context en de prestatie in kwestie en de mate waarin de feedback daarop wordt afgestemd
13. Leggen van de relatie tussen aanpak en resultaat	
14. Het probleem vereenvoudigen en/of in stappen opdelen	
15. Een denkproces of handeling voordoen (modelleren)	
16. Geven van aanpaksuggesties	
17. Geven van hints, cues en prompts	
18. Geven van aanvullende instructie	
19. Stellen van open vragen, bevorderen van reflectie en articulatie	
20. Geven van suggesties voor monitoring en evaluatie en stellen van doelen	

Figuur 5.1 Kenmerken, inhouden en vormen van feedback en hun bijdrage aan het onderwijsleerproces

5.3 Opzet en uitvoering van het onderzoek

Onderzoeksopzet

Het onderzoek is uitgevoerd op vier scholen. Eerst is een startgesprek gevoerd met de schoolleiders gezamenlijk (over de opzet van het onderzoek en de keuze van vakken, leerjaren en docenten) en is door hen individueel een concept map over feedback gemaakt, en daarna is per school een gezamenlijk startgesprek gehouden met de op die school deelnemende docenten (over de opzet en planning van het onderzoek en de keuze van de groep leerlingen) en is door hen individueel een concept map over feedback gemaakt. Vervolgens is door elke docent en door een aantal leerlingen bij deze docent een vragenlijst ingevuld, over onder andere het voor de docent gebruikelijke feedbackgedrag. Daarna is per docent een video-opname gemaakt van een lesuur waarin de leerlingen zelfstandig aan het werk waren en de docent langsliep en feedback gaf, en is aan het einde van het lesuur door de leerlingen een tweede vragenlijst ingevuld, over de door hen in dat uur ontvangen feedback. Na selectie, observatie en scoring per docent van een aantal videofragmenten vond met elke docent een stimulated recall interview plaats. Ten slotte zijn per docent in een eindgesprek alle onderzoeksresultaten aan hem of haar teruggekoppeld en aan de hand daarvan enkele vragen en aandachtspunten besproken.

Respondenten

Binnen de mogelijkheden van tijd en budget konden op vier scholen in totaal 12 docenten meedoen. Drie scholen werken met leerpleinen of domeinen waarin een grotere groep leerlingen door twee tot vier docenten worden begeleid. Op één school deed een vernieuwende afdeling onderbouw havo-vwo mee. Twee scholen hebben leerlingen van vmbo-t t/m vwo. De vierde school is een vmbo-school met verschillende niveaus (basis, kader, theoretisch). Alle scholen richten zich naast het vergroten van de kennisbasis van leerlingen ook op het aanleren van vaardigheden (plannen, samenwerken, kiezen, reflecteren, presenteren) en het stimuleren van de zelfstandigheid van leerlingen.
Gekozen werd voor deelname van drie docenten per school, gespreid over de alpha-, beta- en gammavakken, omdat de aard van het vak en de opdrachten daarbinnen van invloed kunnen zijn op de manier waarop docenten feedback geven. Verder is gekozen voor de leerjaren 3 en/of 4, omdat alle deelnemende scholen over deze leerjaren beschikken en de determinatie van leerlingen in die leerjaren min of meer is afgerond. Tot slot is besloten zowel meer als minder ervaren docenten bij het onderzoek te betrekken, omdat vaak wordt gedacht dat meer ervaren docenten over een breder of effectiever feedbackrepertoire beschikken, maar het de vraag is of dat wel klopt.
Elke schoolleider heeft op basis van de genoemde criteria drie docenten geworven die belang stelden in deelname. Alle docenten deden op basis van vrijwilligheid mee. In de

per docent gevoerde eindgesprekken hebben zij teruggekeken op hun motieven en verwachtingen ten aanzien van dit onderzoek. Ze gaven aan vooral te hebben meegedaan uit nieuwsgierigheid, vanuit de wens er iets van te leren, en vanuit de wens te willen weten hoe hun feedback door de leerlingen wordt ervaren.
De docenten kozen zelf de groep leerlingen die aan het onderzoek zouden meedoen, vooral met het oog op de inpassing van de te maken video-opname in het rooster. Door de gekozen groep leerlingen per docent is op zijn verzoek een vragenlijst ingevuld, een tot twee weken voor het maken van de video-opname. Omdat op voorhand niet bekend was welke leerlingen tijdens de opname feedback zouden krijgen, is de vragenlijst voor de leerlingen ingevuld door de hele betreffende groep leerlingen. Bij één docent (docent 11) is door omstandigheden de vragenlijst afgenomen bij een andere groep leerlingen dan waarbij de opname is gemaakt.
Zie tabel 5.1 voor een aantal kenmerken van docenten, video-opnames en deelnemende leerlingen.

Tabel 5.1: Kenmerken docenten en leerlingen, aantal en duur opgenomen fragmenten en respons vragenlijsten leerlingen vooraf en direct na afloop

School	*Docent*	*m/v*	*Vak*	*onderwijs ervaring totaal*	*Aantal jaar op deze school*	*Schooltype deelnemende leerlingen*	*Leerjaar*	*N vragenlijst vooraf*	*Aantal opgenomen fragmenten*	*Totale tijdsduur fragmenten in minuten*	*N vragenlijst na afloop*
A	1	M	Wiskunde	weinig	2	Havo	4	19	10	22	18
	2	M	Geschiedenis	gemiddeld	9	Havo	4	17	7	8	18
	3	V	Engels	veel	15	Havo	3	22	8	19	21
B	4	M	Latijn	veel	16	Vwo	3	48	7	24	28
	5	M	Biologie	gemiddeld	11	Havo	4	22	6	22	10
	6	M	Economie	weinig	5	Vmbo-t	4	13	8	17	10
C	7	V	Frans	weinig	1	Havo/vwo	3	15	4	18	5
	8	V	Wiskunde	veel	2	Havo/vwo	3	42	5	27	5
	9	V	Techniek	gemiddeld	3	Havo/vwo	3	9	7	15	15
D	10	M	Economie	veel	1	Vmbo	3	20	7	21	4
	11	V	Economie	weinig	2	Vmbo	4	23	5	28	9
	12	V	Engels	gemiddeld	3	Vmbo	3	74	5	22	6

* Mate van ervaring: weinig: ≤ 5 jaar; gemiddeld: > 5 en ≤ 15 jaar; veel: > 15 jaar.

Instrumenten en procedures

Conceptmap schoolleiders en docenten
De schoolleiders hebben in hun gezamenlijke startbijeenkomst individueel een conceptmap gemaakt over het begrip feedback, aan de hand van een grotendeels open opdracht, waarin hen werd gevraagd uit te gaan van drie vragen waarbij ze dienden te bedenken, opschrijven en ordenen welke aspecten volgens hen bij het geven van feedback een rol (dienen te) spelen. De vragen waren: Hoe wil je dat docenten

feedback geven? Hoe zie je dat handelen in relatie tot het schoolconcept? Welke factoren spelen een rol bij het geven van feedback door docenten aan leerlingen?
De docenten hebben, in een gezamenlijke startbijeenkomst per school, een conceptmap gemaakt over het begrip (geven van) feedback, aan de hand van een geheel open opdracht waarbij ze moesten bedenken, opschrijven en ordenen welke aspecten volgens henzelf bij het geven van feedback een rol spelen. Ook werd hen gevraagd relaties tussen de verschillende aspecten aan te brengen.

Vragenlijst docenten
De docenten hebben vervolgens een vragenlijst ingevuld die bestond uit drie delen, een algemeen gedeelte, een gedeelte over de eigen visie en opvattingen, en een schaal met uitspraken over het taakgebied feedback met twee daarbinnen te onderscheiden subtaakgebieden uit een instrument dat eerder is ontwikkeld (zie hoofdstuk 3).
Het algemene gedeelte bevatte open vragen over hun werksituatie (vak, werkzaam in welke afdelingen en leerjaren, aantal jaren ervaring in het onderwijs en op deze school, andere taken, werkdagen) en voorgestructureerde vragen over de verstrekte informatie over het onderzoek, de motieven om mee te doen, hun verwachtingen van de deelname, een eventuele eigen leervraag of aandachtspunt, de wijze waarop tot nu toe de eigen deskundigheid over feedback is opgebouwd, en de vraag of er tussen docenten onderling wordt gesproken over het geven van feedback.
Het gedeelte over de eigen visie en opvattingen bevatte open vragen over de functie van feedback, de doelen die de docent daarmee nastreeft, de aspecten waarop de docent zich richt, de aspecten bij leerlingen waarmee de docent rekening houdt, wat de docent belangrijk vindt bij de timing in het geven van feedback, waarop de docent let in de bewoordingen naar een leerling, wat een docent na het geven van feedback belangrijk vindt, en op wiens initiatief feedback gegeven wordt. Ook werd gevraagd naar voorbeeldsituaties en naar aandachtspunten bij geven van feedback op deze school.

Vragenlijst leerlingen vooraf
De vragenlijst bevatte enkele algemene vragen (geslacht, leeftijd, groep, afdeling), vragen over hun motivatie voor het betreffende vak, de schaal over het taakgebied Geven van feedback met daarbinnen twee subschalen uit eerder onderzoek (zie hoofdstuk 3) en een aanvullend gemaakte schaal over Feedback geven, met daarbinnen twee subschalen over procesgerichte en resultaatgerichte feedback.
Voor het meten van de self-efficacy is de betreffende schaal van de Motivated Strategies for Learning Questionnaire gebruikt (MSLQ; Garcia & Pintrich, 1996), in de vertaling van Severiens (2005). Deze schaal bestaat uit acht uitspraken met daarbij zeven antwoordmogelijkheden, van ‘helemaal niet waar voor mij’ tot ‘helemaal waar voor mij’. Voor het meten van de eerder genoemde vier vormen van motivatie zoals onderscheiden door Deci en Ryan (2000) zijn de betreffende schalen van de SIMS gebruikt (Situational Motivation Scale; Guay, Vallarand & Blanchard, 2000). De antwoord-mogelijkheden op deze items varieerden van “komt helemaal niet overeen”

met waarvoor ik me inspan voor de opdrachten bij dit vak (of leerdomein) bij deze leraar tot "komt helemaal overeen", op eveneens een 7-puntsschaal.
De aanvullende items over Feedback geven zijn gebaseerd op de literatuur in 5.2 (onder andere Kluger & DeNisi 1996; Sadler, 1989; Shute 2008). Net als bij de schaal over het taakgebied (uit hoofdstuk 3) is ook hierbij gevraagd naar zowel het huidige als het gewenste handelen ("past dit bij wat de leraar doet?" en "wil je dat de leraar dit doet"), met vijf antwoordmogelijkheden (van "helemaal niet" tot "helemaal wel "). In 22 aanvullende items zijn de volgende aspecten uitgewerkt:
a. Het focus: nadruk op leren (aanpak-proces-doel; items 6, 10, 17, 20) of op het resultaat (7, 9 en 13)
b. De norm: gerelateerd aan een standaard (items 1, 2, 15 en 18), anderen (11), of de leerling zelf (21)
c. De aansluiting bij het prestatieniveau, de capaciteiten en de inspanning van de leerling (item 16)
d. De nadruk op of iets goed is gedaan (items 3 en 19) of fout is gedaan (items 5 en 12)
e. De mate waarin de feedback gevraagd of ongevraagd wordt gegeven (items 14 en 22)
f. De mate waarin het nut van de opdracht wordt verduidelijkt (items 4 en 8)

Video-opname per docent
Per docent is één opname gemaakt van een lesuur (c.q. domeinuur). De duur ervan varieert van 45 tot 60 minuten, soms werd voorafgaand aan het zelfstandig werken eerst nog enige instructie gegeven. De opnamen zijn gemaakt met een digitale videocamera met zoomlens waarbij de docent een draadloze zendmicrofoon bij zich droeg zodat ook het geluid goed kon worden opgenomen en de onderzoeker op zodanige afstand kon blijven dat de interactie tussen docent en leerlingen niet werd verstoord.

Vragenlijst leerlingen direct na de opname
Direct aansluitend aan de opname is aan de leerlingen die feedback hebben gekregen een vragenlijst uitgereikt. Hen werd gevraagd aan welke opdracht ze werkten, wat ze precies moesten doen, wat goed ging, of ze ergens moeite mee hadden en zo ja, waarmee; of ze een vraag aan de docent hebben gesteld en zo ja, welke; of de docent feedback heeft gegeven, en zo ja, wat de docent heeft gezegd.
Daarna moesten de leerlingen aangeven aan de hand van 20 uitspraken hoe ze de feedback, indien verkregen, hadden ervaren. Ze konden hierbij antwoorden op een 5-puntsschaal van "helemaal niet" tot "helemaal wel". De uitspraken zijn gebaseerd op een procesgeoriënteerde benadering van feedback, waarbij aandacht wordt besteed aan zowel de manier waarop de zender de boodschap brengt als de manier waarop de ontvanger de boodschap percipieert (Ilgen, Fisher, & Taylor, 1979; Shute, 2008). De uitspraken bevatten diverse elementen in het feedbackproces: de perceptie, en de acceptatie en bruikbaarheid van de feedback. Ook is gevraagd waarop de feedback zicht richt (het object). Bij de uitspraken gericht op de perceptie van de feedback is gevraagd of de feedback specifiek was (item 5), faciliterend (18, 19 en 20), doelgericht

(gekoppeld aan einddoelen) (14), en evaluatief (3 en 4) en of de feedback positief of negatief was geformuleerd (6 en 7). Bij de uitspraken gericht op de acceptatie en bruikbaarheid van de feedback is gevraagd of de feedback bruikbaar was (1 en 17), leerzaam (16), terecht (2) en motiverend (15). Zes uitspraken betroffen het object van de feedback (8 t/m 13).

Selectie van opgenomen fragmenten

Ten behoeve van de te houden stimulated recall interviews zijn uit de opnamen fragmenten geselecteerd waarin sprake was van een reactie van de docent op het werk of een vraag van een leerling tijdens het zelfstandig werken. Hierbij konden zich vier typen situaties voordoen.

- *Situaties waarin de docent de interactie start*: 1. De docent loopt rond en start de interactie door feedback te geven op waarneembaar handelen van een of meer leerlingen. 2. De docent loopt rond en start de interactie door het stellen van een vraag. Op het moment dat de leerlingen reageren ontstaat interactie waarin de docent de leerlingen feedback kan geven op wat ze zeggen en/of op waar ze mee bezig zijn (dat wil zeggen: de manier waarop ze bezig zijn en/of het product waaraan ze werken)
- *Situaties waarin leerlingen de interactie starten*: 3. De docent loopt rond en een leerling of groepje leerlingen start de interactie door een gerichte vraag te stellen over een bepaald probleem (ze snappen iets niet, weten niet hoe iets aan te pakken). De docent kan dan feedback geven op wat is benoemd of gedaan. 4. De docent loopt rond en een leerling of groepje leerlingen start de interactie door de docent te vragen wat hij of zij van iets vindt (een open vraag over bijvoorbeeld een product in wording).

Bij de selectie van de fragmenten werd het fragment gestart op het moment dat de interactie tussen de docent en de leerling(en) begon, hoewel dat niet betekende dat direct al sprake was van feedback. Aangezien het in dit onderzoek gaat om het didactisch handelen is de feedback die uitsluitend gericht was op het gedrag of de werkhouding bij de selectie van de fragmenten buiten beschouwing gelaten.

Bij de selectie van de fragmenten is een spreiding aangebracht over verschillende manieren van reageren van de docent, verschillende situaties en de eventueel verschillende lagen in de opdracht (bijvoorbeeld bij wiskunde: het gebruik van een rekenmachine, de aanpak van een som, het begrip, het leerproces). Waar zich dezelfde soorten situaties, vragen en problemen, en typen gegeven feedback voordeden, is slecht eenmaal dergelijk fragment geselecteerd, tenzij dit vanwege gebrek aan variatie in didactische respons en/of gebrek aan verschillende soorten situaties niet mogelijk was. Er werd per docent nadrukkelijk naar gestreefd om de breedte van het feedback-repertoire van de docent tot zijn recht te laten komen. Dit betekende bijvoorbeeld dat eenzelfde probleem dat nogmaals ter sprake kwam bij andere leerlingen dan in de eerste situatie en waarbij de docent in de nieuwe situatie op een andere manier feedback gaf dan eerst, als een nieuw (aanvullend) fragment is geselecteerd. Tevens werd zoveel mogelijk gespreid over de leerlingen die in de interacties waren betrokken.

Op het moment dat de interactie was afgerond, stopte ook het fragment. De interactie betrof vaak niet alleen de feedback op zich (de directe reactie van de docent) maar ging naadloos over in een gesprek waarbij de docent als tutor of coach optrad; zolang daarbij de interactie nog niet was afgerond, is de opname voortgezet. Per docent zijn vijf of meer fragmenten geselecteerd (dit lukte niet bij docent 7 omdat in de opname slechts vier verschillende feedbacksituaties aanwezig waren).
Per docent zijn de eerste vijf geselecteerde fragmenten geobserveerd met behulp van het ontwikkelde observatieformulier. Daarbij is per fragment per activiteit aangegeven of deze activiteit wel of niet voorkwam. Met deze manier van scoren kon zicht worden verkregen op het repertoire van de docent (welke activiteiten komen over vijf fragmenten bezien voor en welke niet) en ook enigszins op de relatieve frequentie per activiteit (hoe vaak over vijf fragmenten).

Observatie van geselecteerde fragmenten
In de literatuur is nagegaan welke docentactiviteiten kunnen worden onderscheiden in situaties waarin leerlingen zelfstandig aan het werk zijn en de docent feedback geeft. We richtten ons op een brede range aan docentactiviteiten, inclusief aspecten van coaching en tutoring. Dit sluit aan bij de realiteit waarin de in een eerste reactie gegeven feedback vaak uitmondt in een dialoog tussen docent en leerling(en). We hebben ons georiënteerd op literatuur over feedback (Shute, 2008), coachen (Veenhoven, 2004), tutoring (Chi, 2001, Chi et al., 2001; King, 2002), docentgedrag (Torrance & Pryor, 1998) en het cognitive apprenticeship model (Collins, Brown & Newman, 1989). Op basis hiervan is een observatieformulier ontwikkeld met 23 docentactiviteiten, verspreid over vijf categorieën: geen feedback, resultaatgerichte feedback, procesgerichte feedback, aanvullende instructie, en overig. Zie figuur 5.2. Van de kenmerken in figuur 5.1 zijn in dit onderzoek a-c (Context en Timing) een gegeven (directe interactie in een sociale setting), terwijl 3-8 (cognitieve en affectieve aspecten) alleen goed kunnen worden onderzocht vanuit de beleving van de leerlingen, hetgeen niet uitvoerbaar was in dit onderzoek en voor de vraagstelling ook niet nodig.

Stimulated recall interview
Voor dit interview is een interviewleidraad ontwikkeld met vragen over de context (vak, leerjaar, onderwijssetting), de taak (type, gestructureerdheid, gelaagdheid, eenduidigheid van de standaard) en de vooraf gegeven instructie (vorm, mate, relatie met leerdoelen), en een aantal vragen, te stellen na het bekijken van elk fragment: a. Wat was het probleem van de leerling? b. Wat deed je als docent (typering, bewust/ onbewust, intenties, gewone manier van reageren)? c. Wat was de basis voor de reactie (kenmerken leerlingen, niveau, specifieke problemen, relatie) of waarom reageerde je zo? d. Wat was het effect bij de leerling (begrip, geholpen)? e. Waarom of naar aanleiding waarvan rondde je de interactie af? Deze laatste vraag betreft gedrag bij feedback-ontvangers dat voor feedbackgevers aanleiding is te concluderen dat de ontvangers iets hebben begrepen (Reynolds, Martin & Groulx, 1996).

Geen feedback
1. zelf overnemen of zelf oplossen als docent
Resultaatgerichte feedback
2. aangeven of bevestigen wat tot nu toe correct of succesvol is zonder informatie waarom iets goed is
3. idem met informatie waarom iets goed is
4. geven van suggesties voor wat beter kan, wat nog ontbreekt/ of wat (nog) fout is zonder informatie waarom
5. idem met informatie waarom iets fout is of ontbreekt
6. aangeven wat de goede oplossing is zonder informatie waarom
7. idem met informatie waarom
8. aangeven wat de standaard is waaraan moet worden voldaan
Procesgerichte feedback
9. de leerling stimuleren of vragen het probleem preciezer te omschrijven en te analyseren
10. vragen naar wat de leerling al heeft gedaan of hoever deze is
11. vragen naar bepaalde kennis
12. stellen van open vragen die het denkproces van bij de leerling stimuleren
13. het probleem verhelderen en/of vereenvoudigen door het op te delen in stappen
14. geven van hints, cues, prompts, voorbeelden
15. geven van suggesties hoe het beter kan
16. voordoen van de aanpak, het probleem, oplossingsproces modelleren
17. vragen of iets duidelijk c.q. begrepen is
18. geven van suggesties of voordoen van manieren om het proces zelf te monitoren en tussentijds evalueren
Aanvullende instructie
19. het verduidelijken of aanvullen van de eerder gegeven instructie
20. een begrip uitleggen
21. uitleg geven over een aanpak om tot een oplossing te komen
Overig
22. feedback geven op gedrag
23. bemoedigende opmerking maken

Figuur 5.2 Categorieën en docentactiviteiten in het observatieformulier voor feedbacksituaties

Gemiddeld vier dagen na de opname vond het stimulated recall interview plaats. In het begin van het gesprek werd aangegeven wat het doel ervan was (een toelichting vragen op het handelen dat op de videobeelden te zien is en inzicht krijgen in overwegingen en intenties). Daarna werden vragen gesteld over de context (vak, leerjaar, onderwijssetting), de taak (type, gestructureerdheid, gelaagdheid, eenduidigheid van de standaard) en de vooraf gegeven instructie (vorm, mate, relatie met leerdoelen). Vervolgens werden één voor één de fragmenten getoond op een beeldscherm en werden na elk fragment de eerder genoemde vragen gesteld. Afhankelijk van de uitgebreidheid van de antwoorden van de docent per fragment kwamen wel of niet alle vijf fragmenten aan bod. De docent kon zelf aangegeven of nog een fragment zou worden bekeken en besproken, en kon desgewenst zelf het fragment kiezen. De interviews zijn opgenomen met geluidsapparatuur en integraal uitgetypt.

Eindgesprek met de docent

Met elke docent is een eindgesprek gevoerd. Daarin werd eerst samen teruggekeken naar de in het begin gemaakte conceptmap en werd de docent gevraagd bepaalde onderdelen nog even toe te lichten.

Vervolgens werden de uitkomsten van de beide leerlingvragenlijsten (voorafgaand aan en direct na afloop van de opname) op itemniveau en op schaalniveau besproken. De docent kreeg deze op papier uitgereikt. Door de onderzoekers werd stilgestaan bij die items waarop de leerlingen relatief hoog of laag hadden gescoord of waarbij de leerlingen gemiddeld aangaven duidelijk meer of minder te willen dan de docent deed (met als criterium een verschil van tenminste 0,5 schaalpunt). De docent kreeg opnieuw de gelegenheid te reageren.

Verschillende uitkomsten werden vergeleken, waaronder de gemiddelde scores van de docent zelf op de subschalen Geven van feedback en Benutten van feedback en de gemiddelden van de leerlingen en ook de gemiddelden van een grote groep docenten in een eerder onderzoek op dezelfde schalen.

Het scoringsprotocol van de videofragmenten werd uitgereikt met daarop ingevuld de frequenties van de feedbackgedragingen van de docent. Eventuele opmerkingen en aandachtspunten vanuit de onderzoekers werden gedeeld met de docent. De docent kreeg uitgebreid de gelegenheid te reageren.

Aan het eind van het gesprek zei een aantal docenten uit zichzelf deze integrale terugrapportage zo informatief te vinden dat ze er op eigen kracht zinvolle leerpunten aan konden ontlenen (en ze benoemden deze ook).

De eindgesprekken zijn opgenomen met geluidsapparatuur en integraal uitgetypt. De inhoud ervan is samen met de gegevens van alle andere instrumenten verwerkt in een individuele terugrapportage die de docent tot slot per post is toegestuurd. Daarin werd de docent ook geïnformeerd over zijn scores in relatie tot de scores van de andere docenten (als groep) die hadden meegedaan aan dit onderzoek en werd per docent een door de onderzoekers opgesteld persoonlijk advies toegevoegd.

Analyses

De kwantitatieve data uit de voorgestructureerde vragenlijsten voor docenten en leerlingen en het voorgestructureerde observatieformulier zijn geanalyseerd met schaalanalyses (criteria: itemrestcorrelaties minimaal .20, Cronbach's alpha minimaal .60), beschrijvende analyses (frequentie, gemiddelde, standaardafwijking) en analyses van verschillen en verbanden (significantieniveau: .05).

De kwalitatieve data zijn verwerkt in matrices per instrument. Elke matrix bevatte in de regels de bevraagde personen en in de kolommen de thema's of categorieën waarin de betreffende data bleken te kunnen worden geordend. In de paragraaf over de resultaten zijn enkele daarvan als tabel opgenomen, die mede kunnen dienen als concrete voorbeelden van de gevolgde aanpak.

Instrument	*Thema's en categorieën per instrument*				
conceptmap schoolleiders	Voorwaarden voor feedback	Beoogde doelen van feedback	Gewenst handelen van docenten	Vormen van feedback	Objecten van feedback
conceptmap docenten	Voorwaarden	Inhoud	Situatie	Vormen, kenmerken, activiteiten	Doelen van feedback
open vragen in vragenlijst docenten	Intenties, doelen	Aanpak, kenmerken	Objecten van feedback	Rekening houden met	Verwijzen naar
stimulated recall interview	Probleem van de leerling	Intentie van de docent	Basis voor de reactie	Effect bij de leerling	Reden afronding interactie

Figuur 5.3 Thema's en categorieën in de analysematrices bij de kwalitatieve data

In figuur 5.3 staat een samenvatting van de thema's c.q. categorieën per instrument. Elke matrix is door één onderzoeker ontworpen en ingevuld en door een andere onderzoeker vergeleken met de data; bij verschil van inzicht, over een categorie of over de toewijzing daarvan, is via discussie consensus bereikt. De beschrijvingen van de resultaten (zie 5.4) zijn teruggekoppeld aan de schoolleiders respectievelijk de docenten (*member-check*, zie Boeije, 2008). Zij gaven aan zich er goed in te kunnen herkennen.

Uitkomsten analyses representativiteit deelnemende docenten en leerlingen
De 12 docenten antwoordden op de aan hen voorgelegde items over de subtaakgebieden Geven van feedback en Feedback benutten bij zowel het huidige als het gewenste handelen gemiddeld even hoog (maximaal 0,1 schaalpunt verschil) als de 141 docenten waarover in hoofdstuk 3 is gerapporteerd. De 324 leerlingen in dit onderzoek scoorden gemiddeld op de subschalen Geven van feedback en Feedback benutten vrijwel gelijk aan de 900 leerlingen in het eerdere onderzoek. Bij het huidig en gewenst handelen van hun 12 docenten op de schaal Geven van feedback scoorden zij gemiddeld 0,2 respectievelijk 0,1 schaalpunt hoger en bij het huidig en gewenst handelen op de schaal Feedback benutten gemiddeld 0,1 respectievelijk 0,2 schaalpunt lager. De 12 docenten vormen dus een representatieve groep.

Uitkomsten schaalanalyses
De vijf motivatieschalen in de vragenlijst vooraf voor leerlingen bleken alle betrouwbaar: self-efficacy .83, intrinsieke motivatie .88, geïdentificeerde regulatie .71, externe regulatie .73, amotivatie .82. De eerste drie schalen vormen samen een schaal voor de motivatie van de leerlingen (alpha .81).
De schalen over feedback die zijn ontleend aan eerder onderzoek (zie de hoofdstukken 3 en 4) bleken ook nu weer voldoende betrouwbaar, behalve bij de subschaal Geven van feedback bij de docenten. Bij deze subschaal bleek sprake van enkele lage itemrestcorrelaties en een kleine standaarddeviatie. De vier subschalen in de vragenlijst

vooraf voor de leerlingen ten aanzien van het huidig handelen van hun docent vormen samen een betrouwbare overkoepelende schaal voor feedbackgedrag (alpha .87). De *verschillen* tussen huidig en gewenst handelen bij de vier subschalen vormen samen een betrouwbare schaal voor de mate waarin de leerlingen meer wensen dan zij nu waarnemen (alpha .89).
De 22 aanvullende items over Feedback geven vormden samen een betrouwbare schaal. Een 12-tal items die kunnen worden opgevat als aspecten van het geven van feedback gericht op het resultaat (de items 1, 2, 3, 5, 7, 9, 11, 12, 13, 15, 18, 19), vormden een betrouwbare subschaal. De overige 10 items die kunnen worden opgevat als aspecten van het geven van feedback gericht op het proces (de items 4, 6, 8, 10, 14, 16, 17, 20, 21, 22), vormden ook een betrouwbare subschaal.
Van de 20 items over Geven van feedback in de vragenlijst voor leerlingen direct na afloop van de opgenomen les vormden 9 items een schaal, op te vatten als maat voor acceptatie en bruikbaarheid van de feedback, en 6 items een schaal over de mate waarin de feedback procesgerichte aanwijzingen bevatte. Zie tabel 5.2.

Tabel 5.2: Schaalanalyses data docenten (n=12) en leerlingen vooraf (n=324) en na afloop (n=149)

Schaal	Aantal items	Docenten		Leerlingen	
		Huidig handelen	*Gewenst handelen*	*Huidig handelen*	*Gewenst handelen*
Taakgebied geven van feedback	11	.67	.77	.90	.89
Subschaal / taak geven van feedback	5	<.60	.60	.86	.86
Subschaal / taak feedback benutten	4	.67	.70	.80	.77
Aanvullende schaal Feedback geven	22	–	–	.91	.88
Subschaal resultaatgerichte feedback	9	–	–	.84	.83
Subschaal procesgerichte feedback	10	–	–	.82	.77
Acceptatie en bruikbaarheid (na afloop)	9	–	–	.81	–
Procesgerichte aanwijzingen (na afloop)	6	–	–	.68	–

Uitkomsten berekeningen interbeoordelaarsovereenstemming observaties
Twee onderzoekers hebben de eerste vijf fragmenten van de eerste drie docenten (dus 15 fragmenten) onafhankelijk van elkaar gescoord. Daarna is de interbeoordelaars-overeenstemming berekend:

a. per docent over alle 23 activiteiten op de totalen per activiteit van de vijf fragmenten samen (dus per activiteit een score tussen 0 en 5), met een jury alpha; de jury alpha's waren: .94, .85 en .91;
b. idem, over de drie docenten samen (69 codes), met een correlatie; deze was .82;
c. per docent per fragment over alle 23 activiteiten (met per activiteit een score van 0 of 1), met een jury alpha; hiervan waren 19 alpha's voldoende (≥ .60) en 4 te laag (< .55).
d. idem, over de drie docenten samen (345 codes), met een correlatie; deze was .61.

De vier activiteiten waarbij de alpha te laag was, zijn nader bekeken. Activiteit 1 kwam amper voor en activiteit 18 bleek niet voldoende eenduidig interpreteerbaar; beide zijn geschrapt. De scorings-afspraken bij de activiteiten 4 en 6 konden worden aangescherpt waarna de alpha's voldoende bleken (>. 70). Afgesproken is dat code 4 positief gescoord moest worden als code 5 positief gescoord werd. Dit leverde een jury alpha voor code 4 op van .76. Ook is afgesproken dat code 6 positief gescoord moest worden als code 7 positief gescoord werd. Dit leverde een jury alpha voor code 6 op van .72.
Als controle op de representativiteit van de geselecteerde fragmenten voor de totale gemaakte opname, is door een van beide onderzoekers tot slot de hele opname van docent 1 gescoord, met als fragmenten de vier delen van de betreffende les waaruit de opname bestond (de opname was daartussen steeds even gestopt omdat zich op dat moment geen voor dit onderzoek relevante docentactiviteit voordeed). Beide scoringen (van de eerste vijf geselecteerde fragmenten en van de vier delen van de hele les) zijn behandeld als twee onafhankelijke scoringen en dit leidde tot een jury alpha van .76.
Naar aanleiding van deze uitkomsten is het observatieformulier aangepast. Het aantal docentactiviteiten is door weglating van twee activiteiten teruggebracht tot 21. De scores van de eerste drie docenten zijn aangepast aan de nieuwe codeerafspraken. De fragmenten van de overige negen docenten zijn door beide onderzoekers onafhankelijk van elkaar gescoord. Bij verschil van scoring zijn de onderzoekers via discussie en zo nodig terugkijken (opnieuw observeren) tot consensus gekomen.

5.4 Resultaten

Visies op feedback van schoolleiders en docenten

De schoolleiders zien gelet op hun conceptmaps als voorwaarden voor het geven van feedback vooral organisatorische en structurele aspecten (goede organisatie en lesopbouw; integraal onderdeel leerproces; tijd nemen; heldere doelen; positief leerklimaat; duidelijk maken wat je als docent onder feedback verstaat) en aspecten die te maken hebben met afstemmen op de leerlingen (weten wat er leeft, goed kunnen luisteren, goed observeren, maatwerk). Dit laatste sluit aan bij wat de docenten aangeven (ruim de helft schetst in hun conceptmap feedback als proces: de docenten 2, 4, 5, 6, 9, 11, 12), maar de structurele aspecten zijn door de docenten niet genoemd.
De doelen van feedback die de schoolleiders noemen, komen overeen met die van de docenten en betreffen zowel de leerstof (verduidelijken, bewerkstelligen van verwerking, uitdaging en verdieping) als de ontwikkeling van leerlingen (voortgang, groei, ontwikkeling en reflectie). Het pedagogisch handelen dat de schoolleiders bij docenten zouden willen zien, betreft geven van feedback vanuit een positieve en betrokken insteek waarbij wordt benadrukt wat goed gaat, vertrouwen wordt gegeven en kwaliteiten van leerlingen worden benoemd. Dit komt overeen met wat de docenten belangrijk vinden. Voor didactisch handelen noemen schoolleiders ook diagnosticeren en relateren aan de standaard maar dat komt bij de docenten minder naar voren.

Beide groepen beschrijven concrete handelingen die ze wenselijk achten in relatie tot de doelen. Dit betreft enerzijds pedagogisch handelen dat nauw samenhangt met het beoogde positieve leerklimaat (stimuleren van zelfvertrouwen, opbouwen van een reëel zelfbeeld, etc.) en anderzijds didactisch handelen dat erop is gericht bij leerlingen zowel kennis als vaardigheden te ontwikkelen (kiezen, samenwerken, plannen, etc.). Hierbij is dus aandacht voor zowel de taak en het leerproces als de persoonlijke ontwikkeling van leerlingen in relatie tot anderen. Zowel de schoolleiders als de meeste docenten streven naar een actieve rol van de leerlingen. Hierbij verschilt echter de mate waarin leerlingen zelfstandig, actief en/of verantwoordelijk mogen of moeten zijn en mogen bijdragen aan de inrichting van hun eigen leerproces per docent, en ook per schoolleider.
De schoolleiders en de docenten komen overeen in de gedachte dat feedback kan worden gegeven op inhoud, aanpak, proces en resultaat en op de persoon van de leerling, en dat leerlingen actief in de interactie moeten worden betrokken door vragen te stellen, leerlingen keuzes te laten maken, samen te analyseren wat goed ging en wat beter kan, samen leerpunten te benoemen, te checken of de leerling het heeft begrepen en aandacht te geven aan de reactie van de leerling.
Beide groepen geven aan dat feedback zowel door de docent als door leerlingen aan elkaar kan worden gegeven. Op een aantal scholen is het geven van feedback als een door docenten te ontwikkelen vaardigheid benoemd. Docenten noemen niet specifiek de feedback die ze van leerlingen zouden kunnen krijgen. Wel noemt een aantal docenten in de vragenlijst dat ze aandacht besteden aan en oog hebben voor het samen reflecteren met de leerling.
Al met al stemmen de opvattingen over feedback van de schoolleiders en de docenten als groep in vrij hoge mate met elkaar overeen, maar is wel sprake van verschillen tussen de docenten, ook per school. De variatie in de aspecten die de docenten noemen is groot en dit geldt ook voor docenten van eenzelfde school. Daardoor kan niet per school worden getypeerd welk feedbackgedrag van docenten op die school door schoolleider en docenten wenselijk wordt geacht. Ook de relatie tussen de genoemde aspecten in de concepmaps van de schoolleiders en de docenten van dezelfde school kan niet één op één worden gelegd.

Geven van feedback door de docenten: opvattingen, intenties en percepties van hun handelen

De antwoorden op de open vragen in de vragenlijst die de docenten in het begin hebben ingevuld, waren een tweede manier om de eigen visie op feedback weer te geven. In dit geval werden de antwoorden enigszins gericht door de gestelde vragen, die weliswaar open waren maar wel de aandacht vestigden op bepaalde aspecten.
In tabel 5.3 zijn de uitkomsten weergegeven, op basis van de inhouden van de antwoorden op de gestelde vragen gezamenlijk. De op deze manier verkregen visies van de docenten op feedback hebben betrekking op hun intenties en doelen, aanpakken

en kenmerken van feedback, objecten van feedback, de aspecten waar de docenten bij het geven van feedback rekening mee houden en waar ze naar verwijzen. Uit de resultaten van deze manier van bevragen blijkt dat de doelen die de docenten hebben en de inbreng die ze van hun leerlingen verwachten, veel breder spelen dan met de concept map in kaart is gebracht. De variatie tussen de docenten blijkt groot te zijn.

Bij de intenties en doelen van de docenten met feedback, komen opnieuw de twee clusters terug die bij de concept maps zichtbaar werden. Een aantal docenten noemt cognitieve doelen (zoals kennis, begrip, aanpak, resultaten) (n=6), en een aantal noemt affectieve doelen (stimuleren, motiveren, versterken zelfvertrouwen, verbeteren houding, succes laten ervaren) (n=8). Daarnaast noemen enkele docenten meer algemeen doelen die te maken hebben met het proces van denken, leren en verdiepen (n=2). Eén docent noemt een doel op metacognitief gebied (zicht op eigen kunnen). Drie docenten noemen (ook) het benutten van de feedback als doel.

Wat betreft de aanpakken en kenmerken van feedback komt een aantal activiteiten terug die ook al zichtbaar werden in de concept maps: afstemmen, checken, controleren, toelichten, tips geven. Aanvullend worden nu genoemd: aandacht richten, focus aanbrengen, laten samenvatten, visualiseren, voorbeelden geven, vragen beantwoorden, via tussenstap komen tot antwoord, coachen, begeleiden, reflecteren, bewijzen, belonen, wijzen op verbeterpunten.

Objecten van feedback hebben deels betrekking op aspecten van de taak: (vak)inhoud, kennis, stof, aanpak, vaardigheden, resultaat, product. Daarnaast worden motivationele, affectieve en sociale aspecten genoemd, zoals werkhouding, zelfvertrouwen, manier van werken, samenwerking, gedrag, proces, concentratie, werktempo. De meeste docenten noemen beide soorten aspecten.

Gevraagd waarmee ze bij het geven van feedback rekening houden, noemen de docenten opnieuw zowel cognitieve als affectieve aspecten: niveau, leerjaar, begrip, taalgebruik en tempo, naast zelfvertrouwen, zelfbeeld, belevingswereld, persoonlijke situatie en specifieke problemen. Ook hier noemen veel docenten beide soorten aspecten. Een enkele docent noemt ook het belang van privacy.

Tweederde van de docenten gaf antwoord op de vraag waarnaar ze in hun feedback verwijzen. De meesten (n=6) noemen verschillende typen bronnen, zoals het leerboek, de door de docent gegeven instructie en uitleg, de gemaakte afspraken, de gegeven opdracht, de gestelde eisen. Enkelen (n=3) verwijzen (ook) naar de leerling zelf: zijn of haar parate kennis, voorkennis, ervaring en leefwereld.

Wat betreft de eigen rol en inbreng van de leerlingen: het belang daarvan wordt door de meeste docenten onderstreept, maar hun antwoorden lijken een enigszins verschillende kleur te hebben. Sommige docenten noemen het bevorderen van de zelfstandigheid van de leerlingen (n=3), waarbij de docent de primaire actor lijkt te zijn, anderen antwoorden in termen van aanspreken op de eigen verantwoordelijkheid en benutten van de actieve houding van leerlingen (n=3), en weer anderen lijken de leerlingen een nog iets grotere eigen rol toe te kennen door te antwoorden in termen van samen reflecteren, leerlingen om hun mening vragen en leerlingen zelf keuzes laten maken.

Tabel 5.3: Visies van docenten op feedback: intenties, aanpak, objecten en activiteiten (open vragen in vragenlijst docenten)

D	*Intenties, doelen*	*Aanpak, kenmerken*	*Objecten van feedback*	*Rekening houden met*	*Verwijzen naar*
1	Vergroten begrip en toepassing, versterken van zelfvertrouwen	Vragen stellen en beantwoorden, checken, individueel/ klassikaal	Begrip, aanpak werkhouding en samenwerking	Niveau, zelfbeeld en zelfvertrouwen van leerlingen	
2	Stimuleren van denken, begrip en leren	Goed gedrag belonen, begrip controleren	Gedrag, zelfvertrouwen, leerproces	Niveau, situatie, zelfvertrouwen van leerlingen	Boek, gegeven uitleg, parate kennis
3	Focus aanbrengen en verdiepen	Tips geven, laten verwoorden, aandacht richten	Kennis, vaardigheden, samenwerking, concentratie, stem	Verschillen tussen leerlingen (tempo)	Kennis, ervaring
4	Stimuleren, motiveren, leren van fouten		Voortgang en diepgang	Individuele verschillen (snelheid van begrip)	
5	Stimuleren van kennis, aanpak, zelfvertrouwen, benutten feedback	Belonen van positief gedrag, vragen, checken, antwoord en suggesties geven	Gedrag, manier van werken, producten, resultaten	Specifieke problemen	Gegeven uitleg, voorkennis, boek, antwoordenboek
6	Kennis en begrip, bijsturen, verdiepen	Via tussenstap komen tot antwoord, aansluiten, checken visualiseren	Werkhouding, aanpak, opdracht	Niveau van de leerling, tijdsduur, taalgebruik, reactie	
7	Focus op ontwikkelpunten. Benutten van feedback.	Aandacht richten, samen werken en reflecteren. Smart formuleren	Resultaat, aanpak, prioriteiten, uitspraak, samenwerking, werkhouding	Context, leerjaar persoonlijke situatie, doelen	
8	Zelfvertrouwen en begrip vergroten; aanpak, houding verbeteren	Ondersteunen, begeleiden,bevestigen, coachen, contact maken en afstemmen		Niveau van leerling en de situatie	Opdracht, voorkennis, houding en inzet
9	Resultaten verbeteren, leren stimuleren, benutten van feedback	Terugkoppelen, toelichten, bewijzen	Gedrag, werkstukken, aanpak, werkhouding	Algemene eisen aan alle leerlingen, individuele kenmerken	Afspraken, eisen, mogelijkheden, oplossingen, verwachtingen en resultaten
10	Leerlingen helpen in proces en motiveren	Wijzen op sterke en verbeterpunten, tips en voorbeelden geven, laten samenvatten	Resultaat, inhoud, proces, samenwerking en reflectie	Belevingswereld van leerling, leerstijl, aanpak	Gegeven instructie, tips
11	Zicht op eigen kunnen, stimuleren en motiveren.	Positieve punten benoemen, vragen stellen, luisteren, focus aanbrengen	Resultaat, werktempo, proces, welbevinden, verbeterpunten	Taalgebruik, privacy, uitstraling	Vergelijkbare situaties, groei
12	Stimuleren van het leerproces, succeservaringen laten opdoen	Contact maken, randvoorwaarden creëren, afstemmen, begeleiden	Vak, lesstof Welbevinden en succeservaringen, gedrag	Positieve insteek, taalgebruik, begrip, mate van zelfreflectie	Opdracht, situatie, gelezen stof, eigen leefwereld

In het vervolg van de vragenlijst gaven de docenten antwoord op de items over het geven van feedback en doen benutten van feedback. Als eerste onderzoeksactiviteit bij de leerlingen lieten zij de groep leerlingen die aanwezig zou zijn tijdens de situatie die later zou worden opgenomen een vragenlijst invullen. Deze vragenlijst bevatte eveneens de schalen over het geven en doen benutten van feedback, en aanvullend de schalen over het geven van procesgerichte en resultaatgerichte feedback Steeds werd gevraagd naar het huidig handelen en het gewenste handelen. Zie tabel 5.4.
De docenten schatten hun huidige handelen bij het geven van feedback vrij hoog in, en bij het doen benutten van feedback meestal veel lager. Het door hen aangegeven gewenste handelen ligt bij veel docenten een stuk hoger dan het huidige handelen (zie de docenten 1-3, 6, 8-12). Bij de meeste docenten komt hun eigen inschatting van hun huidige handelen behoorlijk overeen met de inschatting van hun leerlingen. Sommige docenten onderschatten hun eigen handelen, vergeleken met wat de leerlingen aangeven (docenten 1, 9, 11), anderen overschatten dat (docenten 4, 8).

Tabel 5.4: Percepties van docenten en leerlingen van huidig (H) en gewenst (G) docenthandelen

			DOCENTEN (n=12)				*LEERLINGEN (n=324)*							
School	Docent	Vak	Geven van feedback		Feedback doen benutten		Geven van feedback		Feedback doen benutte		Resultaat-gerichte feedback		Proces-gerichte feedback	
			H	*G*	*H*	*G*	*H*	*G*	*H*	*G*	*H*	*G*	*H*	*G*
A	1	Wiskunde	3.4	4.8	1.8	4.3	3.9	3.9	3.8	3.9	3.9	4.0	3.8	4.0
	2	Geschied.	4.2	4.6	2.8	3.5	4.2	4.5	3.5	4.0	4.0	4.3	3.9	3.9
	3	Engels	3.8	4.2	2.3	3.8	3.9	4.3	3.2	3.9	3.9	4.3	3.6	4.0
B	4 *	Latijn	3.8	-	4.0	-	3.6	4.1	3.0	3.6	3.6	4.4	3.3	4.0
	5	Biologie	3.9	4.4	3.3	3.5	4.1	4.3	2.8	3.3	4.3	4.5	3.7	4.0
	6	Economie	3.9	4.0	2.8	3.8	3.9	4.4	3.0	4.1	3.6	4.6	3.3	4.3
C	7 *	Frans	4.0	-	3.5	-	3.9	4.3	2.8	3.2	3.8	4.2	3.3	3.8
	8	Wiskunde	4.0	4.6	3.5	3.8	3.2	4.0	2.7	3.6	3.4	4.0	3.1	3.8
	9	Techniek	4.0	4.5	3.5	4.5	4.6	4.6	3.9	3.7	4.6	4.6	4.2	4.3
D	10	Economie	4.0	4.4	3.3	4.0	3.8	3.8	3.3	3.4	3.8	4.1	3.7	3.8
	11	Economie	4.0	4.8	1.8	3.5	4.5	4.5	4.0	3.9	4.3	4.4	4.0	4.0
	12	Engels	4.2	4.6	3.8	4.5	4.2	4.2	3.9	3.9	4.0	4.2	3.9	4.0

* Twee docenten gaven geen antwoorden op de vraag naar het gewenste handelen

De leerlingen schatten op alle schalen het huidige handelen van hun docenten gemiddeld in tussen 3 en 4,5, maar feedback gericht op het proces (tussen 3,1 en 4,2) en doen benutten van feedback (tussen 2,7 en 4,0) komen wat lager uit. Een aantal docenten scoort bij de leerlingen vrij hoog (docenten 2, 5, 9, 11 en 12). Tabel 5.4 laat zien dat meer feedback wordt gegeven op het resultaat dan op het proces en dat er nog niet veel aandacht is voor het benutten van feedback.
Het feedbackgedrag van de docenten volgens de leerlingen (de overkoepelende schaal van de vier schalen voor huidig handelen) verschilt significant tussen scholen en tussen docenten. Zie tabel 5.5.

Tabel 5.5: Verschillen in de data van de leerlingen tussen scholen en tussen docenten (gemiddelde en sd)

	Huidig handelen van docenten volgens leerlingen				*Feedback huidig handelen (totaal)*	*Verschil gewenst min huidig (totaal)*	*Acceptatie en bruik-baarheid * (na afloop)*	*Procesge-richte aan-wijzingen * (na afloop)*
	Geven van feedback	*Feedback doen benutten*	*Resultaat-gerichte feedback*	*Proces-gerichte feedback*				
Docent	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.01*
1	3.9 (.57)	3.8 (.64)	3.9 (.48)	3.8 (.61)	3.9 (.52)	0.12 (.26)	4.3 (.52)	3.4 (.98)
2	4.2 (.57)	3.5 (.82)	4.0 (.46)	3.9 (.40)	3.8 (.53)	0.13 (.15)	4.3 (.37)	4.0 (.49)
3	3.9 (.40	3.2 (.84)	3.9 (.47)	3.6 (.33)	3.7 (.37)	0.40 (.27)	3.8 (.37)	3.8 (.49)
4	3.6 (.91)	3.0 (1.05)	3.6 (.72)	3.3 (.67)	3.3 (.74)	0.71 (.77)	3.9 (.59)	3.3 (.75)
5	4.2 (.55)	2.8 (1.06)	4.3 (.32)	3.6 (.39)	3.7 (.30)	0.31 (.41)	4.1 (.65)	3.8 (.85)
6	3.8 (.41)	3.0 (1.14)	3.6 (.69)	3.3 (.72)	3.4 (.64)	0.88 (.88)	4.8 (.20)	3.2 (.79)
7	3.9 (.55)	2.8 (.45)	3.8 (.42)	3.3 (.47)	3.4 (.36)	0.59 (.40)	4.5 (.20)	4.4 (.32)
8	3.3 (1.10)	2.7 (1.04)	3.4 (.67)	3.1 (.72)	3.5 (.52)	0.81 (.94)	3.7 (.79)	3.4 (.58)
9	4.6 (.37)	3.9 (.59)	4.6 (.36)	4.2 (.40)	4.4 (.42)	-0.04 (.14)	4.3 (.59)	3.7 (.78)
10	3.9 (.60)	3.3 (.82)	3.8 (.61)	3.7 (.69)	3.6 (.55)	0.10 (.46)	3.8 (.70)	3.7 (.73)
11	4.4 (.49)	4.0 (.66)	4.3 (.32)	4.0 (.47)	4.2 (.38)	0.05 (.16)	3.6 (1.23)	3.3 (1.23)
12	4.1 (.53)	3.9 (.72)	4.0 (.56)	3.9 (.58)	4.0 (.48)	0.16 (.33)	4.7 (.42)	4.6 (.78)
School	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*p<.001*	*ns **	*ns**
A	4.0 (.52)	3.5 (.81)	4.0 (.46)	3.7 (.46)	3.8 (.47)	0.24 (.27)	4.1 (.49)	3.7 (.70)
B	3.8 (.78)	2.9 (1.05)	3.8 (.69)	3.4 (.65)	3.4 (.67)	0.64 (.74)	4.1 (.64)	3.4 (.80)
C	3.6 (1.04)	2.8 (.97)	3.6 (.69)	3.3 (.75)	3.6 (.56)	0.65 (.84)	4.2 (.63)	3.8 (.74)
D	4.2 (.55)	3.8 (.75)	4.0 (.56)	3.9 (.59)	4.0 (.50)	0.12 (.33)	4.1 (.95)	3.8 (.99)
Totaal	3.9 (.77)	3.3 (.98)	3.9 (.63)	3.6 (.67)	3.7 (.60)	.40 (.64)	4.1 (.61)	3.7 (.78)

* de data over deze twee schalen betroffen bij de docenten 5-8 en 10-12 minder dan 10 leerlingen
** niet significant

Gelet op het verschil tussen hun beoordelingen van het huidige handelen en het door hen gewenste handelen, zijn de leerlingen redelijk tevreden. Dit neemt niet weg dat ze op een aantal punten wel iets meer zouden willen. Dit doet zich bij een deel van de docenten vrij duidelijk voor (de docenten 3, 4, 6 en 8), maar bij andere docenten amper (de docenten 1, 9, 10, 11, 12). Ook het verschil tussen huidig en gewenst handelen van de docenten volgens de leerlingen verschilt significant tussen de scholen en tussen de docenten.

De motivatie van leerlingen voor taken voor het vak bij de betreffende docent

De motivatie van de leerlingen voor het maken van taken voor het vak bij de docent varieert bij elke schaal significant (p<.01) tussen de docenten, zie tabel 5.6.

Tabel 5.6: Gemiddelden per docent op de motivatieschalen (schaal 1-7)

D	Vak	Motivatie (overkoepelend)	Self-efficacy	Motivatie (vier schalen) *			
				A	*B*	*C*	*D*
1	Wiskunde	4.6	5.0	4.0	4.8	4.6	3.8
2	Geschiedenis	4.6	4.8	4.1	5.0	4.9	3.2
3	Engels	4.7	5.2	3.9	5.1	5.2	2.8
4	Latijn	4.3	4.6	3.6	4.7	4.1	3.2
5	Biologie	5.0	5.2	4.9	5.0	4.3	2.5
6	Economie	4.4	4.8	3.8	4.8	4.1	2.6
7	Frans	5.2	6.0	4.7	5.2	4.8	2.5
8	Wiskunde	4.3	4.8	3.4	4.6	4.8	3.5
9	Techniek	6.0	6.5	6.2	5.3	4.8	2.5
10	Economie	4.9	4.9	4.6	5.2	4.4	3.5
11	Economie	5.3	5.4	4.8	5.7	4.8	2.6
12	Engels	4.8	5.4	5.1	5.2	5.1	4.0

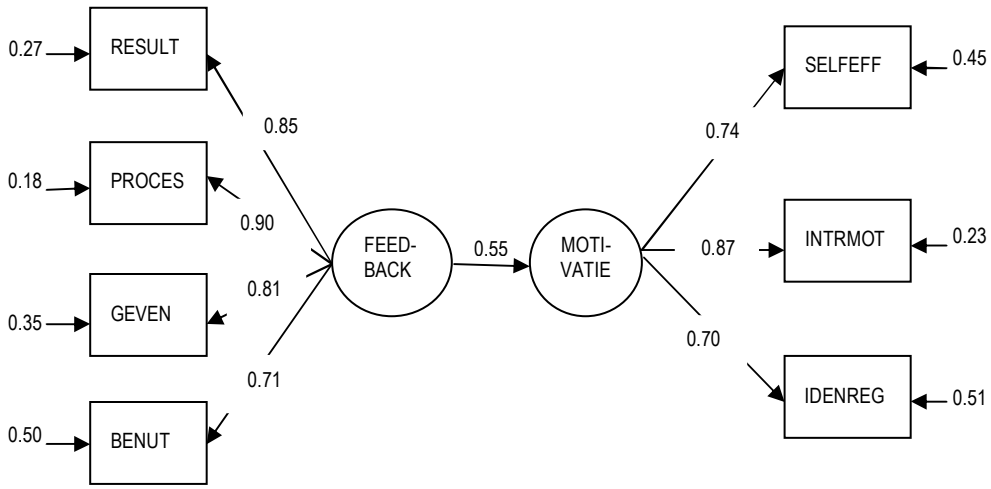
* A=Intrinsieke motivatie, B=Geïdentificeerde regulatie
C=Externe regulatie, D=Amotivatie

De self-efficacy ligt bij de meeste docenten rond de 5 en bij twee docenten hoger (docenten 7 en 9). De motivatie volgens de SIMS heeft bij een aantal docenten vooral de vorm van geïdentificeerde regulatie en externe regulatie (B en C), maar bij andere docenten speelt (ook) de intrinsieke motivatie (A) een voorname rol (docenten 5, 9 en 12). Amotivatie (D) scoorde bij alle docenten laag. Ook de overkoepelende schaal voor motivatie (gebaseerd op self-efficacy, intrinsieke motivatie en geïdentificeerde regulatie) varieert significant tussen de docenten ($p<.01$) en deze varieert van 4,3 voor de docenten 4 en 8 tot 6,0 voor docent 9.

De motivatie van de leerlingen voor het maken van taken bij het vak bij de docent hangt significant negatief samen met de mate van onderwijservaring van de docenten ($r= -.23$, $p<.001$). Bij minder ervaren docenten zijn de leerlingen gemiddeld meer gemotiveerd. In absolute zin is het verschil niet groot (weinig ervaring versus veel ervaring: 4,9 versus 4,4 op de schaal van 1-7).

De motivatie van de leerlingen voor het maken van taken voor het vak bij de docent (de overkoepelende schaal van de drie schalen voor self-efficacy, intrinsieke motivatie en geïdentificeerde regulatie, gem. 4,8 op de 7-puntsschaal, sd 1,00) en de breedte van het repertoire in het huidige feedbackgedrag van de docent volgens de leerlingen (de overkoepelende schaal van de vier feedbackschalen, gem. 3,6 op de 5-puntsschaal, sd 0,68) hangen onderling significant positief samen: de correlatie is .50 ($p<.001$).

Dit verband is aanvullend geanalyseerd en getoetst met LISREL (versie 8.30). Het model is redelijk passend ($\chi2(df=13) = 21,62$ ($p=.062$), RMSEA = 0,058, CFI = 0,98). Het model past vrijwel perfect door de meetfouten van twee feedbackschalen (Result en Benut) te laten correleren ($\chi2$ ($df=12$) = 7,70 ($p= .81$), RMSEA=0,00, CFI=1,00) maar het hier gepresenteerde model is eenvoudiger en de schatting van de relatie tussen Feedback en Motivatie is vrijwel identiek (0,55 versus 0,54). Zie figuur 5.4.



Figuur 5.4: Model van de meting van en relatie tussen Feedback en Motivatie (n=195)

Feedbackgedrag in opgenomen situaties: observaties en leerlingpercepties

Tabel 5.7 biedt een overzicht van de activiteiten waarmee de leerlingen bezig waren tijdens de gemaakte video-opnamen. Soms ging het om het maken van opgaven uit het boek (bij wiskunde en economie, docenten 1, 6, 8). Bij de meeste andere docenten werkten leerlingen aan een opdracht die door de docent zelf was gemaakt. In enkele gevallen ging het om opdrachten waarbij leerlingen samenwerkten of elkaar hielpen (docenten 3 en 9), voor het overige werkten de leerlingen individueel. Bij een aantal docenten was het overigens gebruikelijk of werd door de docent verwacht dat leerlingen elkaar raadpleegden als ze ergens niet uit kwamen (alvorens het aan de docent te vragen).
De aard van de taak c.q. het doel van de activiteit had meestal zowel te maken met het (verder) ontwikkelen van kennis als met het toepassen daarvan en het oefenen van procedures en vaardigheden. De helft van de taken betrof het uitwerken van antwoorden of oplossingen op papier. Daarnaast kwamen voor: het gebruiken van een grafische rekenmachine (docent 1), het in groepjes bespreken van de opzet van een presentatie (3), het maken van een tekening van een cel die was bekeken door een microscoop (5), het oefenen van de uitspraak van woorden in een vreemde taal (7), het maken van een driedimensionaal model van een transportmiddel (9) en het bespreken van ervaringen tijdens een stage (11). In enkele situaties ging het om het nabespreken van een gemaakte toets.

Tabel 5.7: Taken en doelen in de opgenomen situaties

Docent	*Vak*	*Type taak*	*Doel van de les / het werkuur*
1	Wiskunde	Werken aan opgaven uit boek	Toepassen gebruik grafische rekenmachine
2	Geschiedenis	Schema invullen	Begripsontwikkeling, categorisering
3	Engels	Folder, script, presentatie	Informatie zoeken, spreken, samenwerken
4	Latijn	Vertaling en werkblad maken	Leren vertalen en inhoud begrijpen
5	Biologie	Werken aan opdrachten	Opdrachten afmaken of bespreken resultaat
6	Economie	Werken aan opgaven uit boek	Toepassen nieuwe kennis en regels
7	Frans	Oefenen voor presentatie	Oefenen van de uitspraak
8	Wiskunde	Werken aan opgaven uit boek	Bespreken van gemaakte toets
9	Techniek	Model transportmiddel maken	Afmaken model transportmiddel (laatste les)
10	Economie	Werken aan opdrachten	Gelegenheid tot vragen stellen voor toets
11	Economie	Werken aan opdrachten	Gelegenheid na te praten over de stage
12	Engels	Werken aan opdrachten	Gestructureerd werken aan opdrachten

Tabel 5.8 toont de resultaten van de observaties per docent opgeteld over vijf fragmenten, in totaalfrequenties per onderscheiden categorie. De docenten gaven in totaal iets vaker procesgerichte dan resultaatgerichte feedback, maar deze verhouding verschilde nogal tussen de docenten.

Tabel 5.8: Docentactiviteiten in feedbacksituaties, per docent per categorie over 5 fragmenten

Docent / *Categorie*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*Tot*
Resultaatgerichte feedback (7 activiteiten, maximum 35)	16	21	17	12	22	17	21	15	17	7	11	17	193
Procesgerichte feedback (9 activiteiten, maximum 45)	22	14	20	23	20	17	9	28	21	23	18	22	237
Aanvullende instructie + overig (5 activiteiten, maximum 25)	8	6	6	4	11	8	2	11	6	8	3	8	81
Totaal	45	41	43	39	53	42	32	54	44	38	32	47	511

Van de onderscheiden vormen van resultaatgerichte feedback komt het vaakst voor het aangeven of bevestigen van wat goed gaat of is (gemiddeld in 4 fragmenten per docent). Het minst vaak wordt aangegeven wat de goede oplossing is of de standaard waaraan moet worden voldaan (gemiddeld in 1,5 fragment per docent). Het geven van suggesties wat er beter kan zit hier tussenin (gemiddeld in 2,5 fragmenten per docent). De docenten verschillen in hoe vaak zij deze resultaatgerichte vormen van feedback hebben gegeven: van 7 tot 22 keer in vijf fragmenten.
Van de vormen van procesgerichte feedback komen het vaakst voor het geven van hints, cues, prompts en voorbeelden, en het stellen van open vragen die het denkproces bij de leerling stimuleren (beide gemiddeld in bijna 3 fragmenten per docent), en het

minst vaak het voordoen van de aanpak (modelleren), het de leerling stimuleren of vragen het probleem preciezer te omschrijven of te analyseren, en het vragen of iets duidelijk is c.q. begrepen (alle drie gemiddeld in 1,5 fragment per docent). Ook hier verschillen de docenten, van 9 tot 28 keer in vijf fragmenten.
Alle 16 gescoorde vormen van resultaatgerichte en procesgerichte feedback komen regelmatig tot vaak voor. Het getoonde repertoire per docent varieert tussen 13 en 16 vormen, en is dus breed te noemen.
De interactie mondde regelmatig uit in het verduidelijken of aanvullen van de gegeven instructie (gemiddeld in 1 fragment per docent), het uitleggen van een begrip of het uitleg geven over een aanpak (beide gemiddeld in bijna 2 fragmenten per docent). Ook werd regelmatig feedback gegeven op het gedrag van een leerling, of werd een expliciet bemoedigende opmerking gemaakt (beide in gemiddeld 1 fragment per docent). Ook dit verschilde tussen de docenten.
De breedte van het feedbackrepertoire zoals geobserveerd hangt niet samen met de mate van onderwijservaring van de docenten ($F(df = 2, 9) = 1{,}822$, $p = .25$).

Tabel 5.9 laat zien hoe de leerlingen, blijkens de direct na afloop ingevulde vragenlijst, de tijdens de opnamen gegeven feedback hebben gepercipieerd. De leerlingen zijn over het geheel genomen positief over de feedback die ze van hun docenten kregen. De feedback werd vaak duidelijk en terecht gevonden, leerzaam en bruikbaar, en leerlingen wisten vaak wat te doen om verder te kunnen en kregen de kans hun werk te verbeteren. De feedback kwam vaak positief over, en niet vaak negatief.
De feedback ging gemiddeld iets vaker over wat goed was of goed ging, dan over wat nog niet goed was of goed ging, maar dit verschilde wel tussen docenten. Bij een aantal docenten lag het accent op het benoemen van wat goed was (docenten 1, 2, 6, 7, 10, 11, 12), bij enkele anderen meer op het aangeven van wat niet goed was (docenten 3 en 8). De feedback ging vrij vaak over de uitvoering van de opdracht (het meest bij de docenten 11 en 12). De standaard (waar de opdracht aan moest voldoen) werd door drie docenten (2, 8 en 12) vaker genoemd dan door de anderen.
De feedback ging niet vaak over de eigen persoon of werkhouding, behalve bij de docenten 11 (die met individuele leerlingen sprak over hun stage-ervaringen en leerlingen die het moeilijk hadden gehad bevestigde en ondersteunde in wat ze goed hadden gedaan) en 12 (die te maken had met een klas waarmee enkele weken eerder door de school strakke afspraken hadden moeten worden gemaakt).
De mate van acceptatie en bruikbaarheid van de verkregen feedback volgens de leerlingen (de 9-item schaal met gem. 4,1 op de 5-puntsschaal, sd .61) en de mate waarin procesgerichte aanwijzingen zijn gegeven (de 6-item schaal met gem. 3,7, sd .78) (zie tabel 5.2) verschillen niet significant tussen de scholen maar wel tussen de docenten (zie tabel 5.5).

Tabel 5.9: Perceptie van de gegeven feedback per docent (vragenlijst leerlingen direct na de opname (1= helemaal niet, 5= helemaal wel).

Docent	1	2	3	4	5	6	7	8	9	10	11	12	G*
Aantal leerlingen (N)	*18*	*18*	*21*	*21*	*10*	*6*	*5*	*5*	*15*	*4*	*8*	*5*	*135*
1. De feedback was duidelijk	4.3	4.8	4.0	4.4	4.8	4.7	4.8	3.4	4.3	4.3	5.0	4.6	4.5
2. Ik vond de feedback terecht	4.5	4.7	3.9	4.4	4.5	4.5	4.8	4.0	4.1	4.0	3.9	4.6	4.3
3. De feedback ging over wat goed ging of goed was gegaan	4.8	4.5	3.1	3.5	3.8	4.5	3.8	3.6	4.1	3.8	4.7	4.8	4.1
4. De feedback ging over wat nog niet goed was of ging	3.1	3.5	4.1	3.6	4.2	2.7	4.4	2.4	3.8	2.3	3.1	3.8	3.4
5. De feedback gaf precies aan wat goed ging, of anders moest	4.2	3.8	3.5	3.7	4.1	4.5	4.2	3.2	4.1	2.8	3.6	4.6	3.9
6. De feedback kwam positief over	4.6	4.4	3.9	4.0	4.2	5.0	4.8	3.4	4.5	4.0	4.3	4.8	4.3
7. De feedback kwam negatief over	1.4	2.1	2.1	1.9	1.8	1.0	1.2	2.4	1.5	2.8	1.9	1.0	1.8
8. De feedback ging over de uitvoering van de opdracht	4.2	4.3	4.1	3.7	4.1	3.8	4.2	4.0	3.9	4.3	4.7	5.0	4.2
9. De feedback ging over de aanpak van de opdracht	3.6	4.4	4.1	3.9	3.1	3.8	4.6	3.2	3.6	4.3	3.1	4.8	3.9
10. De feedback ging over mijn werkhouding	2.8	2.4	2.3	2.5	1.4	1.0	2.0	1.4	2.3	3.0	3.6	4.6	2.4
11. De feedback ging over mijn persoon	2.1	2.2	2.0	1.7	2.0	1.0	2.0	2.0	2.1	3.0	4.5	3.6	2.4
12. De feedback ging over de samenwerking	3.6	2.9	3.9	1.8	1.7	1.3	2.0	1.6	3.0	2.3	2.5	4.5	2.6
13. De feedback ging over wat ik al had gemaakt	2.8	3.2	3.3	2.8	4.0	1.0	4.4	3.4	2.9	4.0	2.4	4.4	3.2
14. De fb maakte duidelijk waar de opdracht aan moest voldoen	3.3	4.4	3.8	3.2	3.9	3.7	4.6	3.8	4.1	4.3	3.3	4.8	3.9
15.De feedback motiveerde me om verder te werken	3.7	3.8	3.5	3.4	3.4	3.8	4.2	2.8	4.3	3.8	3.0	4.6	3.7
16. De feedback was leerzaam	4.3	4.4	3.8	4.1	4.3	5.0	5.0	3.4	4.4	4.0	3.8	4.6	4.3
17. De feedback was bruikbaar	4.4	4.3	3.9	3.9	4.6	5.0	4.6	3.6	4.5	4.3	3.3	5.0	4.3
18. Door de feedback wist ik wat te doen om door te kunnen	4.1	4.2	4.0	3.5	3.2	5.0	4.0	4.0	4.3	4.0	3.5	4.6	4.0
19.De feedback gaf me de kans mijn werk te verbeteren	4.1	4.0	3.8	3.9	4.3	5.0	4.4	4.4	4.5	3.3	3.4	5.0	4.2
20. Met de fb gaf de leraar suggesties om me te verbeteren	3.0	4.1	3.6	2.7	3.4	4.3	4.4	3.8	3.9	3.3	3.7	4.6	3.7

* gewogen gemiddelden

Intenties, percepties en reflecties van de docenten bij terugzien van de fragmenten

In tabel 5.10 is zichtbaar gemaakt wat volgens de docenten de problemen waren waar de leerlingen tegen aan liepen, wat ze beoogden met de feedback die zij gaven en waarmee ze rekening hielden. Hieruit blijkt dat het probleem van de leerling dat aanleiding was voor de feedback er steeds op neerkomt dat de leerling iets niet begrijpt, iets onduidelijk vindt of ergens mee vastloopt. Dat wat er niet duidelijk is, kan

variëren: de betekenis van een begrip, de gegeven instructie, de te kiezen aanpak, een probleem tijdens de uitvoering, de planning. De docent kan ook uit zichzelf reageren, door te letten op de voortgang, het resultaat, de gemaakte fouten, of de ervaring of beleving van de leerling.
Gevraagd naar de intenties van de docent, komt een soortgelijke reeks intenties en activiteiten naar voren als eerder al vermeld: activeren, stimuleren, zelfstandigheid bevorderen, uitleggen, voordoen, voorbeelden geven, vragen stellen, bewustmaken, verbanden leggen, bevestigen, etc. Docenten blijken over een uitgebreid repertoire te beschikken en daarover te kunnen rapporteren.
In hun feedback houden ze rekening met het niveau van de leerling, zijn of haar mate van zelfvertrouwen, de persoon van en de persoonlijke relatie met de leerling, de specifieke situatie of problemen van de leerling, de uitdaging die de leerling nodig heeft of aankan, de werkhouding, taakgerichtheid en concentratie, en soms worden ook het doel of de eisen van de opdracht genoemd. (Dit correspondeert met de antwoorden die de docenten gaven op de soortgelijke vraag over waar ze bij het geven van feedback rekening mee houden in de lijst met open vragen die ze eerder in het onderzoek hebben beantwoord).
De effecten bij de leerling die de docenten beoogden, hebben vaak betrekking op toename van begrip of inzicht en het "weer verder kunnen" van de leerling. Soms worden ook genoemd bevestiging, zelfvertrouwen, afname van emoties die het leren in de weg staan. De redenen die worden genoemd om de interactie met een leerling af te ronden, zijn hiermee in overeenstemming: als de docent de indruk heeft dat de leerling het heeft begrepen, en als de leerling weer verder kan. Ook geven docenten aan dat ze de beschikbare tijd over de leerlingen willen verdelen. Docenten vinden het niet altijd eenvoudig te bepalen of de leerling het heeft begrepen (en weer verder kan); een aantal docenten meent dit te kunnen zien aan de blik en houding van de leerling, maar meerdere docenten zeggen ook actief te checken of de leerling het echt heeft begrepen.
De helft van de docenten bracht spontaan enkele persoonlijke reflecties in. Enkelen gaven aan dat zij vanwege tijdsdruk soms iets meer sturend waren dan zij zouden willen of geneigd zijn snel dingen over te nemen en (weer) te gaan uitleggen. Enkele anderen reflecteerden op de verschillen tussen vmbo- en vwo-leerlingen of op de balans tussen aandacht voor de onderlinge relatie en aandacht voor de inhoud. Enkele docenten formuleerden een leerpunt voor zichzelf: leerlingen meer vragen stellen, meer luisteren en minder direct reageren, soms alleen bevestigen in plaats van kritisch zijn.
De meeste van de genoemde docenten die spontaan enkele reflecties inbrachten, bleken in de door hen in het begin ingevulde vragenlijst hun eigen handelen tijdens het geven van feedback te onderschatten, vergeleken met de inschattingen van hun leerlingen in de door hen ingevulde vragenlijst vooraf, en/of hun eigen ideaal flink hoger te stellen dan hun eigen inschatting van hun huidige handelen (zie tabel 5.4). Mogelijk zijn deze docenten er op gericht dingen beter te doen dan ze nu doen.

Tabel 5.10: Inschattingen, intenties, redenen en percepties van de docenten, SR-interview

D	*Het probleem van de leerling*	*Intentie van de docent*	*Basis voor de reactie*	*Effect bij de leerling*	*Reden afronding interactie*
1	Niet begrijpen van de aanpak of het gebruik van de reken-machine.	Doorvragen om leerlingen zelf te laten nadenken; stap voor stap uitleggen, geven van concrete voorbeelden.	Niveau, mate van zelfvertrouwen, profielkeuze van de leerlingen.	Meer begrip.	Als is gecheckt of leerling het heeft begrepen.
2	Betekenis van begrippen, onduidelijke instructie, controle antwoorden.	Probeert open vragen te stellen, maar gaat snel over op sturende vragen en uitleg geven. Wil leerlingen laten nadenken en begrippen verduidelijken.	Sluit aan bij niveau. Zet in op relatie en de zwakkere leerlingen.	Meer begrip. Bevestiging.	Als de leerling het begrepen heeft of de goede vragen stelt. Wil ook de tijd verdelen.
3	Iets was niet duidelijk, samenwerking lukte niet, docent checkte voortgang.	Leerlingen zelf problemen laten oplossen en prioriteiten laten stellen. Verduidelijken bij problemen. Bemoedigen, stimuleren ergens trots op te zijn.	Zelfvertrouwen van een leerling. Stem en intonatie van de leerling. Eisen van de opdracht.	Weer zelf verder kunnen.	Als ze het gevoel heeft dat leerlingen zelf weer verder kunnen.
4	Vertaling lukt niet; samen-hang is vaag. Controle op voortgang; vraag over aanpak.	Leerlingen activeren en op goede spoor zetten; niet teveel weggeven. Bewust maken van denkfouten. Logisch nadenken stimuleren. Confronteren (werkhouding). Leerling zelf werk laten beoordelen.	Niveau, snelheid in denken, mate zelfvertrouwen, aanspreekbaar-heid leerling. Relevantie van de vraag.	Meer begrip, weer verder kunnen of aan de slag gaan.	Als het duidelijk is of als de leerling verder kan. Als leerlingen te veel achterlopen. Niet direct helpen.
5	Nabespreking opdracht, vragen over stof of aanpak.	Manier van werken laten begrijpen en laten leren van fouten. Stimuleren eerst bij anderen hulp te vragen. Dingen concreet maken en verbanden leggen tussen hoofdstukken.	Niveau, mate van concentratie of zelfvertrouwen, relatie, specifieke problemen, uitdaging.bieden	Meer begrip, weer verder kunnen, evt samen met anderen.	Als de leerling het begrepen heeft en verder kan. Als de leerling het eens is met hem (en niet gaat zitten mokken).
6	Begrip van de stof, moeite met lezen tekst, aanpak.	Wil door stellen van vragen leerlingen naar juiste antwoord leiden, geeft hints, tips en corrigeert. Wil dat leerlingen denkstappen maken. Benoemt verwachtingen en eisen	Type, aard en problematiek van de leerlingen, behoeften, relatie met de leerlingen.	Meer begrip en inzicht bij leerling.	Als leerling het begrepen heeft (let op blik). Checkt begrip bij twijfel.
7	Uitspraak Frans, schrijffouten.	Wil uitspraak horen en verbeteren. Wil dat leerlingen fonetisch aantekeningen maken en oefenen.	Relatie, gedrag leerling, uitdaging.	Verbeterde uitspraak.	Als uitspraak goed is en leerlingen verder kunnen.
8	Fouten in een gemaakte toets, aanpak.	Wil ingaan op wat goed ging en aanhaken bij probleem. Hakt som in stukjes en stelt vragen. Diept probleem uit. Benoemt stappen.	Begrip leerling, mate van zelfvertrouwen, voortgang.	Meer begrip en inzicht.	Als leerling het begrepen heeft (let op blik en houding). Checkt begrip.
9	Planning, werkstuk afmaken, on-tevredenheid met resultaat.	Leerling leren keuzes te maken, prioriteiten te stellen, dingen in perspectief te plaatsen, eigen werk te beoordelen, samen te werken.	Fase in proces, thuissituatie, zelfvertrouwen, de problematiek van de leerling.	Begrip over opties of antwoord op vraag.	Als leerling het begrepen heeft. Laat leerling soms echt zelf keuzes maken en wacht af.
10	Begrip, extra oefening, navraag stof komende toets.	Wil leerlingen denkstappen laten zetten en zelf tot het goede antwoord laten komen. Stimuleert uitleggen aan elkaar. Benadrukt: fouten maken mag.	Niveau, werkhouding en taakgerichtheid.	Meer begrip en inzicht door extra uitleg en herhaling.	Als leerling het begrepen heeft (let op de ogen) of verder kan. Checkt begrip bij twijfel.
11	Delen van goede/minder goede stage ervaringen.	Besteedt bewust aandacht en tijd aan leerlingen. Benoemt en geeft ruimte aan emoties, bevestigt, geeft tips, wil leerlingen positief benaderen.	Type leerling, emoties, situatie, doel.	Ruimte voor leren doordat emoties Afnemen.	Als het genoeg is of ze niet verder komt of als leerlingen verder kunnen.
12	Begrip, werkhouding.	Probeert leerlingen op goede spoor te zetten. Geeft tips en plaatst dingen in context. Stimuleert samenwerking, begrenst negatief gedrag.	Niveau, mate van zelfvertrouwen, problematiek, relatie.	Serieus aan het werk, begrip, zelf-vertrouwen.	Als leerling verder kan of echt niet wil. Rondt interactie soms niet af.

Reacties van de docenten in het eindgesprek op de terugkoppeling van de resultaten

De reacties van de docenten op de terugrapportage van de inschattingen van de leerlingen van hun gebruikelijke handelen (tabel 5.4, de docenten kregen de gegevens over henzelf) kunnen globaal in vier groepen worden verdeeld, met elk drie docenten:

A. De docent maakt expliciet eigen keuzes maar trekt ook lering uit het onderzoek (docenten 1, 3, 7).
B. De docent verantwoordt zich en verwijst daarbij naar de context (de tijd van het jaar, de opdrachten, de leerlingen, de school) (6, 8, 10).
C. De docent ervaart de informatie als bevestigend voor de eigen kwaliteiten (9, 11, 12).
D. De docent ervaart de informatie als enigszins confronterend (het raakt aan eigen twijfel, aan het idee alles onder controle te hebben, of aan een hoge inschatting van het eigen handelen) (2, 4, 5).

Als we deze indeling vergelijken met de eerder beschreven verschillen tussen het huidige en het gewenste handelen volgens de leerlingen, dan blijken de drie docenten in categorie B te behoren tot de docenten die op hun school relatief laag scoren op hun huidige handelen en waarbij het verschil tussen huidig en gewenst bij de leerlingen relatief groot is, terwijl de drie docenten in categorie C blijken te behoren tot de docenten die relatief hoog scoren op hun huidige handelen en waarbij het verschil tussen huidig en gewenst minimaal is.

Op het gedragsrepertoire in de opgenomen situaties zoals door de onderzoekers gescoord (zie tabel 5.8; de docenten kregen ook de uitgebreide tabel met de scores per observatiecategorie) reageerden drie docenten tevreden (5, 6 en 8): zij zagen een breed repertoire en waren daar blij mee. Drie docenten herkenden hun handelen in het repertoire zoals gescoord, zagen dit als bevestiging van hun eigen handelen zoals zij dat kennen en zien (2, 10 en 12). Vier docenten constateerden dat zij bepaalde vormen van feedback vaak of juist relatief weinig toepassen, herkenden dat ook, en gaven daarvoor een verklaring of nadere uitleg (3, 7, 9 en 11). Deze uitleg kwam er steeds op neer dat ze zich deze accenten bewust zijn en daar ook redenen voor hebben. Eén docent (docent 4) signaleerde dat de uitkomsten van de observatie door de onderzoekers vrij veel procesgerichte feedback laat zien, terwijl de leerlingen in zowel de vragenlijst vooraf als de vragenlijst direct na afloop juist aangaven dat hij weinig zulke feedback geeft. Kennelijk nemen zijn leerlingen niet al zijn procesgerichte feedback waar of nemen zij deze niet op die manier waar. Ten slotte was er nog een docent (docent 1) die in de scores zag dat hij een paar vormen van feedback niet vaak toepast en zich voornam daaraan meer aandacht te gaan geven.

Op de percepties van de leerlingen van de verkregen feedback blijkens de vragenlijst die is ingevuld direct na de opname (zie tabel 5.9), reageerden vijf docenten positief (docenten 1, 3, 8, 11, 12): zij waren daar blij mee en/of zagen dat als bevestiging van hun keuzen. Drie docenten konden de uitkomsten van de perceptievragenlijst niet meteen verklaren: de leerlingen namen niet veel procesgerichte feedback waar

(docenten 2 en 4), respectievelijk niet veel feedback op wat nog niet goed was (docent 7). Kennelijk herkenden (deze) leerlingen zulke feedback niet altijd. Twee docenten stelden vast dat de leerlingen niet alle kenmerken van hun feedback waarderen (docenten 9 en 10) maar zagen dit niet als probleem (hun gedrag was in hun ogen functioneel). Eén docent (6) heeft in het eindgesprek niet op de uitkomsten van de vragenlijst achteraf gereageerd.

5.5 Conclusies en discussie

Conclusies

We geven nu antwoorden op de gestelde onderzoeksvragen.

Welke aanwijzingen bevat de wetenschappelijke literatuur voor het geven van feedback?
In de bestudeerde literatuur worden kenmerken, inhouden en vormen van feedback genoemd die naar verwachting kunnen bijdragen aan de motivatie en het leren van leerlingen. De kenmerken betreffen de context en timing (drie kenmerken), de omvang en aard van de informatie (zes kenmerken) en de affectieve lading (drie kenmerken), waarbij per kenmerk een bepaalde keuze de voorkeur heeft of moet worden gezorgd voor een bepaalde balans c.q. juiste dosering. De inhouden betreffen de aandacht voor de taak, de prestatie en de standaard (in totaal vijf categorieën) waarbij eveneens een bepaalde keuze de voorkeur heeft of de genoemde soort inhoud in het algemeen de voorkeur heeft. De vormen betreffen een negental aanzetten tot potentieel leerzame verdere interactie (feedback als onderdeel van coaching).

Wat zijn de opvattingen en intenties van docenten met betrekking tot het geven van feedback?
De docenten zeggen bij het geven van feedback rekening te houden met zowel cognitieve (niveau van de leerling, kennis en begrip) als affectieve aspecten (zelfvertrouwen, persoonlijke situatie) en ze streven naar zowel cognitieve (kennis, begrip, resultaten) als affectieve doelen (zelfvertrouwen, motivatie). Ze vinden het belangrijk dat feedback wordt gegeven in persoonlijk contact met de leerling en in een situatie van vertrouwen en van voldoende tijd en aandacht, dat de feedback positief en constructief is, en dat leerlingen er actief bij worden betrokken.
In de interviews tijdens het bekijken van de opnamen gaven de docenten aan dat zij steeds feedback gaven in een situatie waarin een leerling iets niet begreep en/of niet goed wist hoe verder te gaan. De docenten benoemden diverse intenties en activiteiten. Zij hielden naar eigen zeggen bij hun feedback op verschillende manieren rekening met de leerling waar het om ging (niveau, zelfvertrouwen, problemen, persoonlijke situatie). Het ging hen er steeds om dat de leerling het begrijpt en weer verder kan, en zij rondden de interactie af als zij de indruk hadden dat dit het geval was.

Bij enkele docenten kwam in het eindgesprek naar voren dat sprake was van bepaalde verschillen tussen de inschattingen van hun leerlingen van hun handelen en/of hun eigen bedoelingen enerzijds en de observatiegegevens anderzijds.

Hoe geven docenten feitelijk feedback, volgens henzelf, volgens hun leerlingen, en volgens observaties door onderzoekers, en hoe hangt dit samen met hun opvattingen en intenties?

Volgens de docenten (gegevens vragenlijst) geven ze zowel feedback op eigen initiatief als naar aanleiding van vragen en problemen van leerlingen, en wel op inhoud, aanpak, uitvoering, resultaat, houding en gedrag, met behulp van een breed scala aan concrete activiteiten (feedbackgedragingen). Volgens de leerlingen (gegevens vragenlijst) geven hun docenten in het algemeen vrij veel en op gevarieerde wijze feedback, zijn zij daarbij iets meer gericht op het resultaat dan op het proces en hebben ze relatief minder aandacht voor het benutten van de feedback. Bij een aantal docenten komen de inschattingen van de docenten zelf van de mate waarin ze feedback geven overeen met de inschattingen van hun leerlingen, terwijl een deel van de docenten het eigen handelen, vergeleken met de inschattingen van de leerlingen, onderschat en een ander deel het eigen handelen overschat.

De opgenomen situaties betroffen uiteenlopende vakken en taken. De taken waren bijna allemaal gericht op leren en oefenen, in overeenstemming met het afgesproken focus van het onderzoek: het geven van feedback terwijl leerlingen zelfstandig aan het werk zijn (bij één docent ging het om het nabespreken van een gemaakte toets). Blijkens de observaties geven de docenten zowel feedback op eigen initiatief als op initiatief van de leerlingen en passen ze een breed scala aan activiteiten toe, waarbij zij iets vaker procesgerichte dan resultaatgerichte feedback laten zien. Vergeleken hiermee lijken de leerlingen de mate waarin de docenten procesgerichte feedback geven enigszins te onderschatten (mogelijk omdat zulke feedback vaak wordt gegeven in de loop van het interactieproces en daardoor door de leerlingen minder als feedback wordt gezien).

De leerlingen waardeerden, blijkens hun antwoorden in de direct na afloop ingevulde vragenlijst, de verkregen feedback over het algemeen positief (deze was duidelijk, terecht, leerzaam, bruikbaar) en wisten over het algemeen wat te doen om verder te kunnen. Hun percepties van de verkregen feedback gaven wel per docent een verschillend profiel te zien.

Bij enkele docenten kwam in het eindgesprek naar voren dat sprake was van bepaalde verschillen tussen de inschattingen van hun leerlingen en/of hun eigen bedoelingen in de opgenomen situaties enerzijds en de observatiegegevens anderzijds. Door twee docenten werd opgemerkt dat de leerlingen hen lager scoorden op procesgerichte feedback dan de mate waarin zij zelf dachten zulke feedback te geven en dan ook bleek uit de observaties.

De in de concept maps, antwoorden op open vragen en interviews verwoorde opvattingen en intenties van de docenten en hun handelen in de praktijk zoals gerapporteerd

in de vragenlijsten die door henzelf en door hun leerlingen zijn ingevuld en zoals geobserveerd door de onderzoekers, laten een vrij coherent beeld zien. De docenten benoemden diverse intenties en activiteiten en dit was conform de geobserveerde variatie in hun repertoire. Het ging hen er steeds om dat de leerling het begrijpt en weer verder kan, en zij rondden de interactie inderdaad af als zij de indruk hadden dat dit het geval was. De docenten die volgens hun leerlingen relatief veel feedback geven zijn voor een belangrijk deel docenten die bij het noemen van voorwaarden en doelen van feedback affectieve aspecten benadrukken. De docenten die bij feedback veel aandacht geven aan affectieve aspecten werden door hun leerlingen voor wat betreft hun feedbackactiviteiten relatief hoog ingeschat.
De opvattingen, intenties, aanpakken en reflecties van de docenten zoals die naar voren kwamen in de conceptmaps, vragenlijsten, interviews en eindgesprekken bestreken veel meer dan alleen de concrete feedbackactiviteiten die in het observatieformulier zijn opgenomen. Daar waar de docenten hun eigen concrete gedrag beschreven, kwam dit echter goed overeen met wat op de opnamen met de observatiecategorieën zichtbaar werd, terwijl hun opvattingen en intenties daar over het algemeen duidelijk bij aansloten. Slechts bij één docent bleek bij het terugzien van de gemaakte opname dat hij feitelijk veel minder het gesprek met leerlingen aanging dan hij zelf vooraf had aangegeven te doen en ook te willen doen. Hij constateerde dat hij zijn intenties niet waarmaakte, en hij schrok daarvan.

Hoe verhoudt zich het handelen van docenten op dit taakgebied in het algemeen tot hun handelen in een specifieke situatie (tijdens een lesuur in een klas)?
Over alle 12 docenten samen genomen gaven de docenten in de geobserveerde situaties feedback in een mate en variatie en met een gerichtheid die overeenkomt met de manier waarop de docenten en hun leerlingen hun gebruikelijke feedbackpraktijk voorafgaand aan de opnamen in de ingevulde vragenlijsten beschreven. De docenten bleken alleen relatief iets meer procesgerichte feedback te geven (zoals tot uitdrukking kon komen in de gebruikte observatiecategorieën) dan de leerlingen in de vragenlijst vooraf aangaven. Mogelijk namen de leerlingen niet alle procesgerichte feedback als zodanig waar of onthielden zij deze minder. Eén docent constateerde in het eindgesprek dat de leerlingen in de vragenlijst vooraf minder positief waren over zijn feedbackgedrag dan in de vragenlijst direct na afloop van de opname. Hiervoor is geen verklaring gevonden.

In hoeverre hangt de motivatie van de leerlingen voor het maken van taken voor het vak bij de betreffende docent samen met hun inschatting en waardering voor zijn handelen op dit taakgebied?
De breedte van het repertoire van het volgens de leerlingen gebruikelijke feedback-gedrag van de docent en de motivatie van de leerlingen voor het maken van taken voor het betreffende vak bij de betreffende docent hangen onderling significant positief samen. Dit kan worden opgevat als een bevestiging van de in de literatuur beschreven

invloed van feedback op motivatie, maar het kan ook zijn dat gemotiveerde leerlingen het pedagogisch-didactisch handelen van hun docenten beter waarnemen en onthouden, hetgeen zou aansluiten bij de in de literatuur beschreven invloed van motivatie op de perceptie en verwerking van feedback. Een nog andere interpretatie is dat docenten aan meer gemotiveerde leerlingen meer (diverse) feedback geven, mogelijk omdat hun werk daartoe aanleiding geeft en/of zulke leerlingen hen daartoe meer stimuleren.

In hoeverre hangen de verschillen tussen docenten in hun opvattingen en handelen op dit taakgebied samen met de scholen waar zij werken?
De opvattingen van docenten over feedback, de aspecten die zij daarbij onderscheiden en de intenties die ze verwoorden, verschillen individueel, ook tussen docenten op dezelfde school. Hun huidige handelen op dit taakgebied zoals gescoord door hun leerlingen in de eerste vragenlijst verschilt significant tussen de scholen (waarbij het gemiddelde op de scholen A en D hoger is dan op de scholen B en C), evenals het verschil tussen het huidige en het door de leerlingen gewenste handelen (dat op de scholen B en C groter is dan op de scholen A en D). Deze verschillen tussen de scholen zijn uiteraard gebaseerd op verschillen tussen de docenten die op de scholen werken en aan dit onderzoek deelnamen. Op de schalen in de data van de vragenlijst voor leerlingen na afloop van de opnames over de zojuist ontvangen feedback verschillen de scholen gemiddeld niet. Ook de resultaten van de interviews en de eindgesprekken tonen individuele verschillen tussen de docenten die zich niet duidelijk laten verbinden met interpreteerbare verschillen tussen de scholen.

Discussie

Uit het onderzoek komt een gedifferentieerd beeld naar voren van het denken en handelen van de docenten met betrekking tot feedback. Al met al lijken de docenten daarbij primair gericht op de leerlingen en de verschillen tussen de leerlingen. Geven van feedback staat voor hen in het teken van persoonlijk contact, vertrouwen en aandacht. Feedback moet voor alles positief en constructief zijn. De docenten willen graag rekening houden met het niveau, het zelfvertrouwen en de persoonlijke situatie van de leerling. Ze vinden het belangrijk leerlingen er actief bij te betrekken. Ze zijn verheugd als leerlingen hun handelen positief waarderen.
Het geven van positieve feedback en bevestiging kan de acceptatie van de feedback bij de leerlingen bevorderen. Wel lijkt het benoemen van wat (nog) niet goed is en van de standaard verhoudingsgewijs iets minder te gebeuren. Het geven van feedback bij deze 12 docenten in de onderzochte situaties lijkt niet sterk te zijn gebaseerd op een duidelijke beoordeling van een geleverde prestatie. Feedback blijkt meer een reactie op een probleem van of bij de leerling. Dit probleem is veelal dat een leerling iets niet begrijpt en (daardoor) niet verder kan. De feedback is er dan op gericht dat de leerling het wel begrijpt en weer verder kan.

In de literatuur wordt feedback op meer manieren benaderd: als formatieve beoordeling (zoals bij Crooks, 1988; Sadler, 1989), als informatie (zoals bij Kluger & DeNisi, 1996; Shute, 2008), als proces van communicatie en interactie (zoals bij Ilgen et al., 1979; Reynolds et al., 1996), als boodschap met impact (zoals bij Ilgen & Davis, 2000; Straub, 1997), en als onderdeel van coaching (zoals bij Chi et al., 2001; Collins et al., 1989). Het geven van mondelinge feedback terwijl leerlingen aan het werk zijn, past in de praktijk van deze docenten meer bij de benadering van feedback als proces van interactie en communicatie en als onderdeel van coaching, dan bij feedback als formatieve beoordeling of (eenmalige) informatie of boodschap. De feedback die ze geven is minder gebaseerd op beoordeling van een prestatie aan een duidelijke standaard en meer op ondersteunen van de leerling en van de voortgang. De feedback draagt hierdoor ook minder bij aan het bevorderen van zelfevaluatie door de leerlingen.

Het voorgaande laat het belang zien van het conceptueel specificeren van didactisch relevante constructen zoals feedback, aangezien de onderwijspraktijk van docenten op basis van een dergelijke specificatie meer nauwkeurig kan worden beschreven. Vergelijk bijvoorbeeld ook wat Van de Pol, Volman en Beishuizen (2010) hebben gedaan ten aanzien van het construct scaffolding.

Een mogelijke interpretatie van de hiervoor beschreven bevindingen is dat de docenten enerzijds sterk zijn gericht op de leerlingen en hun relatie met de leerlingen, en dat anderzijds hun mondelinge feedback in lessituaties is gericht op het (kunnen) doorwerken van de leerlingen c.q. het (gewenste) 'normale' lesverloop (vergelijk Bereiter, 1990; Brown & McIntyre, 1993).

De volgende specifieke bevindingen geven aanleiding tot aanvullende interpretaties.

Dat feedback in de vorm van voordoen/modelleren, de goede oplossing benoemen en de standaard aangeven relatief minder voorkomt, zou kunnen betekenen dat het repertoire van de docenten op dit vlak nog iets minder ver is ontwikkeld (dan in het geven van positieve feedback en bevestiging).

Dat leerlingen procesgerichte feedback iets minder lijken te herkennen en/of onthouden kan betekenen dat zij zelf meer zijn gericht op het (goede) resultaat dan op het proces of de aanpak. Het kan echter ook zijn dat procesgerichte feedback meer dan resultaatgerichte feedback min of meer is opgenomen in het doorgaande interactieproces tussen docent en leerling en daardoor minder opvalt.

Dat bij tenminste één docent, die veel en intensief feedback gaf, de leerlingen de hoeveelheid gegeven feedback achteraf minder hoog inschatten, kan mogelijk worden gezien als bevestiging van wat wel wordt beweerd, dat een docent ook teveel feedback kan geven.

De feedbackpraktijk van de docenten verschilt niet duidelijk tussen de vakken, maar is wel afhankelijk van de specifieke taak, de betreffende leerling en het ervaren probleem of de geleverde prestatie. Zoals een aantal docenten zelf aangaf, is feedback voor hen een kwestie van maatwerk.

De breedte van het geobserveerde feedbackrepertoire hing niet samen met de mate van onderwijservaring van de docenten. De verwachting van de schoolleiders en sommige

deelnemers dat docenten met weinig ervaring over een duidelijk beperkter repertoire beschikken, kwam niet uit.
Ten slotte is duidelijk geworden dat de intenties, afwegingen en activiteiten van de docenten met betrekking tot het geven van feedback voor elke docent een eigen, persoonlijk profiel laten zien. Ook in dit opzicht is het geven van feedback kennelijk maatwerk.
De opvattingen over feedback die de docenten in het onderzoek inbrachten, verschillen zowel tussen scholen als binnen dezelfde school, terwijl de visies die door de schoolleiders zijn geuit niet duidelijk tussen de scholen verschillen. Hierdoor kunnen deze opvattingen en visies niet per school één op één op elkaar worden gelegd en kan niet worden aangegeven in hoeverre de verschillen tussen de docenten moeten worden gezien als individuele verschillen dan wel als verschillen die schoolgebonden zijn.
Het in de vragenlijsten door de leerlingen aangegeven gebruikelijke feedbackgedrag van hun docenten verschilt significant tussen de vier deelnemende scholen bij alle vier hiervoor gebruikte schalen (geven van feedback, zorgen dat feedback wordt gebruikt, resultaatgerichte feedback, procesgerichte feedback). Op twee van de vier scholen (een meer traditionele havo/vwo-school die aan dit onderzoek deelnam met havo-leerlingen, en een sterk vernieuwende vmbo-school) waren de gemiddelde scores op alle vier schalen hoger dan op de andere twee scholen (beide sterk vernieuwende havo/vwo-scholen). Dit wijst vooralsnog niet op een duidelijk verschil in het repertoire van handelen van docenten tussen meer traditionele scholen (scenario 1/2) en sterk vernieuwende scholen (scenario 4).
Duidelijke beperkingen van het onderzoek waren dat het aantal docenten vrij klein was, waardoor de generalisatiemogelijkheden en de mogelijkheden om verschillen en verbanden te analyseren beperkt waren; dat de verschillen tussen vakken en tussen typen taken mede daardoor niet konden worden uitgediept; en dat de hoeveelheid verkregen en teruggekoppelde gegevens per docent de omvang van de feedback die zij zelf zinvol konden verwerken bij sommige docenten bijna te boven ging.
Vervolgonderzoek zou zich kunnen richten op het blootleggen van eventuele vakspecifieke aspecten en op het onderzoeken van verschillen in feedback tussen verschillende typen taken, van de effecten van pogingen van docenten om op grond van feedback op hun feedback hun feedbackrepertoire te verbreden en te differentiëren, en van de effecten van feedback op het leren van de leerlingen.
Mede door de kleine onderzoeksgroep kon niet worden vastgesteld in hoeverre sprake is van verschillen tussen docenten door verschillen tussen de scholen waar ze werken, of van verschillen tussen scholen door verschillen tussen de docenten die daar werken. Ook deze vraag verdient nader onderzoek.

6 Nader onderzoek naar leerzaam beoordelen

6.1 Inleiding

In het vorige hoofdstuk is een onderzoek beschreven naar het geven van mondelinge feedback terwijl leerlingen zelfstandig werken aan een taak. Feedback wordt meestal gegeven met de bewuste intentie bij te dragen aan het leren en presteren van leerlingen. In theorie ligt aan feedback altijd een beoordeling ten grondslag, van het proces (de aanpak die wordt gevolgd, het verloop van het proces) en/of van het resultaat in wording. Bij het geven van mondelinge feedback ontstaat echter al snel een interactieproces waarin docent en leerling met elkaar in gesprek zijn over de taak en de ervaringen en vragen van de leerling. De reacties van de docent zijn daarbij niet steeds gebaseerd op een duidelijke beoordeling; het beoordelingsaspect blijft vaak impliciet. Geven van feedback wordt dan primair gezien als onderdeel van de begeleiding, als docenttaak en als onderdeel van de leeromgeving en het onderwijsproces.
Docenten laten leerlingen regelmatig taken in de vorm van toetsen en opdrachten maken om te kunnen beoordelen wat zij hebben geleerd of om vast te stellen over welke kennis en vaardigheden zij op dat moment beschikken. Hierbij staat de beoordeling als docenttaak en als onderdeel van de leeromgeving en van het onderwijsproces centraal en blijft juist het feedbackaspect vaak impliciet. Als docenten over resultaten communiceren naar leerlingen, is dat om hen mee te delen wat zij hebben gehaald, maar ook vaak in de verwachting dat leerlingen daarvan leren. Beoordelingen zouden dan dezelfde functie hebben of tot dezelfde effecten kunnen leiden als feedback. In principe zouden leerlingen inderdaad ook van beoordelingen moeten kunnen leren, maar de vraag is of dit inderdaad gebeurt, en onder welke voorwaarden dan, en wat zij er dan van leren.
Gemeenschappelijk aan beide invalshoeken is dat leerlingen werken aan taken. Gezien vanuit het geven van feedback staat het leren voorop en blijft de beoordeling vaak meer op de achtergrond. Gezien vanuit het beoordelen staat het presteren voorop en krijgt het leren vaak minder aandacht. Dit neemt niet weg dat docenten en schoolleiders ook met het geven van beoordelingen kunnen beogen bij te dragen aan leren en verbetering. In dat geval kunnen we ook het geven van beoordelingen rekenen onder wat in het eerder uitgevoerde onderzoek is gedefinieerd als feedback: “een reactie op een prestatie op een taak met een standaard, gericht op leren en verbetering” (zie hoofdstuk 5). Het beoordelen van werk van leerlingen en terugkoppelen van de resultaten aan leerlingen wordt daarmee, naast het geven van instructie, het begeleiden van leerlingen en het geven van feedback, een essentieel onderdeel van het onderwijsleerproces. Dit past binnen de in de literatuur beschreven verschuiving van een testcultuur (assessment of learning) naar een assessmentcultuur (assessment for learning) (zie Askew & Lodge, 2000; Black & Wiliam, 1998; Gipps, 1994).

De leereffecten van feedback vallen echter al vaak tegen (zie hoofdstuk 5). Wat mag dan van het geven van beoordelingen worden verwacht, die vaak beperkt blijven tot het meedelen van een cijfer? In de literatuur over formatieve en summatieve assessment in het onderwijs wordt het onderscheid tussen feedback en beoordeling niet altijd scherp gemaakt (Black & Wiliam, 1998; Crooks, 1988; Harlen, 2006). De condities waaronder beoordelingen kunnen bijdragen aan leren zijn zowel in de literatuur als in de onderwijspraktijk minder systematisch verkend dan waar feedback als focus is genomen.
Naar aanleiding hiervan is een onderzoek gedaan om meer inzicht te krijgen in de manieren waarop en de condities waaronder beoordelingen kunnen bijdragen aan leren.
De vraagstelling van dit onderzoek luidt: Op welke wijze en onder welke condities kunnen beoordelingen door docenten van werk en prestaties van leerlingen voor de leerlingen leerzaam zijn?
De onderzoeksvragen zijn:

1) Welke aanwijzingen bevat de literatuur voor beoordelingen die kunnen bijdragen aan leren?
2) Hoe geven docenten hun beoordelingspraktijk vorm en met welke overwegingen?
3) Hoe ervaren leerlingen de beoordelingen die ze krijgen en hoe reageren zij daarop?
4) In hoeverre verschillen de beoordelingspraktijken van docenten en de ervaringen daarmee van leerlingen tussen vakken en typen taken en welke factoren spelen hierbij een rol?

6.2 Relaties tussen taken, beoordelingen en leren: theorie en onderzoek

Leerlingen kunnen op drie momenten leren van een taak. Ten eerste tijdens de voorbereiding op het maken van de taak, doordat de docent de taak toelicht en de leerlingen werken aan de kennis en vaardigheden die voor de taak nodig zijn. Ten tweede door het maken van de taak zelf, als die voldoende duidelijk en motiverend is en vraagt om activiteiten die zelf ook bijdragen aan verdere ontwikkeling en oefening van de kennis en vaardigheden. Ten derde na afloop, als het werk is beoordeeld en daarbij duidelijk wordt, door feedback en/of nabespreking, wat goed en minder goed was, waarom en waardoor, en hoe kan worden gewerkt aan verbetering, en als daarvoor ook gelegenheid wordt geboden.

Taken in soorten en maten

In zowel de onderwijspraktijk als de literatuur is een duidelijk onderscheid zichtbaar tussen taken waarbij de beoordeling kan plaatsvinden in termen van goed of fout (vaak aangeduid met toetsen) en taken met een meer open karakter waarbij een beoordeling nodig is in termen van kwaliteit, aan de hand van bepaalde criteria (vaak aangeduid met opdrachten). In beide gevallen kan behalve het resultaat ook het proces dat tot het resultaat heeft geleid bij de beoordeling worden betrokken, en in beide gevallen zijn voor de beoordeling ook bepaalde standaarden nodig (bij goed/fout taken uit te drukken in bepaalde kwantitatieve grensscores; bij open taken van meer kwalitatieve aard).

Publicaties waarin expliciet wordt ingegaan op voor onderwijzen, leren en beoordelen belangrijke kenmerken van taken, inclusief de formulering, de bijbehorende instructie en de aan het resultaat te stellen eisen, zijn niet heel talrijk (vergelijk Black en Wiliam, 1998), en voor zover beschikbaar divers.

In een inmiddels klassiek geworden artikel stelt Doyle (1983) dat schoolse taken verschillen in de aard van het gewenste resultaat, de denkprocessen die daarvoor nodig zijn (reproduceren van kennis, gebruiken van procedures, tonen van inzicht, beargumenteren van een standpunt), de mate van complexiteit en ambiguïteit, het inhoudsdomein, de te gebruiken bronnen, en de (overige) condities (beschikbare tijd, mogelijke samenwerking, leerklimaat).

Sindsdien is in het onderwijs meer aandacht gekomen voor de waarde van hogere orde denkprocessen en leerdoelen op dat terrein, en het laten werken van leerlingen aan open taken waarvoor zulke denkprocessen nodig zijn respectievelijk waardoor die kunnen worden opgeroepen. In lijn hiermee heeft zich op het vlak van toetsen en beoordelen een verschuiving voorgedaan van een testcultuur naar een assessmentcultuur, waarin enerzijds het werk van leerlingen niet meer eenduidig goed of fout is maar moet worden beoordeeld op kwaliteit en anderzijds deze beoordeling nadrukkelijker wordt geacht bij te dragen aan het leerproces.

Frederiksen en Collins (1989) hebben er in dit verband voor gepleit dat assessments '*systemic valid*' zijn. Daarmee bedoelden ze dat een toets, respectievelijk het maken daarvan, niet alleen de functie moet hebben om leerresultaten vast te stellen maar ook zelf moet bijdragen aan het (verder) ontwikkelen van die kennis en vaardigheden waarop het aan de toets voorafgaande onderwijs is gericht.

In lijn met Doyle en te zien als uitwerking daarvan hebben Elshout-Mohr en Van Hout-Wolters (1995) een aantal regelmatig voorkomende onderwijsleerepisodes beschreven: combinaties van de aard van het proces en van het resultaat en de mate van zelfsturing (reproductie/productie, kennis/vaardigheid, niet/wel metacognitief, nabije/verre transfer). Ook zij merken op dat de instructie, het leerproces en de opdracht op elkaar moeten zijn afgestemd. De beoordeling wordt door hen niet duidelijk uitgewerkt.

Nadien is door verschillende auteurs nader ingegaan op relaties tussen de aard en formulering van de taak en de eisen waaraan het resultaat moet voldoen, processen van leren en denken, de inrichting van de beoordeling, en de taken van de docent en de rol van leerlingen. Baxter en Glaser (1998) beschrijven twee dimensies waarop taken (in hun geval bij science) kunnen variëren: inhoud (smal/rijk) en proces (gesloten/open), en ze beschrijven voor elk van de vier combinaties de cognitieve processen die daarbij een rol spelen. Zo vragen taken met een smalle inhoud en een gesloten proces om het reproduceren van kennis of volgen van een procedure, etc. (vergelijk Doyle en Elshout-Mohr et al.). Taken met rijke kennis en een open proces zijn kenmerkend voor de meer complexe en authentieke taken die in het onderwijs tegenwoordig een grotere rol spelen. Baxter en Glaser leggen er de nadruk op dat voor een valide beoordeling op het al dan niet bereiken van gewenste cognitieve leerdoelen de docent moet weten welke cognities voor het succesvol uitvoeren van taken nodig zijn. De docent moet vervolgens taken

kiezen die ook alleen goed kunnen worden uitgevoerd via de gewenste cognitieve activiteiten en waarmee dit ook goed kan worden beoordeeld. De docent moet daarna bij het beoordelen de getoonde prestatie ook op een zodanige manier scoren dat daarin de beoogde complexiteit of kwaliteit van de cognitie tot uitdrukking komt.
Ook Black en Wiliam (1998) benadrukken dat optimaal uitdagende en betekenisvolle taken nodig zijn, dat docenten een goed beeld moeten hebben van het beoogde leerproces, en dat de taken die zij hun leerlingen laten maken daarop ook een gedifferentieerd zicht moeten bieden. Verder pleiten zij voor het actief betrekken van de leerlingen bij het proces van beoordelen, opdat zij goed begrijpen wat de taak inhoudt en wat van hen wordt verwacht, en waarop en op welke wijze er zal worden beoordeeld. Hoe belangrijk dit is, bleek in het onderzoek van Oshige (Oshige, 2003; zie ook Broekkamp & Van Hout-Wolters, 2007), die vond dat een goed begrip van de taak (op aspecten als doel, aanpak, tijd en bronnen, beoogde kwaliteit, en criteria en standaarden bij de beoordeling) positief samenhing met de prestatie. Hij concludeerde dat leerlingen last hebben van taken die onduidelijk zijn, en dat leraren leerlingen moeten aanleren niet meteen met een taak aan de slag te gaan maar zich eerst goed op de taak te oriënteren.
Volgens Hodgen en Marshall (2005) geldt ook dat hoe beter, respectievelijk duidelijker, een taak is, hoe beter daarop feedback kan worden gegeven. Juist bij meer complexe, open taken is dit van belang, omdat via feedback op een prestatie leerlingen meer begrip kunnen ontwikkelen van het beoogde concept van kwaliteit, dat vooraf moeilijk expliciet is aan te geven. Hodgen en Marshall vergeleken in hun onderzoek het formatief beoordelen van meer complexe opdrachten bij Engels en Wiskunde. Zij concludeerden dat het bij zulke taken gaat om de kwaliteit van het redeneren van de leerlingen (conclusies, argumenten) en dat zulke taken en de beoordeling daarvan tussen beide vakken veel overeenkomsten vertonen (de taken moeten voor de leerlingen duidelijk en betekenisvol zijn en de docenten moeten grondig inzicht hebben in het gewenste leerproces en hoe ze dat kunnen ondersteunen).

Beoordelen als onderdeel van het onderwijsleerproces

Dat beoordelen onderdeel is van het onderwijsleerproces en daarbij geschikte taken nodig zijn en beoordelingen voor leerlingen betekenis moeten hebben, komt goed tot uitdrukking in het cyclische stappenplan van Natriello (1987; zie ook hoofdstuk 5): bepalen van de doelen van de beoordeling, ontwerpen van geschikte taken, vaststellen van beoordelingscriteria, vaststellen van standaarden, verzamelen van gegevens over de prestaties van de leerlingen, beoordelen van de prestaties, geven van feedback, nagaan van de effecten bij de leerlingen.
Volgens Ilgen, Fisher en Taylor (1979) wordt feedback in de vorm van oordelen over gedrag en prestatie al lang onderkend als essentieel voor leren en motivatie, maar is nog weinig kennis verkregen over de effecten ervan. Zij schrijven dit toe aan het feit dat feedback zelf al een gecompliceerde stimulus is en daarnaast ook nog met veel andere factoren samenhangt, waardoor het bijna onmogelijk is de effecten ervan vast te stellen. Zij zien feedback als onderdeel van de interactie die nodig is voor leren, en als speciaal

geval van het algemene proces van communicatie waarbij een zender een boodschap overbrengt aan een ontvanger. De waarneming, de acceptatie en de reactie van de ontvanger worden beïnvloed door kenmerken van de zender, de boodschap en de ontvanger zelf. Volgens Larsen (1984) ontbreken bij veel taken objectieve prestatie-criteria en zijn daardoor anderen de voornaamste bron van feedback. De prestatie waarop feedback moet worden gegeven, is echter niet altijd duidelijk, en ook speelt een rol hoe goed de feedbackgever de ontvanger kent en daardoor de effecten van de feedback kan inschatten.

Volgens Crooks (1988), schrijvend over de praktijk van het beoordelen in schoolklassen, zijn veel "classroom tests" van een relatief laag cognitief niveau, en in die zin minder leerzaam. Het maken van toetsen (en de voorbereiding daarop) kan in het algemeen leerzaam zijn doordat leerlingen daardoor meer tijd, met gerichte aandacht en actief met de leerinhoud bezig zijn, en dit wordt bevorderd als taken leerlingen optimaal uitdagen. Aangezien leerlingen verschillen, zou dit betekenen dat bij het geven van taken moet worden gedifferentieerd. Ook Black en Wiliam (1998) zijn over de dagelijkse praktijk van assessment door docenten niet heel optimistisch. Volgens hen ontbreekt het leraren vaak aan voldoende tijd, inzicht en vaardigheid. Zij pleiten ervoor de effectiviteit van formatieve beoordeling te bevorderen door deze te integreren in bepaalde modellen of leeromgevingen. Dan nog zal volgens hen veel afhangen van de aard van de interactie tussen docent en leerlingen en de manier waarop de beoordeling c.q. feedback bij leerlingen overkomt en wordt verwerkt.

Het onderscheid tussen formatieve en summatieve beoordeling betreft primair het gebruik dat van resultaten van beoordelingen wordt gemaakt. In beide gevallen is het van belang dat de beoordeling betekenisvol is, en staat dus de validiteit voorop: de resultaten moeten weergeven wat leerlingen weten en kunnen. Bij summatieve beoordeling is daarnaast consistentie (vergelijkbaarheid, betrouwbaarheid) belangrijk en bij formatieve beoordeling moeten de resultaten kunnen worden gebruikt om er acties voor verder leren en verbeteren op te baseren (Wiliam & Black, 1996). Zowel bij formatieve als bij summatieve beoordeling kan de betekenis worden verduidelijkt door over de beoordeling interactie aan te gaan. Een summatieve beoordeling moet wel meer dan een formatieve beoordeling ook buiten de betreffende situatie een voor anderen voldoende duidelijke en gedeelde betekenis hebben (Black & Wiliam, 2006). Een summatieve beoordeling die veel stress genereert ("high-stakes") kan een sterk negatief effect hebben op de prestaties (en daardoor de validiteit van de beoordeling verminderen, vergelijk Messick, 1984), en ook op de motivatie en daardoor op het verdere leerproces (Harlen, 2005).

Kwaliteit van beoordelingen en condities voor leerzaamheid

In de literatuur wordt vaak gesteld dat docenten die na een beoordeling aan leerlingen alleen een cijfer verstrekken daarmee weinig bijdragen aan hun leerproces, omdat een cijfer op zich weinig informatief is over de sterke en zwakke punten van de geleverde prestatie en de manier waarop deze kan worden verbeterd. Dat is des te meer het geval waar beoordelingen tot stand komen door middel van vergelijking tussen leerlingen

(*norm-referenced*). Ook waar beoordelingen plaatsvinden op inhoudelijke gronden (*criterion-* of *standards-referenced*) worden cijfers door leerlingen zelf vaak gebruikt voor vergelijking.

Uit een onderzoek van Evans en Engelberg (1988) naar de manier waarop leerlingen ontvangen cijfers interpreteren, bleek dat oudere en beter presterende leerlingen gemiddeld beter begrijpen hoe cijfers tot stand komen dan jongere en minder goed presterende leerlingen, en hun prestaties ook meer toeschrijven aan hun eigen aandeel, in plaats van aan externe factoren waarop zij zelf geen invloed hebben. Dit zou betekenen dat cijfers voor oudere en beter presterende leerlingen potentieel leerzamer zijn dan voor jongere en minder goed presterende leerlingen. Evans en Engelberg vonden echter ook dat oudere leerlingen het behalen van goede cijfers steeds belangrijker gaan vinden, maar over de praktijk van het cijfers geven tegelijk steeds cynischer worden. Zij richten zich steeds meer op het halen van goede cijfers op zich, waardoor het mogelijke leereffect weer afneemt. De auteurs concluderen dat het belangrijk is dat docenten aan hun leerlingen duidelijk maken hoe beoordelingen tot stand komen, wat zij betekenen, en hoe de leerlingen eraan kunnen werken om hoge(re) cijfers te halen.

Het integreren van de beoordeling in het onderwijsleerproces, waarvan het belang hiervoor al aan de orde kwam, kan ook bijdragen aan de kwaliteit en de leerzaamheid ervan. Docenten doen er goed aan hun verwachtingen te verduidelijken door het geven van oefeningen, modelantwoorden, uitgewerkte voorbeelden en commentaar bij eerdere toetsvragen. Ook is het van belang dat leerlingen voldoende ervaring hebben kunnen opdoen met de betreffende taak, dat voldoende duidelijk is wat de bedoeling is, en dat ondersteuning en feedback voldoende substantieel en begrijpelijk zijn en niet te laat komen om nog te kunnen worden gebruikt (Hounsell, McCune, Hounsell & Litjens, 2008).

Als beoordelingen onderdeel gaan uitmaken van het onderwijsleerproces, vervaagt het onderscheid dat in de literatuur vaak is gemaakt tussen formatieve beoordeling (geven van feedback, gericht op leren en verbetering) en summatieve beoordeling (geven van oordelen over prestaties, al dan niet in de vorm van cijfers). Dit zal zich zeker voordoen bij beoordelingen van complexe, open taken, waarbij geen sprake is van goed of fout maar van een bepaalde kwaliteit. Volgens Sadler (1989) hebben docenten daarbij een begrip van kwaliteit nodig dat past bij de taak. De leerling moet dit begrip van kwaliteit ook ontwikkelen, zodat hij de eigen taakuitvoering al tijdens het proces kan volgen en beoordelen en over een repertoire beschikt van te ondernemen acties voor als er moet worden bijgestuurd. Onderwijs moet erop zijn gericht dat leerlingen hun eigen werk op kwaliteit kunnen beoordelen.

Recent heeft Sadler deze thematiek verder uitgewerkt (Sadler, 2009a, 2009b, 2010a, 2010b). De aard van de kwaliteit waaraan het resultaat moet voldoen, hangt mede af van het type taak dat aan de orde is (bijvoorbeeld in termen van cognitieve activiteiten). Van die kwaliteit kunnen criteria worden afgeleid (aspecten van kwaliteit). Het is dus niet andersom, dat het totaal van beoordelingen op criteria leidt tot een oordeel over de kwaliteit van het geheel, laat staan zonder die criteria aan elkaar te relateren en te wegen.

Dit is wat gebeurt als criteria voor elkaar mogen compenseren: een dergelijke beoordeling heeft geen duidelijke betekenis en is dus niet zinvol. Een docent kan ook niet zijn totale beeld van kwaliteit aan leerlingen overdragen. De enige optie om leerlingen zelf een dergelijk beeld te laten ontwikkelen, is door hen zelf ervaring te laten opdoen met beoordelen op kwaliteit, bijvoorbeeld met behulp van peer-assessment aan de hand van meerdere criteria waarbij de oordelen moeten worden gemotiveerd. Een andere voorwaarde voor zinvol beoordelen is volgens Sadler dat de beoordeling alleen betrekking heeft op de uitgevoerde taak en niet ook op andere kenmerken van de leerling of zijn gedrag. Als een beoordeling ook wordt gebaseerd op het voldoen aan eisen ten aanzien van aanwezigheid, tijdig inleveren, vormgeving of getoonde inzet, of compensatie wordt gegeven voor bepaalde belemmeringen of een beloning voor bijvoorbeeld originaliteit, of aanpassingen worden gemaakt met het oog op een gewenst percentage voldoendes, wordt de betekenis ervan onhelder.

Communicatie en betekenisgeving

Volgens Ilgen en Davis (2000) voldoet een prestatie van een leerling op een taak, of het nu gaat om de aanpak en uitvoering of om het resultaat, nu eenmaal wel of niet aan bepaalde standaarden, en moet de leerling als de prestatie (nog) onvoldoende is zich daarvan bewust worden om zich te kunnen verbeteren. Anders gezegd: als het niet goed is, moet je vertellen dat het niet goed is. Dat vinden mensen meestal niet leuk, en docenten vormen daarop geen uitzondering. Zij kunnen daardoor geneigd zijn de negatieve boodschap af te zwakken, vaag te houden of in te kleden. De ontvanger, de leerling, vindt het ook niet prettig om een negatieve boodschap te krijgen, want die wordt gemakkelijk ervaren als pijnlijk en bedreigend voor het zelfconcept, de zelfachting en het zelfvertrouwen. Indien dan toch een negatieve beoordeling wordt gegeven, zullen de ontvangers daarvan volgens Ilgen en Davis eerst kiezen of ze afhaken of hun prestatie willen verbeteren, voordat ze toekomen aan de keuze van de manier waarop ze die verbetering willen bereiken (door hun inzet te verhogen of hun aanpak te veranderen).

Andere affectieve factoren die tijdens het proces van communiceren van en betekenis geven aan beoordelingen bij de ontvanger een rol kunnen spelen, zijn behoefte aan autonomie en het gevoel dat de beoordelaar controle wil uitoefenen. Een beoordelaar kan controle willen uitoefenen door bijvoorbeeld dringend geformuleerde suggesties of adviezen. De dialoog tussen beoordelaar en beoordeelde is daardoor vaak niet helemaal vrij van aspecten van macht (Carless, 2006; Higgins, Hartley & Skelton, 2001). De docent bevindt zich in een lastige dubbelrol van beoordelaar en begeleider.

Shute (2008) vat experimentele onderzoeken samen met betrekking tot de inhoud, dosering en timing van feedback en de cognitieve processen die daarbij een rol spelen. Feedback heeft volgens haar over het algemeen een negatief effect als deze (te) kritisch of controlerend is, wordt gegeven in de vorm van cijfers (wat vooral onderlinge vergelijking oproept), weinig specifiek is, (te) veel informatie bevat, en wordt gegeven terwijl de leerling bezig is de taak uit te voeren en door de feedback wordt gestoord. Shute voegt hieraan toe dat het bij taken die bestaan uit het (leren) uitvoeren van

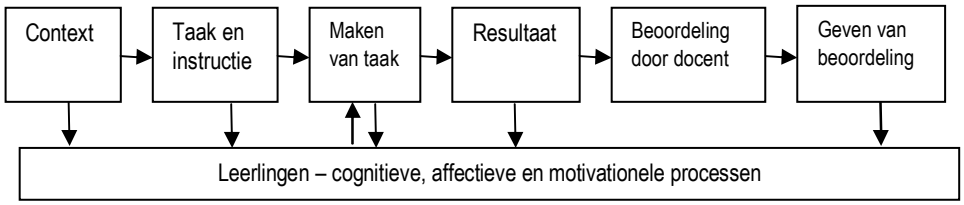
bepaalde procedures, op korte termijn effectief kan zijn om direct feedback te geven, maar op langere termijn juist minder effectief, omdat de transfer (toepassing in andere situaties) dan geringer is. (Mogelijk draagt het niet meteen feedback geven ertoe bij dat leerlingen leren om problemen die zich voordoen zelf op te lossen, waardoor ze dit ook in nieuwe situaties beter kunnen, vergelijk Bjork, 1994).
Vormen van feedback die volgens Shute (2008) vaak een positief effect hebben, zijn: meer specifieke informatie geven dan alleen goed/fout, het goede antwoord geven, en aangeven *waarom* iets goed of fout is. Dit is verklaarbaar door cognitieve mechanismen. Door informatie te geven over de afstand tussen prestatie en standaard wordt de onzekerheid over hoe goed je het doet, verminderd, en daarmee ook de cognitieve belasting (minder onzekerheid is minder belastend). Verder kan specifieke informatie nuttig zijn voor het corrigeren van misconcepties, procedurele fouten en ongeschikte aanpakken.

Motivationele consequenties
Het geven van beoordeling en feedback kan bij ontvangers daarvan ook van invloed zijn op hun motivatie. Ook hiernaar is al geruime tijd onderzoek gedaan. Zo beschrijft Butler (1987) dat mensen bij het uitvoeren van taken regelmatig informatie over hun presteren nodig hebben om hun betrokkenheid bij de taak in stand te houden. Die informatie moet niet leiden tot toenemende betrokkenheid bij het eigen ik ("*ego-involvement*"), zoals door cijfers en lof kan worden opgeroepen, want dat leidt de aandacht af van de taak. Uit door Butler uitgevoerde experimenten bleek dat feedback inderdaad diverse motivationele consequenties kan hebben. Positieve informatie kan, afhankelijk van de precieze formulering, zowel de taak- als de ego-betrokkenheid versterken. Individueel commentaar bij een prestatie is in het algemeen bevorderend voor de interesse en de (volgende) prestatie, maar het verkrijgen van lage cijfers doet beide afnemen.
Voortbouwend op deze en andere inzichten en onderzoeksresultaten, hebben Kluger en DeNisi (1996) een "Feedback Intervention Theory" (FIT) ontwikkeld. Zij nemen aan dat mensen doelen nastreven op verschillende hiërarchisch geordende niveaus en dat feedback de aandacht richt op een bepaald niveau. Ze onderscheiden drie niveaus: het (leren) uitvoeren van de taak, de motivatie voor de taak, en de eigen persoon. Ze stellen dat de aandacht normaal is gericht op het middelste niveau, want bezig zijn met zichzelf (het bovenste niveau) vinden mensen over het algemeen niet leuk, en aandacht voor de taak (het onderste niveau) is vaak niet nodig omdat het uitvoeren ervan min of meer vanzelf gaat. Kluger en DeNisi stellen dat specifieke en procesgerichte feedback de aandacht richt op het niveau van de taak, en dat is vaak wenselijk. De FIT biedt ook een verklaring voor het feit dat geven van een cijfer vaak geen goede vorm van feedback is, doordat een cijfer appelleert aan vergelijking met andere leerlingen, de aandacht daardoor richt op de eigen persoon en daarmee afleidt van de taak. Kluger en DeNisi gaan ook in op de manieren waarop ontvangers op feedback c.q. beoordeling kunnen reageren. Zij kunnen hun gedrag veranderen (harder werken en/of hun aanpak aanpassen), hun persoonlijke standaard veranderen (de lat voor zichzelf lager of hoger leggen), de

feedback verwerpen (bijvoorbeeld als niet geloofwaardig of niet op hen van toepassing), en zich helemaal afkeren van de taak.
Het effect van het geven van beoordelingen kan ook afhangen van de bedoeling van de beoordeling. Uit een aantal onderzoeken is gebleken dat *high-stake* summatieve toetsen bij leerlingen kunnen leiden tot negatieve affectief-motivationele effecten, zoals toenemende onzekerheid, spanning en angst, en afname van motivatie. Zwak presterende leerlingen kunnen hierdoor in een negatieve spiraal terecht komen, maar ook goed presterende leerlingen kunnen hiervan nadelige consequenties ervaren, als zij menen aan hoog gespannen verwachtingen in hun omgeving te moeten voldoen (Paris, Lawton, Turner & Roth, 1991).
De impact van beoordelingen door een docent in een klassensituatie is mede afhankelijk van het klimaat met betrekking tot beoordelingen in de klas bij deze docent. Brookhart en DeVoge (1999) noemen dit de "*classroom assessment environment*" en benadrukken dat de onderdelen en aspecten daarvan, zoals de taak die moet worden gemaakt, de instructie die daarbij is gegeven en de omstandigheden, vanuit de leerlingen gezien één geheel vormen, en daarom ook in hun context moeten worden onderzocht.
Harlen en Deakin Crick (2003) geven een overzicht van resultaten van onderzoek in Engeland rondom de invoering van het landelijke toetsprogramma in het kader van het National Curriculum. In het algemeen worden de resultaten op summatieve toetsen bij leerlingen gemakkelijk het primaire criterium waarop zij zichzelf en elkaar beoordelen. Na de invoering van landelijke summatieve toetsen bleek dat daardoor veroorzaakte subtiele gedragsveranderingen bij leraren in de beleving van de leerlingen het leerklimaat deden verschuiven van toetsen en beoordelen gericht op leren naar toetsen en beoordelen als bron van onzekerheid, spanning en angst (zie ook Harlen, 2006).
Uit de studie van Harlen en Deakin Crick (2003) blijkt verder dat hoog presterende leerlingen over het algemeen de cijfers die ze krijgen beter begrijpen, eerder denken dat de cijfers fair zijn, en er in hun motivatie minder door worden beïnvloed. Bij zwakke leerlingen, die minder vaak begrijpen waar de cijfers op zijn gebaseerd, kan dit leiden tot gevoelens van hulpeloosheid. Als leraren bij het geven van toetsen uitleggen wat hun bedoeling is en op het werk van de leerlingen bruikbare feedback geven, kan dit positief bijdragen aan hun zelfvertrouwen voor zulke taken. Anders gezegd: de manier waarop leraren met toetsen omgaan heeft invloed op de manier waarop leerlingen ermee omgaan. Summatieve beoordelingen kunnen positief bijdragen aan het leren van leerlingen als van hen niet meer wordt geëist dan zij aankunnen (bijvoorbeeld door te toetsen als ze eraan toe zijn), zij bij beslissingen over beoordelingen worden betrokken, zij leren om hun eigen prestaties zelf te beoordelen, een klimaat wordt ontwikkeld dat is gericht op leren en de eigen rol daarin, en de resultaten worden gebruikt om de voortgang zichtbaar te maken en niet voor vergelijking tussen leerlingen.

Een model

Gelet op de vraagstelling en de besproken literatuur kan een aantal voor dit onderzoek belangrijke aspecten als volgt in een model worden weergegeven (zie figuur 6.1).



Figuur 6.1: Een model van het beoordelingsproces

Samengevat komen uit de hiervoor besproken literatuur de volgende aanwijzingen naar voren voor het vergroten van de leerzaamheid van beoordelingen voor leerlingen:

- beoordelingen moeten passen in het beoogde leerproces en moeten daartoe zijn gericht op hetgeen moet worden geleerd; de focus op dat wat moet worden geleerd, en de rol van beoordelingen in het proces, moeten voor docenten en leerlingen duidelijk zijn en door hen worden gedeeld;
- om de vorderingen van de leerlingen in hun leerproces goed te kunnen beoordelen, moeten daarvoor geschikte taken worden gekozen of gemaakt (waar nodig verduidelijkt door aanvullende instructie); deze taken moet passen bij zowel het leerdoel (de kennis en vaardigheden die moeten worden geleerd) als de leerlingen (waar deze zijn in hun leerproces);
- beoordelingen moeten worden gebaseerd op valide oordelen over wat leerlingen weten en kunnen, duidelijk gerelateerd zijn aan de sterke en zwakke punten in hun presteren (af te meten aan duidelijke standaarden) en voldoende dekkend zijn (voor het leerdoel);
- om aan de hand van een taak het weten en kunnen van leerlingen valide te kunnen beoordelen, moeten leerlingen enige ervaring hebben met het betreffende type taak (door een plotseling nieuwe aard of vorm van de taak gaan leerlingen mogelijk minder presteren zonder dat dit kan worden toegeschreven aan de te beoordelen kennis en vaardigheden);
- de boodschap aan de leerlingen moet goed getimed zijn, gedoseerd, informatief, begrijpelijk, relevant en bruikbaar in het verdere leerproces.
- om de boodschap te kunnen begrijpen en te worden ondersteund om deze te gebruiken moeten leerlingen weten of kunnen inzien in hoe deze tot stand is gekomen, wat deze betekent, en hoe ze eraan kunnen werken om zich te verbeteren;
- er moet worden vermeden dat het (moeten gaan) maken van een taak en het verkrijgen van een beoordeling bijdraagt aan negatief geladen affectieve processen die afbreuk doen aan het gewenste leerproces (zoals toename van stress, afname van zelfvertrouwen, toename van faalangst of ervaren van controle ten koste van autonomie);
- de docent moet tot slot letten op de manier waarop de leerling de verkregen feedback of beoordeling ontvangt en benut;
- deze voorwaarden kunnen alleen worden gerealiseerd via processen van interactie tussen docent en leerlingen.

6.3 Opzet en uitvoering van het onderzoek

Onderzoeksopzet

Het onderzoek is uitgevoerd op drie scholen voor voortgezet onderwijs, bij per school drie docenten en per docent een aantal leerlingen. Eerst is de gebruikelijke beoordelingspraktijk van de docenten in kaart gebracht, door middel van een groepsinterview per school, het individueel invullen van een vragenlijst, en het door elke docent laten invullen van een vragenlijst door een 20-tal van zijn of haar leerlingen. Daarna is bij elke docent het laten maken en vervolgens beoordelen van twee of drie taken onderzocht. Steeds werd van de docent de betreffende toets of opdracht verkregen, eventueel voorzien van een toelichting, en de bij vier van zijn of haar leerlingen gegeven beoordelingen. Verder vond bij deze leerlingen een groepsinterview plaats over de taak en de ontvangen beoordeling en feedback en vulden zij daarover een vragenlijst in. Na deze trits van taken is tot slot is bij elke docent door de betreffende vier leerlingen een vragenlijst ingevuld over kenmerken van taken die maken dat zij ervan kunnen leren.

Instrumenten

De instrumenten voor dit onderzoek bestrijken de componenten van het model in figuur 6.1. Eerst zijn bij de docenten en leerlingen hun opvattingen en percepties rond het geven respectievelijk ontvangen van beoordelingen in kaart gebracht (in figuur 6.1: context) alsmede de motivatie van de leerlingen voor het maken van opdrachten voor het betreffende vak van de betreffende leraar (leerlingen). Vervolgens zijn de docenten en leerlingen bevraagd over twee of drie achtereenvolgende taken (taak en instructie) en het werken daaraan (maken van de taak; resultaat; leerlingen). Daarna zijn gegevens verzameld over de vastgestelde beoordelingen (beoordeling door docent), over de aan de leerlingen gecommuniceerde beoordelingen (geven van beoordeling, communicatie), en over de reacties van de leerlingen daarop (leerlingen: cognitieve en affectieve processen). Tot slot is aan de leerlingen gevraagd bij welke kenmerken van taken (inclusief het proces van voorbereiding, beoordeling, nabespreking en vervolgactiviteiten) zij volgens hen het meeste leren (leerlingen).

Groepsinterviews met docenten
In dit interview werden enkele algemene gegevens verzameld (persoonskenmerken, bereikbaarheid en beschikbaarheid) en daarna enkele open vragen gesteld over de organisatie en planning van het afnemen van toetsen en beoordelen van opdrachten, de aard van die toetsen en opdrachten, de manier waarop beoordelingen worden gegeven, en de afstemming op de collega's en op de leerlingen.

Vragenlijst voor docenten
Deze vragenlijst bevatte vier onderdelen:

a) De eigen beoordelingspraktijk: enkele vragen met voorgestructureerde antwoorden over de manieren waarop zij deskundigheid op het gebied van beoordelen hebben opgedaan en de momenten waarop zij met andere docenten spreken over het geven van beoordelingen, en over de in de komende periode geplande toetsen en opdrachten.
b) De eigen opvattingen, intenties en aanpakken daarbij: 10 open vragen in de vorm van af te maken zinnen, bijvoorbeeld: "De functie van het geven van een beoordeling is volgens mij: …".
c) De mate waarin zij als docent bij beoordelen bepaalde activiteiten uitvoeren en dat belangrijk vinden om te doen: 22 activiteiten bij beoordelen met een 5-punts Likertschaal (van 'helemaal niet' tot 'helemaal wel'), bijvoorbeeld: "De leraar bespreekt met de leerlingen waarop hun werk zal worden beoordeeld". Dit betreft de items van het taakgebied Beoordelen uit hoofdstuk 3, aangevuld met items over: het stimuleren van zelfreflectie en zelfbeoordeling bij leerlingen al dan niet in relatie tot de criteria (bijvoorbeeld: de leraar stimuleert de leerlingen kritisch te kijken naar hun eigen werk); het betrekken van leerlingen bij de beoordeling (bijvoorbeeld: de leraar betrekt de leerlingen bij het beoordelen van gemaakt werk); het vooraf aangeven van het belang van de stof, de criteria en standaarden (bijvoorbeeld: de leraar geeft aan waarom het belangrijk is om de stof te beheersen); en het nabespreken van de opdracht (bijvoorbeeld: de leraar gaat na de beoordeling in gesprek met de leerlingen over het resultaat). Alle items zijn in een min of meer logische volgorde gezet, te beginnen met de items die betrekking hebben op de situatie voorafgaand aan het maken van de taak en eindigend met de items die te maken hebben met de nabespreking van de taak.
d) De frequentie waarmee zij bepaalde typen reacties bij leerlingen waarnemen op het moment dat zij een beoordeling hebben gekregen (10 mogelijke reacties met een 5-punts Likertschaal (van 'zelden of nooit' tot 'meestal of altijd'), bijvoorbeeld: "Tevredenheid ("Dat is mooi, het is me gelukt" of "nu ben ik klaar")"). Leerlingen kunnen op diverse manieren reageren op een beoordeling die ze hebben gekregen en de reactie zal mede afhangen van het feit of de beoordeling positief of negatief van aard is en of deze overeenstemt met de eigen verwachtingen. Het is hierbij van belang op welke manier leerlingen de boodschap ontvangen (en accepteren of verwerpen), hoe ze de boodschap interpreteren en waaraan ze het resultaat toeschrijven (attributie), en welke eventuele actieconsequenties ze eraan verbinden (zie Evens & Engelberg, 1988; Ilgen, Fisher & Taylor, 1979; Ilgen & Davis, 2000; Kluger & DeNisi, 1996; Wiliam & Black, 1996). Leerlingen kunnen geen of weinig actie ondernemen, omdat ze de beoordeling verwerpen, deze een gevoel van tevredenheid oproept, of een gevoel van demoralisatie. Leerlingen kunnen ten tweede de interactie aangaan omdat ze de beoordeling niet begrijpen, denken dat deze niet klopt en daarom daarover in discussie willen, of om hulp vragen omdat ze niet weten hoe ze het aan moeten pakken om een betere prestatie neer te zetten. Leerlingen kunnen ten derde hun aanpak aanpassen, hun inspanning verhogen of verlagen of hun eigen standaard aanpassen (de lat voor henzelf hoger of lager leggen). Deze in totaal 10 mogelijke reacties zijn in de vragenlijst opgenomen.

Vragenlijst voor leerlingen vooraf

Deze vragenlijst had tot doel om de docenten en de vakken op deze aspecten te kunnen vergelijken, om te kunnen controleren in hoeverre de vier leerlingen per docent die zijn bevraagd na elke taak representatief zijn voor de leerlingen waarvoor de docenten onderwijs verzorgen, en om na te kunnen gaan in hoeverre de motivatie voor het vak, het gebruikelijke handelen van de docent, en de wensen van de leerlingen samenhangen met hun antwoorden op de vragen over de taken. De vragenlijst bevatte hiertoe vier onderdelen:

a) De self-efficacy (voor taken bij dit vak bij deze leraar). Dit betreft dezelfde schaal als in hoofdstuk 5.

b) De motivatie (voor taken bij dit vak bij deze leraar). Dit betreft dezelfde schalen als in hoofdstuk 5.

c) Het huidige en gewenste handelen van de leraar wat betreft het laten werken aan opdrachten en het geven van feedback. Hiervoor zijn drie onderdelen gebruikt uit twee vragenlijsten die in eerdere onderzoeken zijn ontwikkeld (zie de hoofdstukken 3 en 5). Dit betreft 6 items over het laten werken aan opdrachten, 11 items over procesgerichte feedback en 9 items over resultaatgerichte feedback. Hieraan zijn op inhoudelijke gronden twee nieuwe items toegevoegd (ter aanvulling op het item "De leraar geeft feedback als we een fout hebben gemaakt" is toegevoegd "De leraar geeft feedback als we iets goed hebben gemaakt", en ter aanvulling op "De leraar vergelijkt het werk met hoe een leerling eerder opdrachten heeft gemaakt" is toegevoegd "De leraar vergelijkt het werk van leerlingen met elkaar").

d) Het huidige en gewenste handelen van de leraar wat betreft het beoordelen van werk en prestaties van leerlingen. Hiervoor zijn dezelfde items gebruikt als in de vragenlijst voor docenten (zie hiervoor).

Groepsinterview bij vier leerlingen per docent na afloop van elke taak

Dit interview bestond uit een aantal korte open vragen over het hele proces, vanaf de voorbereiding op en de ervaren duidelijkheid van de toets of opdracht tot de eventuele nabespreking en wat ze van de toets of opdracht hadden geleerd. In het geval van een toets betrof het 10 vragen (bijvoorbeeld: "Wist je vooraf precies waar de toets over zou gaan?" en "Had je tijdens het maken van de toets voldoende tijd?") en bij een opdracht 15 vragen (bijvoorbeeld: "Was de opdracht gemakkelijk of moeilijk?" en "Heeft de leraar je laten terugkijken op de opdracht of hoe je die hebt aangepakt?").

Vragenlijst voor de vier leerlingen per docent na afloop van elke taak

Deze vragenlijst bevatte over zowel de ontvangen beoordeling als de eventueel verkregen feedback een aantal items met een antwoordschaal en een aantal losse vragen.

Over de beoordeling zijn 10 items opgenomen over de eerste reacties van de leerlingen na het ontvangen van de beoordeling, dezelfde reacties als in de vragenlijst voor docenten (zie hiervoor), nu met een 7-puntsschaal. Daarna volgden zes losse vragen over eventuele andere reacties ten aanzien van de beoordeling en over de gemaakte toets of opdracht

(zoals: Had je deze beoordeling verwacht? Vond je de toets of opdracht moeilijk? Sloot de inhoud aan bij wat je eerder hebt gehad bij dit vak?).
Over de feedback werd eerst een aantal losse vragen gesteld (zoals: Had je zelf om feedback gevraagd? Wanneer kreeg je feedback? Hoe kreeg je feedback?). Daarna volgden 20 items over hoe de leerling de feedback had ervaren; dit zijn dezelfde items als gebruikt in hoofdstuk 5 in de vragenlijst direct na de video-opname. In het onderhavige onderzoek staat de perceptie van leerlingen van de verkregen feedback vanuit het perspectief van leerzaamheid centraal en dit leidt mogelijk tot andere schalen.

Afsluitende vragenlijst voor de vier leerlingen per docent
Deze vragenlijst bevatte twee vragen met elk een serie items, ontleend aan de bestudeerde literatuur over kenmerken van taken, aangevuld op basis van eigen inzicht en ervaring van de onderzoekers.
De eerste vraag betrof wanneer (bij welke kenmerken van de taak) en hoeveel de leerling leert bij het maken van taken voor het betreffende vak. Hierbij zijn 16 items opgenomen, zoals: (Ik leer ervan) als de leraar vooraf zegt wat de bedoeling is bij de taak, (Ik leer ervan) als ik tijdens het maken van de taak nog eigen keuzes kan maken, (Ik leer ervan) als de leraar na afloop zegt wat er goed ging en wat niet, en waarom, (Ik leer ervan) als ik wat ik heb geleerd van de taak kan toepassen bij een volgende taak. De leerlingen konden antwoorden op een schaal van 1 ‘weinig’ tot 4 ‘veel’.
De tweede vraag betrof twee te onderscheiden vragen: bij welke taken (bij welke kenmerken van de taak in het algemeen, dus niet specifiek bij het betreffende vak) de leerling zichzelf beter kan voorbereiden, en bij welke taken de leerling er meer van kan leren. Hierbij zijn 10 items opgenomen, zoals: Het is duidelijk waar de taak over gaat, Het is duidelijk hoe de taak kan of moet worden aangepakt, Het is duidelijk aan welke eisen het resultaat moet voldoen. De leerlingen konden bij de eerste subvraag antwoorden op een schaal van 1-3 (niet zo goed, gaat wel, goed) en bij de tweede subvraag eveneens op een schaal van 1-3 (niet zo veel, gaat wel, veel).

Respondenten

Deelnemende scholen, selectie leerjaar en vakken, werving docenten
Twee van de drie scholen (A en B) werken met een zeer vernieuwend onderwijsconcept en bestaan pas een jaar of vijf. De derde school (C) bestaat al geruime tijd. School A is een school voor vmbo-t, havo en vwo en biedt veel onderwijs thematisch aan (zes thema’s per jaar), met uitzondering van gymnastiek, wiskunde en de moderne vreemde talen. Binnen de leergebieden wordt veel gewerkt met complexe opdrachten, los van bestaande methodes. School B is een havo/vwo school met veel aandacht voor het ontwikkelen van vaardigheden, onder meer voor samenwerken en zelfverantwoordelijk leren. De school werkt in de onderbouw grotendeels in leergebieden, waarbij de talen zoveel mogelijk in de thema’s worden geïntegreerd; wiskunde wordt apart gegeven, in een eigen dagdeel. School C is een school voor vmbo-t, havo en vwo die werkt met

vakken en lesuren. In de vakken wordt veel gewerkt vanuit bestaande methoden (schoolboeken). Wel is tijd gereserveerd voor onderzoek en presentaties door leerlingen.
Omdat beoordelingssituaties ook kunnen worden beïnvloed door het type taak en de vakspecifieke kenmerken daarbij, werd in overleg met de schoolleiders besloten voor dit onderzoek docenten te werven van drie verschillende vakken: een moderne vreemde taal, een kennisvak, en een exact vak. Voor de onderlinge vergelijkbaarheid werd afgesproken op elke school dezelfde drie vakken te kiezen. Gelet op de belangstelling en beschikbaarheid van docenten is in onderling overleg gekozen voor de vakken Engels, Aardrijkskunde en wiskunde. Omdat een van de drie scholen een havo/vwo-school is, is voor de onderlinge vergelijkbaarheid voor leerlingen van havo/vwo gekozen. Tot slot hebben de schoolleiders gekozen voor deelname met het derde leerjaar, omdat de leerlingen dan zijn gewend aan de school en de opleiding die ze volgen, maar druk vanuit het examenprogramma dan nog geen rol speelt.
De schoolleiders hebben de docenten persoonlijk benaderd en gevraagd mee te willen werken aan het onderzoek. Dit resulteerde in een groep van negen docenten (drie per school) met 2 tot 7 jaar ervaring als docent (gemiddeld 3,5) waarvan 1 tot 7 jaar op de betreffende school (gemiddeld 3). Drie van de negen docenten waren vrouw (twee van de drie docenten Engels en een van de drie docenten Aardrijkskunde). Na de groeps-interviews met docenten bleek het op school A bij de wiskundedocent helaas toch niet realiseerbaar om leerlingen bij het onderzoek te betrekken.

Keuze per docent van de groep leerlingen en selectie daaruit van vier leerlingen, en keuze van taken

De deelnemende docenten hebben per school onderling afgestemd bij welk deel van de groep leerlingen of welke klas uit de havo/vwo afdeling zij de vragenlijst vooraf voor leerlingen zouden laten invullen, zodat er geen overlap zou zijn in de leerlingen tussen de docenten. Deze vragenlijst is ingevuld door 184 leerlingen: 38 leerlingen op school A, 76 leerlingen op school B en 70 leerlingen op school C.
Acht docenten hebben elk drie taken (toetsen en/of opdrachten) gekozen die de leerlingen in het najaar van 2010 zouden gaan maken. Bij enkele docenten bleek de derde taak pas te worden gemaakt na afloop van de onderzoeksperiode. In totaal konden bij acht docenten 21 toetsen en opdrachten in het onderzoek worden betrokken, zie figuur 6.3 verderop. Meestal kon de toets of opdracht zoals die, op papier of via de digitale leeromgeving, aan de leerlingen is verstrekt, worden verkregen.
Vervolgens hebben deze docenten elk uit de groep leerlingen die voor hem of haar de vragenlijst heeft ingevuld vier leerlingen geselecteerd voor de onderzoeksactiviteiten rond de taken, namelijk, met het oog op voldoende spreiding over het prestatieniveau van de leerlingen: een zwakke leerling, twee gemiddelde leerlingen en een goede leerling. Deze 32 leerlingen hebben meegedaan (met incidentele afwezigheid door ziekte) aan de groepsinterviews en de vragenlijsten per taak.
Bij 18 van de 21 taken zijn bruikbare groepsinterviews met leerlingen gerealiseerd (drie keer bleek de geluidsopname te zijn mislukt). Bij de taken die in groepjes waren

gemaakt, hadden de leerlingen die aan deze interviews meededen meestal in verschillende groepjes gewerkt. De vragenlijst aan het eind van het onderzoekstraject is ingevuld door 29 leerlingen (drie leerlingen waren door ziekte afwezig).
De geselecteerde groep van 32 leerlingen bleek in analyses op de verzamelde data een representatieve steekproef te vormen uit de totale groep van 184 leerlingen: slechts in 3% van de per docent uitgevoerde t-toetsen op alle beschikbare schalen was sprake van een significant verschil in gemiddelde. Ook bleek uit analyses van de van de docenten verkregen beoordelingen van gemaakte toetsen en opdrachten dat het bij de vier leerlingen per docent inderdaad gaat om een bij het betreffende vak zwak presterende leerling, twee gemiddeld presterende leerlingen, en een goed presterende leerling.

Analyses

De verwerking en analyse van kwalitatieve data
De groepsinterviews met de drie docenten per school zijn opgenomen en uitgewerkt. De gegevens over de organisatie van de toetsen zijn per school samengevat. De antwoorden op de vragenlijst voor docenten zijn beschreven. Deze teksten zijn gebaseerd op bewoordingen en gegevens van de docenten zelf en niet geabstraheerd of geïnterpreteerd, en naar hen teruggekoppeld (*member-check,* zie Boeije, 2008). Enkele docenten gaven een aanvulling door en deze aanvullingen zijn in de tekst in 6.4 verwerkt.
De kenmerken van de door de docenten verstrekte toetsen en opdrachten en de antwoorden zoals verkregen in de groepsinterviews met vier leerlingen per docent per taak zijn per taak samengevat in een gestructureerd schema met een omvang van een halve tot een hele bladzijde. De rubrieken in het schema waarin de antwoorden van de leerlingen zijn samengevat, zijn samengesteld uit combinaties van inhoudelijk bij elkaar behorende vragen zoals die in de groepsinterviews zijn gesteld. Zie figuur 6.4 (in paragraaf 6.4) voor een voorbeeld van een ingevuld schema van een opdracht.

De verwerking en analyse van kwantitatieve data
De gegevens die zijn verkregen met de voorgestructureerde vragenlijsten zijn ingevoerd en geanalyseerd in SPSS. Eerst zijn bij alle itemsets die waren voorzien van antwoordschalen schaalanalyses uitgevoerd (criteria: itemrestcorrelaties ten minste .20, liefst ≥.35; Cronbach's alpha ten minste .60, liefst ≥.70). Daarna zijn van alle schalen de gemiddelden en standaarddeviaties berekend. Met t-toetsen en enkelvoudige variantie-analyses is nagegaan in hoeverre sprake is van statistisch significante verschillen tussen scholen, tussen vakken, tussen docenten en tussen groepen leerlingen (significantieniveau .05; in analyses met kleine aantallen aangepast tot .10, dit wordt ter plekke vermeld). Tot slot zijn correlaties berekend en padanalyses uitgevoerd (met LISREL, versie 8.30) om de relaties te exploreren tussen het door de leerlingen waargenomen handelen van hun docent, hun motivatie voor het maken van taken voor het vak bij deze docent, en de voor de gemaakte taken behaalde cijfers.

Schalen in de vragenlijsten voor leerlingen: betrouwbaarheid, scorebereik, gemiddelde, spreiding

Bij alle schalen was gemiddeld per item sprake van ongeveer 1% item-nonrespons. Per schaal waren de ontbrekende waarden steeds verspreid over items en over leerlingen, en ging het om een klein percentage van de leerlingen (gemiddeld minder dan 10%) en per betreffende leerling om slechts om één item of hooguit enkele items. Alle schaalanalyses zijn eerst uitgevoerd op de data zoals verkregen, dus per schaal alleen over de leerlingen waarvan de betreffende gegevens compleet waren. Vervolgens zijn per schaal de ontbrekende waarden vervangen door de gemiddelde score op de items in die schaal (afgerond op een halve schaalpunt). Op het resulterende bestand zonder ontbrekende waarden zijn alle schaalanalyses herhaald. Bij alle schalen was het verschil in berekende Cronbach's alpha tussen de eerste en de tweede analyse maximaal éénhonderdste. Daarna zijn alle schaalgemiddelden tussen beide versies van het bestand (met en zonder ontbrekende waarden) vergeleken. Bij één schaal bleek het verschil 0,2 schaalpunt te zijn, bij vier schalen 0,1 schaalpunt en bij de overige 14 schalen kwamen de gemiddelden precies overeen. Op grond van de uitkomsten van deze vergelijkingen is besloten de analyses en rapportage verder te baseren op bestand zonder ontbrekende waarden, dus over alle 184 leerlingen.

Schalen in de vragenlijst vooraf (n=184)

De schalen voor *self-efficacy* en de motivatieconstructen in de SIMS waren intern consistent. Om te kunnen beschikken over een samenvattende schaal voor motivatie is nagegaan of de afzonderlijke schalen samen één schaal vormen. De eerste drie schalen (voor self-efficacy, intrinsieke motivatie en geïdentificeerde regulatie) vormen samen een voldoende betrouwbare schaal, met Cronbach's alpha .70 en gemiddelde 4,7 (op de betreffende schaal van 1-7). Zie tabel 6.1.
De 6 items over de percepties van de leerlingen van het handelen van de docent bij het *laten werken aan opdrachten,* de 11 items over *geven van procesgerichte feedback* en de 11 items over *geven van resultaatgerichte feedback* bleken geen voldoende betrouwbare schalen te vormen. Een factoranalyse op alle 28 items, gevolgd door nieuwe schaalanalyses, resulteerde in drie voldoende betrouwbare schalen, met dezelfde inhoud als beoogd en hiervoor genoemd maar met respectievelijk 7, 7 en 9 items. Zie tabel 6.2.

Tabel 6.1: Psychometrische informatie over de schalen voor het meten van motivatie voor het maken van opdrachten voor het vak van de betreffende docent (k = aantal items) (schaal 1-7) (N=184)

Schaal	*k*	*Voorbeelditem*	*r_{ir}'s*	*α*	*min*	*max*	*gem*	*sd*
Self-efficacy	8	Ik ben vol vertrouwen dat ik mijn opdrachten en toetsen bij dit vak (of leerdomein) heel goed ga maken	.70-.80	.93	1.1	7.0	5.1	1.1
SIMS	Waarom span je je in voor de opdrachten bij dit vak (of leerdomein) bij deze leraar?							
Intrinsieke motivatie	4	Omdat ik denk dat het werken aan deze opdrachten interessant is	.61-.80	.87	1.0	7.0	3.9	1.4
Geïdentificeerde regulatie	3*	Omdat ik denk dat deze activiteit belangrijk voor me is	.32-.53	.65	1.7	7.0	5.2	1.1
Extrinsieke regulatie	4	Omdat het van me verwacht wordt	.45-.57	.70	1.5	7.0	4.9	1.1
A-motivatie	4	Ik weet het niet, ik zie niet wat het werken daaraan mij oplevert	.53-.67	.79	1.0	7.0	3.0	1.2

* Eén item ("Omdat ik het zelf besloten heb"), met een itemrestcorrelatie van .18, is verwijderd.

Voor het meten van aspecten van *beoordelen van het werk en de prestaties van leerlingen* waren in totaal 22 items opgenomen. De 11 items voor het huidig handelen van de docent op het taakgebied beoordelen (zie hoofdstuk 3) bleken opnieuw een betrouwbare schaal te vormen, maar om na te gaan of dit domein in onderhavig onderzoek meer gedifferentieerd zou kunnen worden gemeten, is op alle 22 items een factoranalyse uitgevoerd (principale componentenanalyse met varimax-rotatie, met als besliscriteria factorlading ≥ .40, eenvoud van structuur en goede interpreteerbaarheid in het licht van dit onderzoek). Bij extractie van vier factoren bleken twee items op twee factoren te laden (het ene item met ladingen .41 respectievelijk.49, het andere met .48 respectievelijk .61). In beide gevallen is het item toebedeeld aan de factor met de hoogste lading. Dit waren de enige afwijkingen van de beoogde eenvoud van structuur, de vier factoren bestreken alle 22 items, de factoren bleken goed interpreteerbaar, en de schalen bleken betrouwbaar.

Deze schaalanalyses zijn in eerste instantie uitgevoerd op de antwoorden van de leerlingen op de vraag over het huidig handelen omdat tussen leerlingen daarover eerder c.q. meer overeenstemming valt te verwachten dan bij de vraag naar het gewenste handelen. Schaalanalyses op de overeenkomstige schalen voor het gewenste handelen leidden inderdaad in alle gevallen tot lagere alpha's maar alle schalen op één na waren ook nu voldoende betrouwbaar. Bij die ene schaal voldeed na verwijdering van één item met een te lage item-restcorrelatie (< .20) de alpha precies aan het gestelde minimum (.60). Zie tabel 6.2.

De schalen voor het huidig handelen vormen samen ook een schaal (Cronbach's alpha .85, gemiddelde 3,2), evenals de schalen voor het gewenste handelen (Cronbach's alpha .78, gemiddelde 3,7).

Tabel 6.2: Psychometrische informatie over de schalen voor het meten van het huidige en gewenste handelen van de docent rond het laten werken aan opdrachten, het geven van feedback en het beoordelen, volgens de leerlingen (schaal 1-5) (N=184)

*Schaal**	*k*	*Voorbeelditems*	*α*	*min*	*max*	*gem*	*sd*
Geven van feedback							
Bespreking vooraf van de opdracht**	7	De leraar geeft aan waarom het belangrijk is deze opdracht te maken.	.73	1.7	4.6	3.4	.66
Bespreking vooraf van de opdracht	*7*	De leraar bespreekt met de leerlingen hoe een opdracht kan worden aangepakt.	.63	2.4	5.0	4.0	.56
Procesgerichte feedback	7	Tijdens het werken geeft de leraar feedback op de aanpak van de opdracht.	.81	1.4	4.7	3.4	.66
Procesgerichte feedback	*7*	De leraar laat leerlingen terugkijken op een opdracht en hoe ze die aanpakten.	.79	1.6	5.0	3.8	.58
Resultaatgerichte feedback	9	We krijgen achteraf feedback op het resultaat van een opdracht.	.74	1.6	5.0	3.8	.63
Resultaatgerichte feedback	*9*	De leraar geeft aan waarom een antwoord goed of fout is.	,65	1.9	5.0	4.1	.59
Beoordelen							
Bespreking vooraf van criteria en standaarden	7	De leraar bespreekt met de leerlingen waarop hun werk zal worden beoordeeld.	.75	1.4	4.7	3.3	.69
Bespreking vooraf van criteria en standaarden	*7*	De leraar laat de leerlingen nadenken over wanneer iets goed is. De leraar geeft vooraf aan wat de leerlingen precies moeten beheersen.	.64	2.3	5.0	3.8	.62
Betrekken van leerlingen bij beoordeling	4	De leraar laat de leerlingen meebeslissen over waarop zal worden beoordeeld.	.77	1.0	4.5	2.2	.93
Betrekken van leerlingen bij beoordeling	*4*	De leraar betrekt de leerlingen bij het beoordelen van elkaars werk.	.73	1.0	5.0	2.9	1,0
Ook beoordelen van aanpak, proces en progressie	4	De leraar beoordeelt ook de manier waarop de leerling heeft gewerkt.	.67	1.3	5.0	3.4	.76
Ook beoordelen van aanpak, proces en progressie	*3*	Als de leraar werk beoordeelt, let hij\zij er ook op of het al beter is dan eerder. Bij groepswerk beoordeelt de leraar ook de manier waarop is samengewerkt.	.60	1.0	5.0	3.7	.84
Nabespreking van opdracht en beoordeling	7	De leraar bespreekt de beoordeling klassikaal. De leraar legt leerlingen uit hoe de beoordeling van hun werk tot stand kwam.	.79	1.1	4.9	3.2	.77
Nabespreking van opdracht en beoordeling	*7*	De leraar geeft aan waar in het vervolg nog aan gewerkt kan worden.	.73	1.1	4.9	3.5	.59

* Rechtop = huidig handelen. Cursief = gewenst handelen.

** Betreffende het doel, het belang, de bedoeling, de aanpak.

Schalen in de vragenlijsten na elke taak (n=32)

De items over de op de taken verkregen beoordeling (op één extreem geformuleerd item na) bleken te kunnen worden samengevat in drie schalen. De items over de ontvangen feedback bleken te kunnen worden samengevat in twee schalen, die samen 14 van de 20 items bestreken (één item ging over samenwerking bij een opdracht, maar dit was vaak

niet van toepassing; vier items gingen over feedback die negatief overkwam of betrekking had op de persoon of houding van de leerlingen; één item was bij nader inzien weinig informatief geformuleerd). De 16 items in de afsluitende vragenlijst over condities en aspecten bij maken van taken bij het betreffende vak, bleken, afgezien van één item (dat weinig informatief was geformuleerd) te kunnen worden samengevat in drie interpreteerbare (zij het deels zwakke) schalen. De uitkomsten van deze schaalanalyses staan samengevat in tabel 6.3.
De 11 uitspraken in de afsluitende vragenlijst over bij welke typen taken ze zichzelf beter kunnen voorbereiden en er meer van leren, bleken geen schaal te vormen.

Tabel 6.3: Psychometrische informatie over de schalen in de vragenlijst na elke taak (N=32)

Schaal	*k*	*Voorbeelditem*	*r_{ir}'s*	*α*	*min*	*max*	*gem*	*sd*
Over de beoordeling (na elke taak) Met schaal 1-7 (helemaal niet van toepassing – helemaal wel van toepassing)								
Leerling begrijpt beoordeling niet	3	De beoordeling staat er wel, maar dit kan niet kloppen.	.58-.67	.77	1.0	6.3	2.0	1.29
Leerling is tevreden met beoordeling	3	Dat is mooi, het is me gelukt. Ik ben klaar en tevreden	.53-.62	.75	1.0	7.0	4.3	1.56
Leerling wil beter resultaat	3	Ik ga de volgende keer harder werken, zodat ik dan een beter resultaat haal.	.51-.66	.72	1.0	6.7	3.0	1.43
Over de feedback (na elke taak) "Kloppen deze uitspraken?" Met schaal 1-5 (helemaal niet – helemaal wel)								
Leerling vond feedback positief	5	De feedback was duidelijk. De feedback kwam positief over.	.31-.71	.74	2.2	4.8	3.8	.68
Leerling vond feedback leerzaam	9	Door de feedback wist ik wat ik moest doen om verder te kunnen.	.31-.69	.82	1.0	4.7	3.5	.67
Over leren van een taak (als afsluiting van het onderzoek) "Ik leer ervan als" Met schaal 1-4 (weinig – veel)								
Leert als de taak duidelijk is	5	Als de leraar vooraf zegt wat de bedoeling is bij de taak.	.30-.66	.67	1.8	3.8	2.9	.53
Leert als de taak uitdaagt	4	Als ik de taak vooraf interessant en uitdagend vind.	.50-.60	.74	1.8	4.0	3.1	.62
Leert als je erbij moet nadenken	6	Als ik tijdens het maken van de taak moet nadenken over de aanpak.	.25-.60	.65	1.5	3.8	2.7	.51

6.4 Resultaten

Groepsinterviews met drie docenten per school

De groepsinterviews zijn per school beschreven (1,5 pagina per interview) in steeds dezelfde rubrieken: organisatie, beoordeling, toetsen en opdrachten, onderlinge afstemming. Hieruit bleken verschillen tussen de scholen, de vakken en de docenten. Deze worden hier samengevat.
De scholen verschillen in de duur van de blokken waarin wordt gewerkt (6 weken op de scholen A en C en wisselend op school B), en in de duur van de periode waarover of

waarna beoordelingen plaatsvinden (op school A in principe eens per 6 weken, maar afhankelijk van de leerling en de docent ook korter (met mogelijkheid tot verbeteren) of langer; op school B 3, 4 of 6 weken; op school C steeds per vak na elk hoofdstuk of onderdeel, 4 weken bij wiskunde, 6 weken bij Engels, 8 weken bij aardrijkskunde). Ook zijn er verschillen in de manier waarop taken, beoordeling en feedback worden gegeven (op school A wisselend via de elektronische leeromgeving (elo), op papier, mondeling, op school B via elo of mondeling, op school C via papier of mondeling), en in dat waarin de beoordeling wordt uitgedrukt (op de scholen A en B in woorden, op school A in vijf categorieën en op school B in vier, met daarnaast drie extra, en op school C in cijfers).
Ook zijn er verschillen tussen de vakken. Bij Engels en aardrijkskunde komen meer verschillende toetsvormen voor en, het meest bij aardrijkskunde, open en complexe opdrachten die lastig zijn om te beoordelen. Bij wiskunde is er meer eenvormigheid in toetsvorm en meer eenduidigheid bij het beoordelen, en op twee scholen laten de docenten wiskunde de leerlingen zelf gemaakte taken nakijken.
Veel toetsen worden door de docenten zelf gemaakt (dit geldt op één school voor alle toetsen bij aardrijkskunde en op alle drie scholen voor alle toetsen bij wiskunde). De samenwerking met collega's verschilt van sectie tot sectie (op de ene school meer bij Engels, op de andere meer bij aardrijkskunde, etc.). In het algemeen wordt tussen docenten niet veel gesproken over toetsing en beoordeling.
Beoordelen komt in de interviews naar voren als een sterk individuele docentactiviteit, waarbij naast het vakgebied (de kennis en vaardigheden die de leerlingen moeten leren) ook de relatie met de leerlingen een belangrijke rol speelt. Voor een aantal docenten staat beoordelen mede in het teken van het kunnen volgen van de ontwikkeling van de leerlingen en het kunnen inschatten van hun niveau (havo of vwo). De meeste docenten zouden graag willen dat leerlingen van beoordelingen leren, en ze proberen toetsen en opdrachten na te bespreken of er op andere manieren feedback op te geven.
Ter illustratie van de verschillen tussen de docenten volgt nu een meer uitgebreide beschrijving van de rubriek beoordeling.
Op school A. In de onderbouw wordt op deze school beoordeeld met onvoldoende, matig, voldoende, goed en uitmuntend. Of en hoe uitgebreid de docenten daarbij leerlingen nadere feedback geven (verbeterpunten), hangt af van de docent, de leerling en de beschikbare tijd. Bij de docent aardrijkskunde hangt het ook af van het soort opdracht, en krijgen leerlingen als ze de opdracht tijdens de periode hebben ingeleverd de kans om nog dingen te verbeteren. De docent wiskunde bespreekt elke toets na afloop gedetailleerd met de klas na. Voor hem is toetsing een heel belangrijk onderdeel van het onderwijsleerproces: als docent kun je daardoor volgen wat leerlingen wel en niet begrijpen, en leerlingen kunnen ervan leren door hen te laten zien waar ze het goed doen en waar het fout gaat. Op de school leveren leerlingen hun toetsen en opdrachten deels in via de elektronische leeromgeving en deels op papier, en de beoordeling en eventuele feedback door de docenten gebeurt eveneens wisselend: via de elo, op papier en/of mondeling (individueel en/of klassikaal). Dit verschilt tussen docenten en tussen opdrachten. De docent wiskunde laat de proefwerken op papier maken (net als bij een

eindexamen) en schrijft er tijdens het nakijken ook opmerkingen bij (wat soms kan leiden tot nawerk).

Op school B. Binnen een thema is de docent in principe geheel vrij hoe hij beoordeelt, bijvoorbeeld met alleen een eindtoets of ook het nodige maakwerk tussendoor. Beoordeeld wordt met: nog niet voldoende, voldoende, goed of uitstekend. Een nog niet voldoende opdracht moet worden verbeterd, en wordt dan na afronding genoteerd als 'voldoende na verbetering'. Na meerdere herkansingen wordt overgegaan op de notering 'behaald', want dan moet de opdracht worden aangepast, omdat heel veel hulp (moeten) geven niet meer voldoet aan wat in eerste instantie van de leerling werd gevraagd. Ook zijn er procesopdrachten of tussenproducten die alleen moeten zijn ingeleverd, en dat wordt genoteerd met 'gezien'. Het geven van beoordelingen gaat in principe altijd digitaal (via de elo). Daarbij is ook de mogelijkheid om er als docent commentaar bij te geven. De docent Engels doet dit heel veel en richt zich bij beoordelen mede op inzet en groei. Ze geeft aan dat je eigenlijk voortdurend bezig bent om de ontwikkeling van leerlingen te volgen en te beoordelen, en ook kijkt ze of sprake is van een havo of vwo leerling. De docent aardrijkskunde geeft wisselend commentaar via de elo of bij het teruggeven van een product individueel mondeling of bij het nabespreken van een toets met alle leerlingen mondeling, zodat de leerlingen kunnen zien welke soorten fouten ze maken en ook vragen kunnen stellen. Vooral meer open en complexe opdrachten kunnen voor leerlingen lastig zijn. Ook al zijn vooraf criteria bekend, de leerling moet zelf een aanpak kiezen. Ook zij houdt de ontwikkeling van de leerlingen in de gaten, waar ze staan en welke stappen ze nog moeten zetten, en dat betreft ook de inschatting havo of vwo. De docent wiskunde laat de leerlingen de door hem gemaakte toetsen zelf nakijken met behulp van een door hem gemaakt uitgebreid correctievoorschrift en hij verwacht van hen digitaal daarvan een verslagje met daarin de gemaakte fouten en of dit kennisfouten betrof of vergissingen.

Op school C. De docenten op deze school beoordelen met een cijfer. De docent Engels bespreekt de methodetoetsen en de schrijfvaardigheidtoetsen altijd na en komt dan nog een keer terug op dat waarop laag is gescoord of geeft advies over dat waaraan nog moet worden gewerkt. Het komt ook voor, bij opdrachten zonder duidelijke criteria vooraf, dat leerlingen hun werk komen inkijken en uit zichzelf vragen wat ze er nog aan kunnen doen om het te verbeteren. Bij werkstukken zoals een boekverslag zijn er duidelijke criteria en gebruikt ze een rubric. Bij diagnostische toetsen laat ze de leerlingen zelf nakijken. De docent aardrijkskunde bespreekt toetsen die heel goed zijn gemaakt alleen op aanvraag en een toets die heel slecht is gemaakt klassikaal, omdat hij de leerlingen slechts één uur per week ziet. Praktische opdrachten bespreekt hij met leerlingen individueel, buiten de lessen om. De leerlingen krijgen dan tips om het de volgende keer anders aan te pakken en daardoor hopelijk een hoger cijfer te halen. Verbeteren van een beoordeelde opdracht zelf gebeurt alleen bij de praktische opdrachten. Het beoordelen van open opdrachten, waarbij leerlingen binnen een algemeen kader veel eigen keuzes kunnen maken, vindt hij, net als zijn collega's, nog heel lastig. De docent wiskunde vindt dat klassikale nabespreking veel tijd kost en dan ook nog eens niet alle leerlingen zal

interesseren. Daarom bespreekt hij een toets bijvoorbeeld een halve les na met een kleine groep, maar hij heeft voor zichzelf nog niet de optimale manier gevonden om een toets na te bespreken. De minder gemotiveerde leerlingen richten zich na een beoordeling van een toets met bijvoorbeeld een 7 direct op het volgende hoofdstuk. Leerlingen die echt willen weten wat ze niet goed hebben gedaan, kunnen naar de steunles komen, die gebruikt hij ook wel daarvoor, maar dat betreft dan enkelingen. Aan het eind van het jaar is er ineens wel veel meer interesse om een toets in te zien, maar dan om er nog een puntje uit te onderhandelen en zo het cijfer op te hogen, niet om er iets van te leren. Hij zou wel graag willen weten hoe hij het aantrekkelijker kan maken voor leerlingen om een terugkoppeling te krijgen op wat ze goed hebben gedaan en wat niet, zodat ze op volgende toetsen en in volgende jaren niet steeds dezelfde fouten blijven maken.

Vragenlijst voor de docenten over hun beoordelingspraktijk

Opvattingen, intenties en aanpakken bij beoordelen (antwoorden op open vragen)
In de figuren 2 A–C zijn de antwoorden op de open vragen per school samengevat.

Docent	*1*	*2*	*3*
Vak	Engels	Aardrijkskunde	Wiskunde
Aspect ↓			
1 De functie van het geven van een beoordeling is	Leerlingen inzicht geven in hun leerpunten Leerlingen stimuleren	Peilen waar leerling staat, qua niveau en leerpunten	Bepalen of leerdoelen zijn behaald, niveau groep zien
2 Wat ik daarmee wil bereiken bij de leerling is	Fouten herkennen Patronen zien Gestimuleerd zijn	Iets leren, weten waar te verbeteren gestimuleerd zijn	Dat de leerling aantoont de doelen te hebben bereikt
3 Bij geven beoordeling houd ik rekening met	Mijn voorkennis van niveau en de leerpunten van de leerling	Duidelijke argumenten, opbouwende feedback	Fouten, compleetheid, netheid, inzicht *
8 In relatie tot de lessen is het belangrijk dat	De beoordeling verwijst naar eisen en valkuilen	Beoordeling aansluit bij inhoud en criteria in les	Beoordeling overeenkomt met essentie van de les
10 Op deze school is bij beoordeling belangrijk dat	Leerlingen weten hoe ze hun werk kunnen verbeteren	Ook letten op aanpak, vorm en individu	Leerling moet essentie beheersen en netjes werken
4 Bij geven van toets of opdracht bespreek ik	Criteria (eisen) Valkuilen	Bedoeling, aanpak, eventuele extra uitleg	Niets (alles vooraf doorgesproken)
5 Voor de timing vind ik belangrijk dat	Alle leerlingen de beoordeling tegelijk krijgen	Binnen 2 weken Toetsen klassikaal nabespreken	Ik dit (nakijken) in alle rust moet kunnen doen
7 Een beoordeling is goed als (criteria)	Duidelijk en beknopt Stimulerend	Met duidelijke argumenten en opbouwend	Perfecte toets en perfect correctievoorschrift. Met ijver een 6
6 Als ik leerling beoordeling geef, is belangrijk dat	er bij volgende opdracht op wordt teruggekomen	hij deze begrijpt en gemotiveerd is voor volgende	correctievoorschrift gezien, bij fouten feedback gehad
9 Een beoordeling draagt bij aan leren als	deze duidelijk is, op maat, en wordt vervolgd	deze duidelijk is en opbouwend	genoeg feedback is gegeven en oordeel wordt als eerlijk ervaren

** vraag 3 is door docent 3 opgevat in termen van de beoordelingscriteria*

Figuur 6.2 A: Opvattingen, intenties en aanpakken bij beoordelen, per aspect per docent (antwoorden op open vragen in de vorm van aanvulzinnen), school A

Docent	*4*	*5*	*6*
Vak	Engels	Aardrijkskunde	Wiskunde
Aspect↓			
1 De functie van het geven van een beoordeling is	Zichtbaar maken niveau, verder helpen leerproces	Niveau bepalen Waarderen inzet Prestatie duidelijk	Beeld geven aan de leerlingen van capaciteit en inzet
2 Wat ik daarmee wil bereiken bij de leerling is	Inzicht in niveau en aanpak en hoe te verbeteren	Ervan leren. Tevreden zijn. Kunnen afronden	Goed beeld krijgt, goede conclusies trekt
3 Bij geven beoordeling houd ik rekening met	[resultaat, aanpak, inzet, groei]	Leerlingen kunnen toets inkijken, zo nodig feedback	Hun antwoorden en het niveau (havo/vwo)
8 In relatie tot de lessen is het belangrijk dat	Leerling inzicht krijgt in wat te verbeteren	Leerlingen hebben kennis en vaardigheden kunnen oefenen	de beoordeling aansluit bij de mogelijkheden van de leerlingen
10 Op deze school is bij beoordeling belangrijk dat	Aandacht voor resultaat en werkaanpak	Feedback geven!	Letten op niveau en profiel (t.b.v. determinatie)
4 Bij geven van toets of opdracht bespreek ik	Criteria en eisen	Inhoud, vragen, criteria, regels, planning nakijken	Ik laat oefentoets maken en geef regels voor gebruik hulpmiddelen
5 Voor de timing vind ik belangrijk dat	Binnen 2 weken Feedback kort voor herkansing	Alle leerlingen de beoordeling tegelijk krijgen	Zodra ik klaar ben met nakijken geef ik de beoordeling
7 Een beoordeling is goed als (criteria)	Criteria duidelijk. Tips+tops, beoordeling duidelijk	Opdracht helder Criteria en normen helder	Goede inzet leidt tot een 6, inzicht tot hogere cijfers
6 Als ik leerling beoordeling geef, is belangrijk dat	hij er iets mee kan en dit ook doet (groei)	toets inzien, feedback, vragen kunnen stellen	leerling naar mij toekomt als hij anders verwachtte
9 Een beoordeling draagt bij aan leren als	er feedback bij staat en leerling gemotiveerd raakt	leerling inzicht krijgt in wat wel en niet goed ging	leerling daardoor doel-gericht aan de slag gaat

Figuur 6.2 B: Opvattingen, intenties en aanpakken bij beoordelen, per aspect per docent (antwoorden op open vragen in de vorm van aanvulzinnen), school B

Docent	*7*	*8*	*9*
Vak	Engels	Aardrijkskunde	Wiskunde
Aspect↓			
1 De functie van het geven van een beoordeling is	Leerling niveau toont en via fouten stof weer bekijkt	Indicatie geven van kennis leerling en kwaliteit product	Tonen wat wordt beheerst en laten leren van fouten
2 Wat ik daarmee wil bereiken bij de leerling is	Naar de fouten kijkt en zo de stof nogmaals bekijkt	Voldoening als goed, zien wat nodig als slecht	Inzicht in wat goed was en wat minder goed
3 Bij geven beoordeling houd ik rekening met	Inhoud, spelling resp. uitspraak, grammatica *	Bij toets goed/ fout, bij opdracht inhoud/vorm *	Hoe de stof is behandeld en hoe moeilijk die is
8 In relatie tot de lessen is het belangrijk dat	beoordeling past bij wat in de les is behandeld	beoordeling moet motivatie verhogen	aansluiting op de kennis en vaardigheden in de les
10 Op deze school is bij beoordeling belangrijk dat	Rekening houden met dyslexie	Zie antwoorden bij de andere vragen	n.v.t.
4 Bij geven van toets of opdracht bespreek ik	Wat ze hiervoor moeten kennen en weten	Bij toets wat wel/ niet mag, bij opdracht bedoeling	De belangrijkste dingen die ze moeten weten
5 Voor de timing vind ik belangrijk dat	Op die dag geen andere toets is	leerling voldoende tijd heeft gehad voor voorbereiding	leerlingen snel inzicht krijgen in hun capaciteiten
7 Een beoordeling is goed als (criteria)	Als naast cijfer ook uitleg over goed en fout	Toets moet representatief zijn!	Zo objectief mogelijk, en helder/duidelijk
6 Als ik leerling beoordeling geef, is belangrijk dat	leerling weer naar werk kijkt en weet wat nog te oefenen	voldoening als goed, zien wat nodig als slecht	leerling begrijpt wat goed en fout was, +zich erin herkent
9 Een beoordeling draagt bij aan leren als	de leerling nogmaals naar zijn fouten kijkt	als leerling feedback krijgt op goede en foute antwoorden	als leerling inzicht krijgt in wat goed/fout was

** vraag 3 is door de docenten 7 en 8 opgevat in termen van de beoordelingscriteria*

Figuur 6.2 C: Opvattingen, intenties en aanpakken bij beoordelen, per aspect per docent (antwoorden op open vragen in de vorm van aanvulzinnen), school C

In figuur 6.2 (A-C) is te zien dat de meeste docenten bij het geven van beoordelingen belangrijk vinden dat de beoordeling a) aan de leerlingen duidelijkheid verschaft over hun prestatie, in termen van inhoud (de leerstof, kennis en vaardigheden, aanpak, goed/fout, criteria), b) voor de leerlingen opbouwend, motiverend en stimulerend is, door hen wordt herkend en als eerlijk wordt ervaren en hen voldoening geeft, en c) bijdraagt aan het verdere leerproces. Verder leggen de meeste docenten een relatie met het begrip feedback.
Bij de eerste vraag is te zien dat een deel van de docenten bij de functie van beoordelen primair denkt aan de functie voor de leerling, terwijl anderen de functie voor de docent zelf noemen (zoals: zicht krijgen op de mate waarin de doelen zijn behaald, een beeld krijgen van het prestatieniveau van de klas).
Een andere vraag die door de docenten verschillend is geïnterpreteerd, betreft de derde vraag over waarmee ze bij beoordelen rekening houden. Sommige docenten noemen hierbij de grondslag waarop de taak of de beoordeling is gebaseerd (zoals het niveau van de leerlingen), anderen noemen de beoordelingscriteria die zij hanteren, en nog weer anderen noemen aspecten van de manier waarop de beoordelingen aan de leerlingen worden gecommuniceerd.
Elke docent legt wel eigen accenten. Zo stelt bijvoorbeeld docent 3 het uitgangspunt dat de leerling moet aantonen de doelen te hebben bereikt en het belang van een goed correctievoorschrift centraal, terwijl docent 7 herhaaldelijk benoemt dat het zich verdiepen van leerlingen in gemaakte fouten betekent dat zij zich nog weer eens in de stof verdiepen.
Tot slot zijn in figuur 6.2 enkele accentverschillen tussen de drie scholen te zien. Op school A komt het belang van het stimulerend en motiverend zijn van beoordelingen voor leerlingen relatief duidelijk naar voren, op school B het onderdeel uitmaken van beoordelingen van het onderwijsleerproces, en op school C (waarop wordt gewerkt met vakken, lesuren en cijfers) de aandacht (van docenten en leerlingen) voor wat goed is en wat fout. Op de scholen A en B (waar wordt gewerkt in leergebieden, dagdelen en grotere groepen en in het 3e leerjaar de verdeling tussen havo en vwo nog moet plaatsvinden) wordt het zicht op het 'niveau' van de leerling relatief vaak genoemd.

Het eigen handelen bij beoordelen

Op de meeste in de vragenlijst opgenomen 22 uitspraken over het handelen van docenten bij beoordelen scoren de meeste docenten op de mate waarin zij dit nu doen op een schaal van 1 "helemaal niet" tot 5 "helemaal wel" een 3 of hoger. Het hoogste gemiddelde heeft de uitspraak "De leraar beoordeelt werk van leerlingen aan de hand van duidelijke criteria" (uitspraak 8, gemiddelde 4,7), het laagste "De leraar laat de leerlingen meebeslissen over waarop zal worden beoordeeld" (uitspraak 5, gemiddelde 1,7).
Zeven uitspraken scoren gemiddeld lager dan 3. Dit zijn, naast uitspraak 5, de volgende: De leraar laat leerlingen eerst hun werk zelf beoordelen in relatie tot de criteria (4), De leraar betrekt de leerlingen bij het beoordelen van gemaakt werk (6), De leraar betrekt de leerlingen bij het beoordelen van elkaars werk (7), Als de leraar werk beoordeelt,

vergelijkt hij de leerlingen met elkaar (11), De leraar bespreekt de beoordeling klassikaal (20) en De leraar bespreekt de beoordeling mondeling met leerlingen (21).
Bij tien uitspraken verschillen de scores van de docenten meer dan twee schaalpunten. Zie tabel 6.4.

Tabel 6.4 Aspecten van handelen bij beoordelen waarop de docenten sterk verschillen (scores per item)

School	*A*			*B*			*C*		
Vak	E	A	W	E	A	W	E	A	W
Docent	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
6 De leraar betrekt de leerlingen bij beoordelen van gemaakt werk	3	2	5	3	2	4	3	1	1
9 De leraar beoordeelt ook de manier waarop de leerling heeft gewerkt	4	3	5	5	3	1	2	4	5
10 Als de leraar werk beoordeelt, let hij ook op of het al beter is dan eerder	5	3	-	5	4	2	2	4	4
11 Als de leraar werk beoordeelt, vergelijkt hij leerlingen met elkaar	1	3	5	4	4	4	2	1	1
12 Bij groepswerk beoordeelt de leraar ook de manier waarop is samengewerkt	4	3	5	4	4	4	4	3	1
15 De leraar laat extra oefenen voordat de leerlingen de opdracht gaan maken	4	2	4	4	4	5	5	3	5
17 De leraar geeft na de beoordeling aan wat er nog niet goed ging	5	4	5	5	4	2	4	3	2
18 De leraar geeft aan waar in het vervolg nog aan kan worden gewerkt	5	4	4	5	3	3	4	2	2
20 De leraar bespreekt de beoordeling klassikaal	2	3	5	2	3	1	3	4	2
22 De leraar legt leerlingen uit hoe de beoordeling van hun werk tot stand kwam	3	4	5	4	4	5	2	2	3

In tabel 6.4 is te zien dat deze variatie geen duidelijk verband houdt met het vak, maar wel enigszins met de school. De uitspraken 9-12 gaan over het betrekken bij het beoordelen van andere criteria dan vakinhoudelijke (namelijk aanpak/proces, progressie, vergelijking met andere leerlingen, samenwerking) en de uitspraken 17-22 gaan over de feedback en de interactie die aan beoordelen kunnen zijn gekoppeld. Door beoordelingen van vakinhoudelijke prestaties alleen te baseren op inhoudelijke criteria kunnen leerlingen in principe een beter zicht krijgen op hun inhoudelijke prestatie, wat kan bijdragen aan het trekken van lering uit de beoordeling. Het aan beoordelingen verbinden van feedback en interactie kan eveneens bijdragen aan het trekken van lering uit de beoordeling. In beide gevallen scoren de docenten op school C gemiddeld lager dan die op de scholen A en B: zij betrekken minder niet-inhoudelijke criteria bij de beoordeling, maar ook minder feedback en interactie.

Door docenten waargenomen reacties bij leerlingen

Van in de vragenlijst opgenomen 10 mogelijke reacties van leerlingen op een beoordeling nemen de docenten het vaakst de reactie "Tevredenheid" waar ("Dat is mooi", "Het is me gelukt", "Nu ben ik klaar"). Op de schaal van 1 'zelden of nooit' tot 5 'meestal of altijd' scoorde deze reactie gemiddeld 3,7. Nog twee reacties werden gemiddeld ten minste regelmatig waargenomen, namelijk dat de leerling de inspanning verhoogt en dat de leerling de aanpak aanpast (beide gemiddeld 2,7). De overige reacties (de feedback verwerpen, de lat lager of juist hoger leggen, en verschillende manieren om de interactie met de docent aan te gaan) worden slechts sporadisch waargenomen (gemiddelde rond 1,5), met uitzondering van de reactie "Demoralisatie" ("Ik geef de moed op"). Deze reactie heeft een gemiddelde van 2,3 (tussen 'af en toe' en 'regelmatig') en de scores daarop variëren nogal. De drie wiskundedocenten (de docenten 3, 6 en 9) scoren hierbij het hoogst (met 3 of 4).

Vragenlijst vooraf voor de leerlingen

De motivatie voor de taken bij dit vak bij deze leraar

De leerlingen hebben gemiddeld vrij veel vertrouwen in eigen kunnen (*self-efficacy*), waar het gaat om het maken van een taak voor het betreffende vak bij de betreffende leraar. Het gemiddelde op de schaal van 1-7 is 5,1 (sd 1,1). Dit verschilt niet significant tussen de scholen, de vakken en de docenten.

Van de vier gradaties op de dimensie intrinsieke-extrinsieke motivatie scoren de leerlingen gemiddeld het hoogst op geïdentificeerde regulatie (geleerd om in te zien dat het van belang is) (5,2, sd 1,1) en externe regulatie (4,9, sd 1,1). Het gemiddelde op intrinsieke motivatie is wat lager (3,9, sd 1,4), en op amotivatie nog lager (3,0, sd 1,2). De intrinsieke motivatie verschilt significant tussen de drie scholen (3,9 op school A, 4,3 op school B en 3,5 op school C), maar niet tussen de drie vakken (gem. 4,0 sd 1,4 bij Engels, gem. 4,0 sd 1,3 bij aardrijkskunde, gem. 3,7 sd 1,6 bij Wiskunde). School B scoort gemiddeld dus hoger op intrinsieke motivatie dan de scholen A en C. Dit zien we terug per docent in tabel 6.5.

Tabel 6.5: Intrinsieke motivatie voor taken bij deze leraar, gemiddelde (sd)

Schaal	*School*	*A*			*B*			*C*	
	Vak	E	A	E	A	W	E	A	W
	Docent	*1*	*2*	*4*	*5*	*6*	*7*	*8*	*9*
Intrinsieke	3.9	4.2	3.8	4.2	4.3	4.4	3.6	4.0	3.0
motivatie	(1.4)	(1.3)	(1.1)	(1.4)	(1.3)	(1.6)	(1.3)	(1.3)	(1.2)

De docent als feedbackgever en als beoordelaar, volgens de leerlingen

De docenten verschillen op een aantal schalen significant. Zie de tabellen 6.6 A en 6.6 B, waarin deze schalen met # zijn gemarkeerd.

Tabel 6.6 A: Gemiddelde (sd) voor huidig (rechtop) en door leerlingen gewenst (cursief) docenthandelen m.b.t. <u>feedback geven</u> schaal per docent (schaal 1-5)

		Vak	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
		Docent	1	2	4	5	6	7	8	9
Schaal *		*N*	*17*	*21*	*23*	*30*	*23*	*24*	*22*	*24*
		Totaal								
Bespreking vooraf van de opdracht **	3.4	#	3.3 (.75)	2.9 (.56)	3.1 (.69)	3.2 (.70)	3.8 (.50)	3.3 (.70)	3.5 (.57)	3.7 (.51)
Bespreking vooraf van de opdracht	4.0		3.9 (.58)	3.9 (.54)	3.8 (.60)	4.1 (.70)	4.1 (.48)	3.9 (.40)	4.1 (.58)	3.9 (.58)
Procesgerichte feedback	3.4	#	3.5 (.51)	3.0 (.74)	3.5 (.46)	3.2 (.63)	3.8 (.56)	3.3 (.67)	3.1 (.76)	3.6 (.54)
Procesgerichte feedback	3.8		3.9 (.42)	3.7 (.60)	3.7 (.58)	4.1 (.63)	4.0 (.48)	3.6 (.71)	3.8 (.56)	3.9 (.46)
Resultaatgerichte feedback	3.8	#	4.1 (.53)	3.9 (.49)	3.8 (.47)	3.7 (.69)	4.0 (.61)	3.6 (.70)	3.5 (.71)	3.7 (.60)
Resultaatgerichte feedback	4.1	#	4.3 (.43)	4.3 (.37)	4.0 (.54)	4.4 (.69)	4.3 (.48)	3.9 (.68)	4.0 (.61)	4.0 (.65)

* Rechtop = huidig handelen. Cursief = gewenst handelen.

** Betreffende het doel, het belang, de bedoeling, de aanpak.

De verschillen in gemiddelde tussen de docenten zijn significant ($p < .05$).

Tabel 6.6 B Gemiddelde (sd) voor huidig (rechtop) en door leerlingen gewenst (cursief) docenthandelen m.b.t. <u>beoordelen</u> per schaal per docent (schaal 1-5)

		Vak	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
		Docent	1	2	4	5	6	7	8	9
		N	*17*	*21*	*23*	*30*	*23*	*24*	*22*	*24*
Schaal *		*Totaal*								
Bespreking vooraf van criteria en standaarden **	3.3	#	3.1 (.58)	2.7 (.61)	3.3 (.66)	3.4 (.70)	3.5 (.90)	3.3 (.74)	3.3 (.57)	3.4 (.46)
Bespreking vooraf van criteria en standaarden	3.8	#	3.4 (.54)	3.4 (.51)	3.6 (.69)	4.1 (.69)	4.1 (.59)	3.7 (.55)	3.9 (.53)	3.8 (.47)
Betrekken van leerlingen bij beoordeling	2.2	#	2.2 (.83)	1.8 (.77)	2.8 (.77)	2.5 (.88)	2.2(1 .09)	2.1 (.87)	1.7 (.60)	2.4(1 .20)
Betrekken van leerlingen bij beoordeling	2.9	#	2.3 (.73)	2.3 (.98)	3.2 (.80)	3.5 (.88)	2.5(1 .14)	2.8 (.90)	2.5 (.81)	3.4(1 .07)
Ook beoordelen van aanpak, proces, progressie	3.4		3.7 (.67)	3.3 (.57)	3.6 (.78)	3.4 (.85)	3.5 (.73)	3.4 (.86)	3.4 (.66)	3.1 (.78)
Ook beoordelen van aanpak, proces, progressie	3.7		4.0 (.73)	3.5 (.76)	3.6 (.74)	4.0 (.85)	3.8 (.91)	3.6 (.84)	3.7 (.89)	3.6 (.75)
Nabespreking van opdracht en beoordeling	3.2	#	3.3 (.50)	2.5 (.60)	3.3 (.80)	2.9 (.88)	3.7 (.63)	3.4 (.75)	3.0 (.71)	3.3 (.58)
Nabespreking van opdracht en beoordeling	3.8	#	3.8 (.54)	3.0 (.67)	3.8 (.72)	3.8 (.88)	4.0 (.59)	3.9 (.43)	3.7 (.60)	4.0 (.44)

* Rechtop = huidig handelen. Cursief = gewenst handelen.

** Betreffende het doel, het belang, de bedoeling, de aanpak.

= De verschillen in gemiddelde tussen de docenten zijn significant ($p < .05$).

Samenvattende schalen voor 'huidig handelen' en 'gewenst handelen'
De gemiddelde schaalscores op de schaal van 1-5 voor huidig docenthandelen verschillen significant tussen vakken en tussen docenten, en die voor het gewenste docenthandelen verschillen significant tussen scholen en tussen docenten. Zie de tabellen 6.7 en 6.8.

Tabel 6.7: Schaalgemiddelde (sd) voor huidig en gewenst docenthandelen per school en vak

			School				Vak		
Schaal	*Totaal*		*A*	*B*	*C*		*E*	*A*	*W*
Huidig handelen	3.2		3.1	3.3	3.2	#	3.3	3.1	3.4
	(.53)		(.50)	(.54)	(.53)		(.53)	(.51)	(.50)
Gewenst handelen	3.7	#	3.5	3.9	3.7		3.7	3.7	3.8
	(.47)		(.45)	(.51)	(.47)		(.43)	(.55)	(.39)

De verschillen in gemiddelde tussen de scholen c.q. vakken zijn significant (p <.05).

Tabel 6.8: Schaalgemiddelde (sd) voor huidig en gewenst docenthandelen per docent

		Docent	1	2	4	5	6	7	8	9
Schaal	*Totaal*	*Vak*	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
Huidig	3.2	#	3.3	2.9	3.3	3.2	3.5	3.2	3.1	3.3
handelen	(.53)		(.46)	(.44)	(.47)	(.59)	(.50)	(.62)	(.43)	(.49)
Gewenst	3.7	#	3.7	3.4	3.7	4.0	3.8	3.6	3.7	3.8
handelen	(.47)		(.38)	(.49)	(.50)	(.57)	(.37)	(.41)	(.40)	(.41)

De verschillen in gemiddelde tussen de docenten zijn significant (p < .05).

In tabel 6.7 laat het vak wiskunde van de drie vakken het hoogste gemiddelde zien op zowel huidig als gewenst handelen, en heeft school A, waar de beoogde wiskundedocent uiteindelijk niet meedeed, het laagste gemiddelde op zowel huidig als gewenst handelen. Het valt daarom niet uit te sluiten dat indien op school A bij deze wiskundedocent wel leerlingen hadden meegedaan, de scholen niet alleen op het huidig handelen maar ook op het gewenst handelen niet significant zouden hebben verschild. In dat geval resteert in die tabel alleen het significante verschil op het huidig handelen tussen de vakken, en dit verschil is dan wel significant (p=.001) maar in absolute zin niet groot (het gemiddelde varieert tussen 3,1 en 3,4). De verschillen tussen de docenten zijn significant voor zowel huidig als gewenst handelen, zie tabel 6.8. Op het huidig handelen varieert het gemiddelde tussen 2,9 en 3,5 (p=.004) en op het gewenst handelen tussen 3,4 en 4,0 (p=.002). Het verschil tussen 'huidig' en 'gewenst' varieert tussen de docenten van 0,3 tot 0,8 schaalpunt.

De onderzochte toetsen en opdrachten

In totaal konden bij acht docenten 21 toetsen en opdrachten in het onderzoek worden betrokken, zie figuur 6.3. In bijna alle gevallen kon de toets of opdracht zoals die, op papier of via een digitale leeromgeving, aan de leerlingen is verstrekt, worden verkregen, en bij een deel van de toetsen en opdrachten is enige aanvullende informatie verstrekt. We duiden hierna toetsen en opdrachten beide aan met het woord taken.

Uit de op basis van de verkregen informatie gemaakte samengevatte beschrijvingen (0,5 tot 1 A4) van elk van de gemaakte en beoordeelde taken en de reacties van de leerlingen daarop blijkt dat de taken op een aantal aspecten sterk variëren, vooral bij de vakken Engels en Aardrijkskunde. Ook blijkt dat de reacties van de leerlingen heel uiteenlopend kunnen zijn, ook tussen leerlingen onderling bij dezelfde taak. Figuur 6.4 bevat een voorbeeld van een beschrijving van een taak, een opdracht bij Engels op school A.

School	Vak	Docent	Nummer taak	Typering van de taak (toets of opdracht)
A	E	1	1	1 Een open schrijfopdracht (dagboek van een gefingeerde reis), in tweetallen
			2	2 Een open schrijfopdracht (brief aan een tijdschriftredactie), individueel
			3	3 Een complexe serie taken rond schrijven (voor een inzending voor een competitie via internet), in een groep van vier met wisselende rolverdeling, digitaal en klassikaal
	A	2	4	1 Een open schrijfopdracht (brief aan de overheid van een land), individueel
			5	2 Discussiëren als voor- of tegenstander (over een verkeersprobleem), schriftelijke toelichting standpunt, beoordeling bijdrage medeleerling, schriftelijk verslag discussie
			6	3 Vergelijkend onderzoek (van twee havens) en maken van een verslag, in tweetallen
	W	3		-
B	E	4	7	1 Lees- en schrijftaken op de digitale leeromgeving (over manuals), individueel
			8	2 Lezen van teksten, beantwoorden van vragen en maken van samenvatting, individueel
			9	3 Toets (luistervaardigheid)
	A	5	10	1 Toets (reeks diverse vragen over geschiedenis, aardrijkskunde en economie)
			11	2 Een complexe open opdracht (maken van een businessplan en een website en verzorgen van een presentatie), groep van 4-5 leerlingen
	W	6	12	1 Toets na hoofdstuk over kwadraten, tafels en breuken
			13	2 Toets na hoofdstuk over oplossen van vergelijkingen
			14	3 Toets na hoofdstuk over letterrekenen
C	E	7	15	1 Toets (vertalen van woorden en zinnen, vervoegen, tekst lezen, email schrijven)
			16	2 Toets over woordjes, grammatica en vergrotende trap
			17	3 Toets over woordjes en grammatica
	A	8	18	1 Toets over topografie wereld
			19	2 Toets over atlasvaardigheden
	W	9	20	1 Toets na hoofdstuk over lineaire verbanden
			21	2 Toets na hoofdstuk over gelijkvormigheid

Figuur 6.3: Overzicht van de taken die de leerlingen moesten maken

Taak 1. Opdracht, School A, Engels, Docent 1 (Beschikbare beschrijving: 2 pp.)	
Inhoud & opzet	Creative writing task in pairs. Maak in een duo samen een *cool* dagboek van een denkbeeldige reis, *backpacking* door een land naar keuze.
Vorm & omvang	Lengte dagboek tenminste 400 woorden, over tenminste 4 dagen.
Aanwijzingen	Kies een land, motiveer je keus, bedenk wat je wilt zien. Eerst met elkaar bespreken, dan beschrijven. Beschrijf om en om een dag, morgen+middag+avond. Verbind de dagen met elkaar tot een doorlopend verhaal. Schrijf in *full English*, gebruik de verleden tijd.
Beoordeling	In te leveren: de planning, het dagboek, een korte evaluatie. Geen aparte expliciete beschrijving van criteria.
Resultaten interview met 4 leerlingen	
Kenmerken, duidelijkheid, uitleg, keuzes	De docent heeft de opdracht klassikaal uitgelegd en het was wel duidelijk wat we moesten doen. Voor één leerling was niet duidelijk dat het in de verleden tijd moest worden geschreven, zij moest de tekst op de valreep nog aanpassen. De bedoeling van de opdracht (waarom de docent deze gaf) was niet duidelijk, maar de leerlingen vonden het wel nuttig om te oefenen met de verleden tijd. Je mocht zelf kiezen waarover je wilde schrijven.
Eisen, criteria	Je moest tenminste 400 woorden schrijven en tenminste 4 dagen beschrijven. Verder stonden er alleen wat punten die in de tekst moesten staan. Je mocht zelf weten of je plaatjes of foto's wilde toevoegen (dat heeft ook niet iedereen gedaan).
Motivatie, moeilijkheid	De leerlingen vonden het nuttig om te oefenen met de verleden tijd en leuk dat ze hun fantasie konden gebruiken (je mocht het zo gek maken als je wilde). De leerlingen vonden de opdracht wel gemakkelijk (al viel het sommigen tegen hoeveel fouten er nog in bleken te zitten).
Begeleiding, feedback	De leerlingen hebben niet echt begeleiding gekregen (volgens één leerling had de docent zoiets van die snappen het wel). Je kon wel vragen stellen, maar dat hebben ze niet gedaan. Wel moest iedereen op een bepaalde dag de eerste versie inleveren en gaf de docent er mondeling feedback op. Één leerling had veel fouten gemaakt die allemaal waren omcirkeld; daardoor voelde het voorgekauwd, ze had liever gehad dat de docent het had uitgelegd en haar de tekst zelf had laten verbeteren. De andere leerlingen hoefden weinig meer te veranderen. Eén leerling zou, ook al staan er geen of weinig fouten in, toch willen weten wat ze beter kan doen. Omdat het digitaal ging, heeft ze de docent niet om een toelichting gevraagd. Als hij het mondeling had gezegd, had ze dat wel gedaan.
Proces, problemen	De leerlingen zijn geen problemen tegengekomen. Dat komt ook doordat ze in hun tweetallen goed met elkaar konden samenwerken (als het niet klikt, is het lastiger).
Beoordeling	De leerlingen hadden uitmuntend of goed (één leerling wist het niet meer). De docent heeft nog digitale feedback gegeven (één leerling had daar nog niet naar gekeken), maar daar stond in wat hij al had gezegd.
Geleerd	De leerlingen hebben goed geoefend met de verleden tijd en de vormen van woorden. Één leerling heeft geleerd dat woorden meer vormen kunnen hebben ook al klinken ze hetzelfde; ze spreekt nu sommige woorden voor zichzelf uit als ze die schrijft en vraagt zich af of het klopt, en dat werkt goed. Over de manier van werken hebben ze niet geleerd, ze werken vaak samen.

Figuur 6.4: Voorbeeld van een opdracht en de reacties van de leerlingen (vak Engels, school A)

Groepsinterviews met leerlingen: reacties op de gemaakte taken en de beoordeling daarvan

Door het beperkte aantal taken in relatie tot hun diversiteit, kunnen niet gemakkelijk duidelijke patronen worden gevonden van bepaalde kenmerken van taken en docenthandelen die samengaan met bepaalde reacties van leerlingen. Wel bieden de antwoorden in de groepsinterviews een staalkaart van manieren waarop leerlingen tegen opdrachten en toetsen en daarop verkregen beoordelingen kunnen aankijken en hoe zij daarop kunnen reageren. We vatten dit hieronder op hoofdlijnen samen, aangevuld met voorbeelden. De getallen hierna achter de hoofdletter T verwijzen naar de nummers van de taken in figuur 6.3. De voorbeelden achter de opsommingstekens zijn bedoeld als illustratie (aan het eind tussen haakjes aangevuld met het nummer en een korte omschrijving van de betreffende taak).

De duidelijkheid van toetsen en opdrachten voor leerlingen

In hun kenschets van de toets of opdracht die ze moesten maken, noemen de leerlingen nadrukkelijk *de mate van duidelijkheid* (zoals zij die hebben ervaren). De taak was 'wel duidelijk' (bijvoorbeeld T 1, 5, 8), 'eerst een beetje vaag maar gaandeweg wel duidelijk' (bijvoorbeeld T 3), 'vooraf goed uitgelegd' (bijvoorbeeld T 10, 12, 17), 'heel duidelijk', 'precies' (bijvoorbeeld T 14, 16). Dit is voor de leerlingen een belangrijk aspect.
Bij een aantal taken hebben docenten *verschillende manieren om informatie te verstrekken* gebruikt: op internet, op papier, via mondelinge instructie (bijvoorbeeld T 3, 5). Uit de reacties van leerlingen bij zulke taken blijkt dat als informatie die al op internet of op papier beschikbaar is, ook nog eens door de docent wordt benoemd of toegelicht, dit kan bijdragen aan de duidelijkheid. Uit de reacties bij taken waarbij de informatie slechts op één manier beschikbaar was, blijkt echter dat ook in zulke situaties voor leerlingen een taak voldoende duidelijk kan zijn. Een taak wordt niet persé duidelijker doordat op meer plaatsen of meer momenten informatie beschikbaar is, maar wel onduidelijker als informatie vaag is of tegenstrijdig, of ontbreekt.
Vooral de in het onderzoek betrokken *opdrachten verschillen in duidelijkheid* voor de leerlingen (bijvoorbeeld T 1 en 3, beide bij Engels, en T 5 en 6, beide bij aardrijkskunde). Ook kan bij een opdracht wel duidelijk zijn wat leerlingen moeten doen maar kan bij de beoordeling alsnog onduidelijkheid ontstaan over waar het resultaat aan moet voldoen (zie bijvoorbeeld T 1, 3, 5).

- Volgens één leerling moest je in de tekst drie oplossingen aandragen, maar in de feedback stond dat hij ook het openbaar vervoer moest noemen. Volgens een andere leerling had de docent gezegd wat je moest doen en als je dat deed was het goed. Een nog andere leerling had ook feedback gekregen over dingen die ze helemaal niet in de opdracht had zien staan. Bij de discussie vonden alle leerlingen niet duidelijk waar die aan moest voldoen. Je moest je standpunt vertellen, maar je moest maar afwachten tot je een keer aan de beurt kwam, en sommigen kregen steeds de beurt en anderen helemaal niet. Eén leerling had slechts een voldoende, omdat hij steeds hetzelfde had gezegd, maar dat kwam doordat steeds dezelfde

vraag werd gesteld. Verder moest je ook andere leerlingen beoordelen met een formulier, maar een aantal leerlingen had voordat zij aan de beurt kwamen het formulier nog niet kunnen lezen en een aantal anderen wel, waardoor die nog konden letten op de punten die daarop stonden. (T 5, samengestelde individuele opdracht bij aardrijkskunde)

In de set van in dit onderzoek betrokken taken worden toetsen door leerlingen vaker als duidelijk ervaren dan opdrachten. Bij toetsen kan de duidelijkheid behalve door de informatie over de toets ook worden bevorderd door *de voorbereiding op de toets* door de voorafgaande instructie over de stof en/of het oefenen daarmee in lessen of via huiswerk-opdrachten (bijvoorbeeld T 12, 13 en 14 (wiskunde), 17 (Engels), 18 en 19 (aardrijks-kunde)), door het maken van een oefentoets (bijvoorbeeld T 9 (aardrijkskunde), T 13, 20 en 21 (wiskunde), T 15 en 16 (Engels)) of doordat vooraf door de docent is aangegeven wat voor soort vragen de leerlingen konden verwachten (bijvoorbeeld T 9, 10 en 18 (aardrijkskunde), T 14 (wiskunde), T 16 (Engels)).

- Volgens de leerlingen was het een heel brede toets, over aardrijkskunde, geschiedenis en economie, waarvoor ze vooral veel begrippen moesten leren, en het ging ook over globalisering. Ze geven aan vooraf goed te zijn geïnformeerd. De docent heeft vooraf gewoon gezegd wat voor soort vragen ze zouden krijgen, en ook dat de toets moeilijk zou zijn maar dat als je zou leren hij te doen was. Ook kregen ze drie dagen voor de toets een samenvatting. De leerlingen moesten leren uit readers en informatie haalde je verder uit de links op de ELO en uit de opdrachten die je tijdens de les maakte. Ze hebben ook meerdere workshops over het onderwerp gehad, en de docent staat altijd op de groep zodat je vragen kunt stellen als dat nodig is. Ze begrepen de stof voor de aanvang van de toets goed, en alleen de opdrachten maken was al een goede voorbereiding. (T 10, samengestelde toets, bij aardrijkskunde)
- De toets was aangekondigd en was een basistoets, over kwadraten, tafels en breuken, om het geheugen weer even op te frissen na de vakantie. De docent heeft vooraf goed uitgelegd wanneer ze de toets hadden en wat ze moesten leren. Het stond ook op de site. De leerlingen wisten ook wat voor vragen ze zouden krijgen, gewoon dingen zelf uitrekenen (geen meerkeuze), met wiskunde krijgen ze vaker zulke toetsen. Aan het begin van het jaar is de stof nog wel een beetje uitgelegd, vooral de breuken, dat wisten sommigen niet meer. Verder was het gewoon een beetje herhalen, deze basisstof gebruik je eigenlijk steeds en je hoefde niet echt veel te leren. De toets sloot gewoon aan op de opdrachten, de rijtjessommen die ze hadden gehad. Er is altijd ruimte om een vraag te stellen, maar daar hadden ze geen behoefte aan. (T 12, toets bij wiskunde)
- Elke toets wordt twee weken van te voren opgegeven. De les daarna maak je dan een d-toets (diagnostische toets), die lijkt op de toets en met nog een moeilijke toepassingsvraag. Alle toetsen hebben dezelfde opbouw, de vorm is steeds hetzelfde. In het boek staat op de laatste twee bladzijden van het hoofdstuk altijd precies welke woorden en zinnen je moet leren. De docent geeft altijd ver van te voren op wat we moeten leren, alle grammatica wordt een paar keer behandeld, ze schrijft veel op het bord en als je iets niet snapt kun je het vragen. Alle leerlingen geven aan goed geleerd te hebben voor de toets (eentje heeft niet veel moeite met Engels en hij leest het gewoon een beetje door, een ander heeft veel moeite met Engels en leert er altijd goed voor, de grammatica samen met zijn vader). De leerlingen begrepen de stof ook goed tijdens de laatste les voor de toets; de vierde leerling een paar dingen nog niet maar die lees je dan in het boek na en dan lukt dat wel. (T 15, toets bij Engels)

Verschillen tussen leerlingen in motivatie en ervaren moeilijkheid

Bij opdrachten is gevraagd of ze daarvoor *gemotiveerd* waren en of ze de opdracht *makkelijk of moeilijk* vonden en bij toetsen alleen of ze de toets *makkelijk of moeilijk* vonden. De antwoorden verschillen per groepje tussen de leerlingen.

- Eén leerling vond het wel een leuke opdracht, om dingen uit te zoeken en te kijken wat voor oplossingen er zijn. Een andere leerling vond het wel een leuk onderwerp en hij vond ook de discussie leuk. Een nog andere leerling vond het niet echt leuk, vond het onderwerp saai. Het was moeilijk om informatie te vinden, daar moest je veel moeite voor doen en daar raakte ze ongemotiveerd door. Ze weet ook niet wat ze nu heeft aan dat knooppunt en waarom ze moet weten dat daar files staan, het maakt haar niet zoveel uit. De vorm vond ze wel leuk, weer eens iets anders dan normaal. Alle leerlingen vonden een deel van de opdracht niet gemakkelijk: informatie zoeken, oplossingen bedenken, tijdens de discussie er tussen komen, en dergelijke (T 5, samengestelde individuele opdracht bij aardrijkskunde)
- Eén leerling vond de opdracht leuk. Bij een toets wordt je niet geholpen en bij zo'n opdracht leer je heel veel, tijdens het werken eraan. Ook er een brief erbij moeten schrijven vond zij op deze manier veel fijner (meer complex en uitdagend) dan als ze zeggen ga maar leren hoe je een brief moet schrijven. De tweede leerling vond het wel leuk en was gemotiveerd, bij een toets kun je gewoon leren en dit is echt een heel ingewikkelde opdracht (de informatie was wel lastig te vinden, dat kostte veel tijd). De derde leerling vond de opdracht heel irritant; als je een groepje hebt met mensen die allemaal iets anders willen is dat demotiverend, en er werd ook Nederlands aan gekoppeld en het was zoveel en in weinig tijd dus hij vond het niet echt wat. (T 11, complexe groepsopdracht bij aardrijkskunde)

Geven van betekenis aan ontvangen beoordelingen en leren van toetsen en opdrachten

Bij veel opdrachten en toetsen zijn leerlingen achteraf actief bezig met het duiden van de verkregen beoordelingen.

- Eén leerling had een voldoende voor de tekst en voor de discussie; met die eerste voldoende was hij tevreden maar met die voldoende voor de discussie niet. Een andere leerling (…). Een derde leerling had voor alle onderdelen voldoende; in de feedback bij de tekst stond eerst dat deze heel goed was en daarna van en dat en dat kan beter; dat vond ze wel gek en daardoor was haar ook niet duidelijk van wat er nu goed was en wat niet. Volgens de eerste leerling doet de docent dat altijd zo, en is het duidelijker als gewoon eerst wordt genoemd wat er niet goed is; de anderen beamen dit. (T 5, samengestelde individuele opdracht bij aardrijkskunde)
- Twee leerlingen hadden voldoende, eentje had goed en eentje had uitstekend. Ze hebben geen verdere feedback gekregen. De leerlingen die daarna al wel weer Engelse les hebben gehad geven aan dat de toets niet is nabesproken, ze moesten gewoon verder met de opdrachten. De leerlingen geven aan van de toets niet iets te hebben geleerd. (T 9, luistertoets bij Engels)
- De eerste leerling had een 7, ze had het wel hoger verwacht maar dat heeft ze altijd (ook bij andere vakken), en dan maakt ze allemaal hele kleine foutjes en dan gaat er meteen al veel af. De tweede leerling had een 8 en is wel goed in Engels maar dan leert hij te weinig en schrijft hij woordjes verkeerd. De derde leerling had een 3 en had toch wel een voldoende verwacht. Ze snapte het echt niet, ze begon al niet met een goed gevoel aan de toets en dan gaat het helemaal verkeerd. Ze moet nu gewoon bij de volgende toets extra hard haar best doen. (T 17, toets bij Engels)

- De leerlingen hadden een 9 (zij had eerder een 7 verwacht omdat zij het heel veel vond), een 6,5 (hij had veel lager verwacht omdat hij zich vooral op Europa had gericht), een 10 (hij wist voor de toets al bijna alles en is al op veel plaatsen geweest) en een 8,5 (zij had dat wel verwacht, ze had goed geleerd). De leerlingen hebben de toets teruggekregen maar hadden direct na afloop al in het boek gekeken om te zien wat ze fout hadden gedaan. (T 18, kleine toets (topografie) bij aardrijkskunde)

Als leerlingen bezig zijn ontvangen beoordelingen te duiden, is sprake van een leerproces. Aanvullend is in elk interview, in het verlengde daarvan, expliciet gevraagd wat de leerlingen van de toets of opdracht hadden geleerd. Hieronder tot slot daarvan een aantal voorbeelden.

- Eén leerling heeft geleerd dat de haven van Rotterdam heel groot is. Een andere leerling heeft geleerd dat ze eerder had moeten beginnen, want er was in de laatste week te weinig tijd om het af te maken. (T 6, opdracht in tweetallen bij aardrijkskunde)
- Eén leerling geeft aan dat je door tekst te lezen altijd wel wat leert, over de inhoud maar bijvoorbeeld ook woorden. Een andere leerling vind de feedback altijd heel fijn. De derde leerling zou zich de volgende keer meer verdiepen in de opdracht, in wat je precies moet doen, hij heeft volgens hem gewoon ergens overheen gelezen. (T 8, serie korte individuele opdrachten bij Engels)
- Eén leerling heeft door het werken in een groepje geleerd om zich aan te passen. De tweede leerling heeft veel moeten opzoeken en geleerd om door te zetten. De derde leerling heeft in zijn groepje wel veel gediscussieerd en dat vond hij wel leerzaam. (T 11, complexe samenwerkingsopdracht bij aardrijkskunde)
- Door het maken van de toets fris je de stof weer even op, volgens een van de leerlingen. Hij zou de volgende keer wel eerder beginnen met leren. Een andere leerling zou zich beter voorbereiden. De toets is de week erna ook klassikaal nabesproken, dan legt de docent alles nog een keer uit. De leerlingen vinden zo'n nabespreking handig, je ziet als je niet geleerd had hoe het had gemoeten en je kunt vragen stellen. (T 12, toets bij wiskunde)
- Twee leerlingen hebben door het moeten maken van de toets grammatica en woordjes geleerd, een andere leerling heeft geleerd de volgende keer iets eerder te leren, zodat hij meer tijd heeft (T 15, toets bij Engels)
- De leerlingen brachten verschillende ervaringen en opvattingen in met betrekking tot het krijgen van lage cijfers. Zij zien (bij dit vak) geen duidelijk verband tussen hoe hard ze ervoor werken en het cijfer dat ze krijgen. Een leerling: als je in de les altijd goed meedoet en je huiswerk maakt en op de toets heb je een 4 dan staat dat echt niet in verhouding. Een andere leerling: het is altijd een momentopname, als je een dag niet lekker bent dan maak je de toets niet goed en heb je pech, je zou een onvoldoende moeten mogen herkansen. Een weer andere leerling: door die slechte cijfers heb ik ook niet echt het idee van dan ga ik harder werken. (T 20, toets bij wiskunde)

Concluderend lijkt het erop dat, mede door de vooraf verstrekte informatie aan en de voorbereiding door de leerlingen, bij toetsen de inhoud en de beoordelingen gemiddeld duidelijker zijn dan bij opdrachten en dat leerlingen daardoor bij toetsen achteraf actiever bezig zijn om verkregen beoordelingen te duiden en ook beter kunnen aangeven wat ze hebben geleerd. De motivatie voor en de ervaren moeilijkheid van de taken verschilde veelal per individuele leerling, evenals de betekenis die achteraf aan verkregen

beoordelingen werd gegeven, mede afhankelijk van de hoogte van de beoordeling. Zowel de door de leerlingen gepercipieerde duidelijkheid van de taak als de persoonlijke ervaring en betekenisgeving droegen bij aan wat leerlingen ervan leerden.
De reacties van de leerlingen op de verkregen beoordelingen kunnen als volgt worden getypeerd: a) de beoordeling voor kennisgeving aannemen; b) niet goed begrijpen hoe de beoordeling tot stand is gekomen, evt. verwijzend naar onduidelijkheden in de taak of de beoordeling; c) verklaren van een goede beoordeling door goed zijn in het vak of goed hebben geleerd; d) verklaren van een tegenvallende beoordeling door niet goed zijn in het vak en/of er onvoldoende voor te hebben gedaan, of slordigheid tijdens het maken van de taak, of de moeilijkheid van de taak (en/of de samenwerking), of gebruik door de docent van bijkomende beoordelingscriteria; e) zich voornemen om voortaan harder te werken, anders te leren, eerder te beginnen (bij toetsen met leren, bij opdrachten ook *met het maken van* de opdracht) en/of een betere prestatie neer te zetten; f) zich niets voornemen.

Vragenlijst voor leerlingen na elke taak over de verkregen beoordeling en feedback

Het eerste deel van de vragenlijst die na afloop van elke cyclus door de vier leerlingen per docent werd ingevuld, bevatte een serie van 10 items met een schaal van 1-7 over de reacties van de leerlingen op de verkregen beoordeling, en zes losse vragen over hun overige reacties. Dit deel is 74 keer ingevuld.

<u>Reacties op de verkregen beoordeling</u>
Van de 10 genoemde reacties herkenden de leerlingen bij zichzelf het vaakst reactie 1: tevredenheid (gemiddelde 4,7 op de schaal van 1-7), en ook vrij vaak reactie 4 (goed genoeg) (gem. 4,0). De reacties 5 en 6, beide gericht op een beter resultaat, scoorden ook hoog (gem. 4,3 en 4,1). De overige reacties scoorden rond de 2. De antwoorden op de 10 reacties konden worden samengevat in drie voldoende betrouwbare schalen (zie tabel 6.3), over de mate waarin de leerling respectievelijk: de beoordeling niet begrijpt, tevreden is met de beoordeling, en streeft naar een beter resultaat. Zie de tabellen 6.9 en 6.10. In de tabellen is te zien dat bij alle drie typen reacties sprake is van verschillen, ook tussen docenten van hetzelfde vak.

Tabel 6.9: Schaalgemiddelde (sd) van de schalen m.b.t. beoordeling, per school en vak

		School				Vak		
Schaal	*Totaal*	*A*	*B*	*C*		*E*	*A*	*W*
Leerling begrijpt de	2.0 (1.3)	1.6	2.1	2.2		2.0	2.1	2.0
beoordeling niet		(0.8)	(1.4)	(1.4)		(1.4)	(1.3)	(1.1)
Leerling is tevreden met de	4.3 (1.6)	4.3	4.6	3.9		4.5	4.5	3.8
beoordeling		(1.6)	(1.2)	(1.9)		(1.5)	(1.7)	(2.0)
Leerling wil beter resultaat	3.0 (1.4)	2.8	2.8	3.3	#	2.9	2.5	3.8
		(1.2)	(1.4)	(1.7)		(1.2)	(1.1)	(1.8)

De verschillen in gemiddelde tussen de scholen c.q. vakken zijn significant ($p < .05$).

Tabel 6.10: Schaalgemiddelde (sd) van de schalen m.b.t. de beoordeling. per docent

		Docent	1	2	4	5	6	7	8	9
Schaal	*Totaal*	*Vak*	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
Leerling begrijpt de	2.0 (1.3)		1.2	2.0	2.0	2.4	2.0	2.7	1.8	2.0
beoordeling niet			(0.2)	(0.9)	(1.4)	(1.9)	(1.2)	(1.7)	(1.2)	(1.0)
Leerling is tevreden	4.3 (1.6)		4.4	4.2	4.7	4.8	4.5	4.3	4.8	2.6
met de beoordeling			(1.9)	(1.3)	(0.9)	(0.9)	(1.7)	(1.8)	(1.3)	(2.0)
Leerling wil beter	3.0 (1.4)	#	3.0	2.7	2.8	2.1	3.1	2.9	2.6	4.8
resultaat			(1.3)	(1.2)	(1.3)	(1.0)	(1.6)	(1.3)	(1.2)	(1.8)

De verschillen in gemiddelde tussen de docenten zijn significant (p <.05).

Overige reacties

De leerlingen die dit deel van de vragenlijst hebben ingevuld, hadden in de meeste gevallen (84%) geen andere reactie dan de reacties die al in de genoemde serie van 10 reacties waren opgenomen.

In 66% van de gevallen hadden de leerlingen de ontvangen beoordeling wel verwacht. Dit verschilde bij een significantieniveau (α) van .10 significant tussen de scholen (p=.06; A 86%, B 63%, C 52%). Er waren geen significante verschillen tussen de vakken.

In 62% van de gevallen vonden de leerlingen de toets of opdracht niet moeilijk (in de overige 38% wel). Dit verschilde niet significant tussen scholen of vakken.

In 45% van de gevallen heeft de leraar volgens de leerlingen toegelicht waarom de betreffende stof moet worden gekend of de betreffende opdracht moet worden gemaakt. Dit verschilde significant (α =.10) tussen de scholen (p=07; A 10%, B 67%, C 48%).

In 65% van de gevallen sloot de inhoud van de toets of opdracht volgens de leerlingen aan bij wat ze eerder bij het vak hadden gehad. Dit verschilde significant tussen de scholen (p=.01; A 38%, B 70%, C 83%). In 27% van de gevallen sloot de inhoud aan bij (ook) andere vakken. Dit verschilde significant (α =.10) tussen de scholen (p=.09; A 43%, B 27%, C 13%), en ook tussen de vakken (p=.02; Engels 17%, Aardrijkskunde 46%, Wiskunde 16%).

Samengevat is volgens de leerlingen op school A door de docenten minder vaak toegelicht waarom de stof moest worden geleerd of opdracht gemaakt en sloot de inhoud van de taak minder aan bij wat ze eerder hadden gehad, dan op beide andere scholen, maar sloot die inhoud vaker aan bij andere vakken en hadden de leerlingen vaker de ontvangen beoordeling wel verwacht dan op beide andere scholen.

In het tweede deel van de vragenlijst waren vijf losse vragen opgenomen over het krijgen van feedback, en een serie items met een schaal van 1-5 over de aard van de feedback (indien verkregen). Dit deel is 54 keer ingevuld (van de maximaal mogelijke 21 (taken) x 4 (leerlingen) = 84 keer).

Het krijgen van feedback

Tijdens het werken aan de lesstof voor de toets of opdracht heeft 59% van de responderende leerlingen feedback gekregen van de leraar. Dit verschilt significant tussen de scholen (p=.04; school A 62%, B 71%, C 22%) en tussen de vakken (p<.001; vak E

22%, A 100%, W 16%). Op de scholen A en B kwam dit meer voor dan op school C, en het kwam vooral voor bij het vak aardrijkskunde. In de meeste gevallen nam de leraar het initiatief om feedback te geven (79%), in de overige gevallen wilde de leerlingen graag zelf weten of zij goed bezig waren en wat zij goed hadden gedaan en wat verkeerd (21%). In de meeste gevallen werd de feedback pas na de beoordeling ontvangen (68%), en gebeurde dit één keer (79%). De manier waarop de feedback werd gegeven, varieerde sterk: mondeling en/of schriftelijk, individueel en/of klassikaal of via de digitale leeromgeving.

De aard van de feedback

Op basis van de antwoorden op 20 uitspraken (met een schaal van 1-5) konden twee voldoende betrouwbare schalen worden gemaakt, over de mate waarin de leerling de feedback positief vond, en de mate waarin de leerling de feedback leerzaam vond (zie tabel 6.3). Omdat deel B van de vragenlijst op school C abusievelijk niet is ingevuld, is hierbij niet getoetst op verschillen tussen scholen, vakken en docenten. Zie tabellen 6.11 en 6.12. In de tabellen is te zien dat sprake is van enige verschillen tussen de scholen, de vakken en de docenten, en dat het patroon enigszins verschilt voor de mate waarin de leerlingen de feedback positief vinden en de mate waarin ze deze leerzaam vinden.

Tabel 6.11: Schaalgemiddelde (sd) van de schalen m.b.t. feedback, per school en vak

		School			Vak		
Schaal	*Totaal*	*A*	*B*	*C **	*E*	*A*	*W*
Leerling vond de feedback	3.8 (.68)	3.6	4.0		4.2	3.3	4.1
positief		(.75)	(.59)		(.72)	(.52)	(.52)
Leerling vond de feedback	3.5 (.67)	3.4	3.7		3.3	3.6	3.8
leerzaam		(.67)	(.67)		(.95)	(.42)	(.57)

* Op school C zijn deze vragen abusievelijk niet beantwoord

Tabel 6 12: Schaalgemiddelde (sd) van de schalen m.b.t. feedback. per docent

		Docent	1	2	4	5	6	7	8	9
Schaal	Totaal	Vak	E	A	E	A	W	E *	A *	W *
Leerling vond de	3.8 (.68)		4.5	3.2	3.8	3.9	4.1			
feedback positief			(.39)	(.41)	(.89)	(.46)	(.50)			
Leerling vond de	3.5 (.67)		3.2	3.5	3.3	3.9	3.8			
feedback leerzaam			(1.1)	(.39)	(.94)	(.42)	(.57)			

* Op school C zijn deze vragen abusievelijk niet beantwoord

De gegeven beoordelingen (cijfers)

Van alle docenten zijn de beoordelingen verkregen van de twee of drie door hun leerlingen gemaakte opdrachten en/of toetsen die in het onderzoek zijn meegenomen. Op school C worden cijfers gegeven, op de bekende schaal van 1 tot 10, met één decimaal. Door de docenten op de scholen A en B worden de beoordelingen in woorden uitgedrukt, namelijk onvoldoende, matig, voldoende, goed, uitstekend. Om alle beoordelingen op dezelfde wijze kwantitatief te kunnen analyseren, zijn deze beoordelingen in woorden ook in cijfers omgezet, en wel als volgt: onvoldoende = 4,5; matig = 5,5; voldoende = 6,5; goed = 8; uitstekend = 9. Enkele beoordelingen met "behaald" of "gezien" (bij één docent) zijn gescoord als voldoende (6,5). Eén beoordeling met "voldoende na verbetering" is gescoord als voldoende (6,5). Twee beoordelingen met "nog niet voldoende" zijn gescoord als matig (5,5). Tweemaal "(nog) niet ingeleverd" is gescoord als missing.
De beoordelingen waren bij elke opdracht sterk gespreid (range 2-10 bij de beoordeling van de eerste opdracht, met een gemiddelde van 7,1 (sd 1.6) en range 3-9 bij de beoordelingen van de tweede en de derde opdracht. met gemiddelden van 6,4 (sd 1.7) en 6,6 (sd 1.5)). De drie beoordelingen vormen samen een betrouwbare schaal (item-rest correlaties .49-.74; Cronbach's alpha .78). Dit gemiddelde per leerling varieerde van 3,3 tot 9,0 en is gemiddeld 6,6 (sd 1.3). De vier leerlingen per docent bleken inderdaad te zijn verdeeld over zwak (gemiddeld 5,8). gemiddeld (6,5) en sterk (7,7).
Het cijfer verschilde niet significant tussen de scholen (F = 1.08, df = 2, 27, p = .35), maar wel tussen de drie vakken (F = 5.40, df = 2, 27, p =.01). Het gemiddelde voor Engels was 7,0 (sd 1.2), voor aardrijkskunde 7,1 (sd 0.8), voor wiskunde 5,5 (sd 1.4). In tabel 6.13 staan de range, het gemiddelde en de standaarddeviatie van de cijfers per docent. De range varieert tussen de docenten van 1,5 tot 3,8.

Tabel 6.13: De in cijfers uitgedrukte beoordelingen per docent

		Docent	1	2	4	5	6	7	8	9
	Totaal	*Vak*	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
Range	3.3 -		5.8-	5.8-	6.5-	5.8-	5.2-	5.3-	6.7-	3.3-
	9.0		8.2	7.3	9.0	8.0	7.0	7.3	8.3	7.1
Gemiddelde	6.6		7.4	6.5	7.5	7.2	6.1	6.3	7.5	4.8
Standaarddeviatie	1.3		1.3	0.7	1.3	0.9	0.8	1.1	0.7	1.8

De afsluitende vragenlijst voor leerlingen over kenmerken van taken waarvan ze kunnen leren

Tot slot is aan de leerlingen gevraagd waarvan zij hoeveel leren bij het maken van taken voor het betreffende vak (16 uitspraken over condities en aspecten) en bij welke typen taken (los van het vak) ze zichzelf beter kunnen voorbereiden en er meer van leren (11

uitspraken over kenmerken van taken). Op basis van de antwoorden op 15 van de 16 uitspraken over condities en aspecten, met een schaal van 1-4 (weinig tot veel) (één item was weinig informatief geformuleerd) konden drie voldoende betrouwbare schalen worden gemaakt (zie tabel 6.3) over de mate waarin de leerling leert als de taak duidelijk is, als de taak uitdaagt, en als je bij de taak moet nadenken. De set van 16 bleek ook als geheel een betrouwbare schaal te vormen, maar voor dit onderzoek zijn de drie te onderscheiden subschalen informatiever. De gemiddelden daarop verschillen niet significant tussen scholen, tussen vakken en tussen docenten. Wel is zichtbaar dat leerlingen bij wiskunde relatief meer leren van een opdracht als die duidelijk is en bij Engels en aardrijkskunde meer als die uitdaagt. Zie de tabellen 6.14 en 6.15.

Tabel 6.14: Schaalgemiddelde (sd) van de schalen leren van een taak, per school en vak

		School			Vak		
Schaal	*Totaal*	*A*	*B*	*C*	*E*	*A*	*W*
Leerling leert als de taak duidelijk is	2.9 (.53)	2.7(.34)	3.0(.60)	2.9(.65)	2.8 (.55)	2.9 (.68)	3.1 (.37)
Leerling leert als de taak uitdaagt	3.0 (.62)	3.3(.59)	3.3(.54)	2.7(.54)	3.2 (.68)	3.3 (.45)	2.7 (.61)
Leerling leert als je erbij moet nadenken	2.7 (.51)	2.4(.53)	2.8(.48)	2.9(.61)	2.7 (.56)	2.8 (.56)	2.7 (.52)

Tabel 6.15: Schaalgemiddelde (sd) van de schalen over leren van een taak, per docent

		Docent	1	2	4	5	6	7	8	9
Schaal	*Totaal*	*Vak*	*E*	*A*	*E*	*A*	*W*	*E*	*A*	*W*
Leerling leert als de taak duidelijk is	2.9 (.53)		2.8 (.20)	2.6 (.43)	2.9 (.12)	2.9 (.90)	3.3 (.48)	2.6 (1.1)	3.2 (.67)	3.0 (.10)
Leerling leert als de taak uitdaagt	3.0 (.62)		3.6 (.52)	3.1 (.59)	3.4 (.38)	3.5 (.35)	2.9 (.66)	2.5 (.66)	3.2 (.38)	2.4 (.55)
Leerling leert als je erbij moet nadenken	2.7 (.51)		2.4 (.82)	2.4 (.35)	2.7 (.29)	2.8 (.67)	2.8 (.49)	2.9 (.54)	3.3 (.25)	2.5 (.45)

De antwoorden op de 11 uitspraken over typen taken (los van het vak) hingen niet duidelijk onderling samen, noch ten aanzien van "beter voorbereiden" noch ten aanzien van "meer leren". We presenteren de uitkomsten in percentages per kenmerk per antwoord. zie tabel 6.16. Daarin is te zien dat de leerlingen aangeven vooral zich goed te kunnen voorbereiden op en veel te kunnen leren van taken die duidelijk zijn (items 1, 6, 8, 9) en die individueel worden gemaakt (items 2). De gemiddelde scores op deze items bij beide vragen (met betrekking tot voorbereiden en leren) verschilden niet betekenisvol tussen de scholen en tussen de vakken (van de 44 toetsingen waren slechts vier significant, bij verschillende items en met verschillende scholen of vakken die (iets) hoger of lager scoorden).

Tabel 6.16: Kenmerken van taken en zich kunnen voorbereiden en ervan kunnen leren (percentages)

Taken ↓ Vragen →	Hoe goed kun je je op zo'n taak voorbereiden?			Hoeveel kun je van zo'n taak leren?		
	Niet zo goed	*Gaat wel*	*Goed*	*Niet zo veel*	*Gaat wel*	*Veel*
1 Het is duidelijk waar de taak over gaat en waarom die van belang is	3	31	66		35	65
2 De taak moet individueel worden gemaakt	7	31	62		41	59
3 De taak moet in samenwerking met anderen worden gemaakt	17	55	28	7	66	28
4 Voor de taak is ruim voldoende tijd beschikbaar		41	59		52	48
5 De taak duurt niet te lang, niet langer dan nodig is	17	45	38	10	59	31
6 Het is duidelijk hoe de taak kan of moet worden aangepakt	3	31	66	3	41	55
7 Het is duidelijk wat je bij het maken van de taak mag of moet gebruiken		45	55	10	59	31
8 Het is duidelijk wat je moet maken, wat het resultaat moet zijn	3	17	79	3	35	62
9 Het is duidelijk aan welke eisen het resultaat moet voldoen	7	21	72	7	38	55
10 De taak kan op meer manieren worden aangepakt	13	66	21	10	52	38
11 Bij de taak kunnen meer antwoorden of oplossingen goed zijn	14	55	31	24	38	38

Relaties tussen motivatie, waargenomen docenthandelen, en verkregen beoordelingen

De motivatie van de leerlingen voor taken voor het vak bij de docent enerzijds en hun beoordeling van het handelen van hun docent op zes van de zeven schalen (zie tabel 6.2) voor het geven van feedback en het beoordelen zoals in september gemeten met de klassikale vragenlijst (n=184), hangen onderling significant positief samen (correlaties rond .20-.30, $p<.05$).

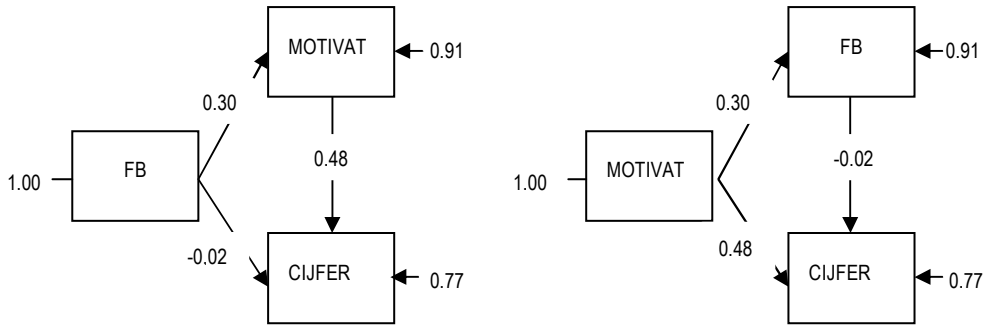
Het voor de gemaakte taken in het najaar behaalde gemiddelde cijfer (n=32) hangt significant positief samen ($p<.10$) met zowel de in september gemeten motivatie (correlatie .48) en de onderliggende schalen voor self-efficacy (.56), intrinsieke motivatie (.38) en geïdentificeerde regulatie (.37), als met de in september gevraagde beoordeling door de leerling van de mate waarin de docent in het algemeen resultaatgerichte feedback geeft (.37) en de mate waarin de leerling dat graag wil (.34).

De mate waarin de leerling in de eindvragenlijst (n=32) aangeeft in het algemeen te leren van taken waarbij hij moet nadenken, hangt positief samen met de in september gemeten motivatie van de leerling voor taken voor dit vak bij de docent (.55, $p<.01$) en de beoordeling door de leerling van het handelen van de docent op de verschillende schalen voor het geven van feedback en het beoordelen (correlaties rond .50, $p<.05$), en het in het najaar voor de gemaakte taken behaalde gemiddelde cijfer (.34, $p<.10$).

De mate van amotivatie van de leerling hangt negatief samen met het behaalde gemiddelde cijfer (-.33, $p<.10$) en positief met de mate waarin de leerling zegt te leren van taken die duidelijk zijn (.42, $p<.05$).

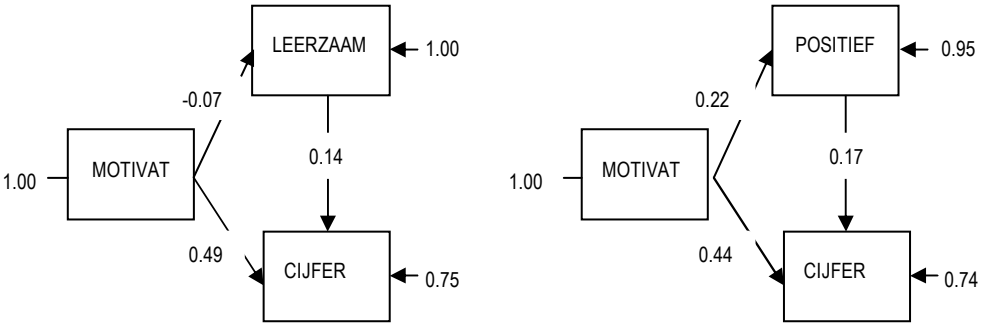
Ter verdere exploratie van de relaties tussen het door de leerlingen waargenomen handelen van hun docent, hun motivatie voor het maken van taken voor het vak bij deze docent, en de voor de gemaakte taken behaalde cijfers, zijn padanalyses uitgevoerd (met

LISREL, versie 8.30). In hoofdstuk 5 kon een sterk verband worden geconstateerd tussen het feedbackgedrag van de docenten en de motivatie van de leerlingen. Figuur 6.5 toont de resultaten van de padanalyses op basis van het in onderhavig onderzoek in de klassikale vragenlijst voor leerlingen vooraf vastgestelde gebruikelijke feedbackgedrag van de docent en de eigen motivatie voor het maken van taken voor het vak bij deze docent, en als extra variabele het gemiddelde voor de gemaakte (twee of drie) taken behaalde cijfer. Ook nu weer blijkt er een duidelijke relatie te zijn tussen Feedbackgedrag en Motivatie, terwijl aanvullend een duidelijk verband zichtbaar wordt tussen Motivatie en Cijfer. Er blijkt geen *direct* verband te zijn tussen het (bij de docent gebruikelijke) Feedbackgedrag (gemeten in september) en het gemiddelde Cijfer voor de (in oktober-december) gemaakte taken, alleen *via* Motivatie. Bij de vergelijking tussen de analyse in hoofdstuk 5 en die in dit hoofdstuk moet worden bedacht dat de operationalisering van het Feedbackgedrag enigszins verschilt (in hoofdstuk 5 op basis van vier schalen en in dit hoofdstuk op basis van drie schalen, met echter een duidelijke overlap in de twee schalen voor procesgerichte respectievelijk resultaatgerichte feedback).



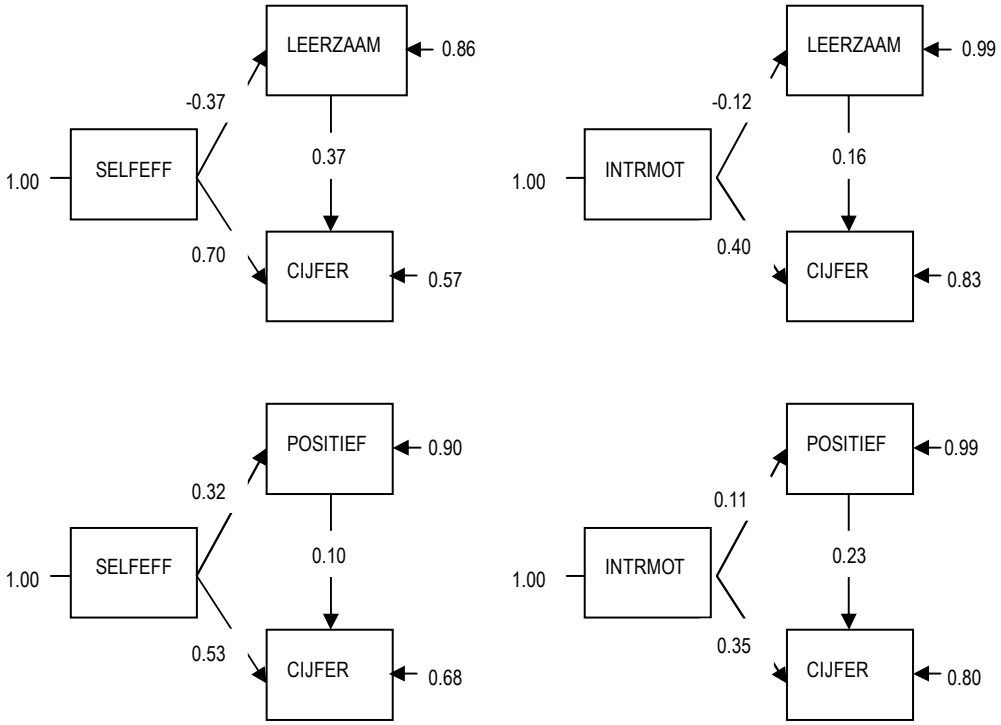
Figuur 6.5: Uitkomsten padanalyses met Feedbackgedrag docent, Motivatie leerling en Cijfer (n=30)

Zoals in hoofdstuk 5 al is opgemerkt, kan in plaats van een effect van feedback op motivatie ook sprake zijn van een effect van motivatie op feedback (meer gemotiveerde leerlingen kunnen meer feedback waarnemen en/of bij hun docent meer feedback oproepen), en figuur 6.5 bevat ook de (in dit model uiteraard identieke) uitkomsten van de op die manier gespecificeerde analyse.
De vragenlijst die de leerlingen na afloop van elke gemaakte taak hebben ingevuld, bevatte ook een aantal items over de feedback die zij (eventueel) van hun docent hadden gekregen bij het werken aan de betreffende taak, waaruit twee betrouwbare schalen konden worden gevormd: over de mate waarin de leerlingen de feedback positief respectievelijk leerzaam hadden gevonden (zie tabel 6.3). Figuur 6.6 bevat de uitkomsten van de analyses met deze taakspecifieke meting van het feedbackgedrag van hun docent.



Figuur 6.6: Motivatie leerling, taakspecifiek Feedbackgedrag docent, en Cijfer (n=20)

Bij deze taakspecifieke meting van het feedbackgedrag van de docent (zoals gepercipieerd door de leerling) worden wel (positieve) relaties zichtbaar met het (gemiddelde) voor de betreffende taken behaalde cijfer. Anderzijds blijkt het verband tussen de motivatie en de ervaren feedback nu duidelijk lager of zelfs ontbrekend. Een alternatief voor het model in figuur 6.6 zou overigens zijn om een effect te verwachten van cijfer op ervaren feedback (dus met ook cijfer als exogene variabele), uitgaande van de redenering dat een hoger presterende leerling de feedback mogelijk anders waarneemt en/of bij de docent andere feedback oproept. In de onderzochte situaties is de feedback echter in een aantal gevallen verkregen *tijdens* het werken aan de taak, vooral bij de opdrachten, en toen was er nog geen cijfer voor gegeven. Door het beperkte aantal leerlingen dat in dit deel van het onderzoek participeerde is het niet zinvol om de analyse apart te doen voor de leerlingen die opdrachten maakten en zij die toetsen maakten (waarbij niet alleen het cijfer maar ook de (eventuele) feedback achteraf werd verkregen).
Tot slot kan wel ook nog de operationalisering van de motivatie worden gespecificeerd. De in de analyses in de figuren 6.5 en 6.6 gebruikte maat voor motivatie is immers een overkoepelende schaal van drie bestaande schalen (zie de toelichting bij tabel 6.1). De analyses in figuur 6.6 zijn apart herhaald met self-efficacy en intrinsieke motivatie als representanten van twee te onderscheiden aspecten van motivatie, in termen van kunnen en in termen van willen. Figuur 6.7 toont de resultaten daarvan.
Uit figuur 6.7 blijkt dat leerlingen met een hoge self-efficacy de van de docent verkregen feedback gemiddeld minder leerzaam en vaker positief vinden, terwijl leerlingen die de verkregen feedback leerzaam vinden gemiddeld een hoger cijfer behalen. Uit de padcoëfficiënten in figuur 6.7 kan worden afgeleid dat het totale effect van Selfeff op Cijfer in beide modellen even groot is (.56) en het totale effect van Intrmot op Cijfer in beide modellen eveneens (.38).
Het motivatieaspect selfefficacy blijkt een betere voorspeller voor het behaalde gemiddelde cijfer dan de intrinsieke motivatie, terwijl de mate waarin de leerlingen de verkregen feedback als leerzaam hebben ervaren daaraan meer toevoegt dan de mate waarin de leerlingen de feedback positief vonden.



Figuur 6.7: Motivatie leerling (twee maten), Feedbackgedrag docent (twee maten), en Cijfer (n=20)

6.5 Conclusies en discussie

Dit onderzoek is gericht op de volgende vragen:
1. Welke aanwijzingen bevat de literatuur voor beoordelingen die kunnen bijdragen aan leren?
2. Hoe geven docenten hun beoordelingspraktijk vorm en met welke overwegingen?
3. Hoe ervaren leerlingen de beoordelingen die ze krijgen en hoe reageren zij daarop?
4. In hoeverre verschillen de beoordelingspraktijken van docenten en de ervaringen daarmee van leerlingen tussen vakken en typen taken en welke factoren spelen hierbij een rol?

Om op deze vragen antwoorden te verkrijgen hebben op drie scholen in totaal negen docenten aan dit onderzoek meegewerkt, bij drie vakken: Engels, aardrijkskunde en wiskunde. Deze docenten hebben mondeling en schriftelijk vragen beantwoord over hun opvattingen, intenties en aanpakken bij beoordelen en de manieren waarop zij over de

resultaten met hun leerlingen communiceren. Ook hebben per docent gemiddeld 23 leerlingen een schriftelijke vragenlijst beantwoord over hun motivatie voor het maken van taken bij dat vak voor die docent en over de manieren waarop de docent feedback geeft en beoordeelt. Vervolgens heeft elke docent tijdens de onderzoeksperiode de leerlingen enkele taken laten maken en die beoordeeld. Per docent per taak is steeds nadat de leerlingen de beoordelingen hadden verkregen met vier leerlingen een groepsinterview gehouden en is door hen aanvullend een aantal schriftelijke vragen beantwoord. Tot slot hebben deze leerlingen een korte vragenlijst ingevuld over kenmerken van taken waarvan zij volgens henzelf kunnen leren.

Welke aanwijzingen bevat de literatuur voor beoordelingen die kunnen bijdragen aan leren?

Leerlingen kunnen op drie momenten leren van een taak. Ten eerste tijdens de voorbereiding op het maken van de taak, doordat de docent de taak toelicht en de leerlingen werken aan de kennis en vaardigheden die voor de taak nodig zijn. Ten tweede door het maken van de taak zelf, als die voldoende duidelijk en motiverend is en vraagt om activiteiten die zelf ook bijdragen aan verdere ontwikkeling en oefening van de kennis en vaardigheden. Ten derde na afloop, als het werk is beoordeeld en daarbij duidelijk wordt, door feedback en/of nabespreking, wat goed en minder goed was, waarom en waardoor, en hoe kan worden gewerkt aan verbetering, en als daarvoor ook gelegenheid wordt geboden.

Uit de literatuur komt naar voren dat om te kunnen bijdragen aan het onderwijsleerproces beoordelingen daarvan onderdeel moeten uitmaken: ze moeten zijn gericht op dezelfde kennis en vaardigheden en worden voorafgegaan door en gevolgd door een consequent doorgaand doelgericht leertraject. Dit vergt dat de taken die de leerlingen krijgen daarvoor geschikt zijn, dat wil zeggen een beroep doen op de bedoelde kennis en vaardigheden, onderscheid kunnen maken tussen leerlingen die deze kennis en vaardigheden goed en minder goed beheersen, en zodanig kunnen worden beoordeeld dat kan worden vastgesteld en uitgelegd wat goed was en wat niet, en waarom en waardoor. Het vergt verder dat de beoordeling op valide wijze plaatsvindt, dat de beoordeling zo wordt gecommuniceerd (door formulering, feedback, nabespreking, interactie) dat deze voor leerlingen begrijpelijk is, betekenis krijgt en kan worden gebruikt in het verdere leerproces, dat docenten letten op hoe de beoordeling bij elke leerling overkomt (wordt de beoordeling begrepen, waaraan wordt de prestatie toegeschreven, wat doet dit met de leerlingen), en dat leerlingen ook de gelegenheid krijgen er iets mee te doen, bij het alsnog verbeteren van de prestatie en/of bij een volgende vergelijkbare taak.

Hoe geven docenten hun beoordelingspraktijk vorm en met welke overwegingen?

De meeste docenten in dit onderzoek denken desgevraagd bij het onderwerp beoordelen aan de criteria aan de hand waarvan ze beoordelen, de communicatie over de beoordeling (feedback, nabespreking) en de opbrengsten of gevolgen daarvan bij de leerlingen. Beoordelen heeft voor hen zowel de functie dat zij als docenten een beeld krijgen van de

leerlingen als dat de leerlingen een beeld krijgen van zichzelf: de kennis en vaardigheden die ze al wel en die ze nog niet (voldoende) beheersen, en hoe ze zich kunnen verbeteren. De docenten geven aan dat taken en beoordelingen moeten aansluiten op het voorafgaande onderwijs en de kennis en vaardigheden die de leerlingen hebben kunnen ontwikkelen, en dat bij beoordelen duidelijke criteria nodig zijn die ook voor de leerlingen inzichtelijk zijn. Beoordelingen moeten volgens de docenten voor de leerlingen duidelijk en begrijpelijk zijn, bijdragen aan inzicht, en motiveren om te werken aan verbetering.
Binnen het onderscheiden repertoire van docentactiviteiten bij beoordelen geven de docenten desgevraagd in relatief beperkte mate aan dat zij leerlingen bij de beoordeling betrekken en beoordelingen achteraf mondeling met de leerlingen bespreken.
De taken die de docenten hun leerlingen geven die in dit onderzoek zijn betrokken, waren in de helft van de gevallen toetsen, met gesloten en open vragen van diverse aard (kennis, toepassing, inzicht), en daarnaast open schrijfopdrachten, samengestelde opdrachten (waarbij individueel of in tweetallen verschillende taken moeten worden uitgevoerd, al dan niet met behulp van een elektronische leeromgeving of op basis van een excursie) en enkele vrij omvangrijke en complexe opdrachten voor groepen van vier of vijf leerlingen (met schriftelijke, mondelinge en digitale componenten).

Hoe ervaren leerlingen de beoordelingen die ze krijgen en hoe reageren zij daarop?
De deelnemende leerlingen hebben blijkens de klassikale vragenlijst (n=184) gemiddeld behoorlijk vertrouwen in eigen kunnen als het gaat om het maken van taken voor het betreffende vak bij de betreffende leraar (5,1 op de schaal van 1-7). Hun inschatting van de mate waarin hun docenten over het algemeen de in de vragenlijst opgenomen activiteiten rond feedback geven en beoordelen uitvoeren, ligt op de schaal van 1-5 gemiddeld rond de 3,3 voor activiteiten die te maken hebben met het vooraf bespreken van de taak, het vooraf bespreken van criteria en standaarden, het geven van procesgerichte feedback, het bij de beoordeling betrekken van aanpak, proces en vooruitgang, en het nabespreken van de taak en de beoordeling, en ligt rond de 3,8 voor het geven van resultaatgerichte feedback. Van het betrekken van leerlingen bij de beoordeling is duidelijk minder sprake (2,2).
Rond tweederde van de leerlingen die na elke taak en beoordeling daarover schriftelijke vragen hebben beantwoord (n=32) had de verkregen beoordeling wel verwacht. Eveneens tweederde vond dat de taak aansloot bij het voorafgaande onderwijs en dat de taak niet moeilijk was. Ze waren gemiddeld behoorlijk tevreden over de beoordeling (4,3 op de schaal van 1-5) maar wilden vaak wel de volgende keer een beter resultaat (3,0). In een beperkt aantal gevallen begrepen leerlingen een beoordeling niet goed (2,0). Rond tweederde van de leerlingen gaf aan ook feedback te hebben gehad. Deze leerlingen hadden de feedback gemiddeld ervaren als vrij positief (3,8 op de schaal van 1-5) en ook wel als leerzaam (3,5).
Uit de gehouden groepsinterviews komt naar voren dat bij toetsen voor de leerlingen de inhoud en de beoordelingen gemiddeld duidelijker zijn dan bij opdrachten en dat ze bij

toetsen achteraf actiever bezig zijn om verkregen beoordelingen te duiden en ook beter kunnen aangeven wat ze hebben geleerd. De in de interviews geregistreerde reacties van leerlingen op verkregen beoordelingen kunnen als volgt worden getypeerd: a) de beoordeling voor kennisgeving aannemen; b) niet goed begrijpen hoe de beoordeling tot stand is gekomen, evt. verwijzend naar onduidelijkheden in de taak of de beoordeling; c) verklaren van een goede beoordeling door goed zijn in het vak of goed te hebben geleerd; d) verklaren van een tegenvallende beoordeling door niet goed zijn in het vak en/of er onvoldoende voor te hebben gedaan, of slordigheid tijdens het maken van de taak, of de moeilijkheid van de taak (en/of de samenwerking), of gebruik door de docent van bijkomende beoordelingscriteria; e) zich voornemen om voortaan harder te werken, anders te leren, eerder te beginnen (bij toetsen met leren, bij opdrachten ook *met het maken van* de opdracht) en/of een betere prestatie neer te zetten; f) zich niets voornemen.
Uit de afsluitende vragenlijst blijkt dat de leerlingen volgens henzelf van taken bij het betreffende vak (kunnen) leren als die duidelijk zijn, als die hen uitdagen, en als ze erbij moeten nadenken. Goed voorbereiden op en veel leren van een taak kunnen de leerlingen volgens henzelf het meest van taken die voor hen duidelijk zijn (waar gaat het over, hoe moet het worden aangepakt, wat moet het resultaat zijn, en waar moet dat aan voldoen).

In hoeverre verschillen de beoordelingspraktijken en de ervaringen van leerlingen tussen vakken en typen taken en welke factoren spelen hierbij een rol?
De vergelijkingen tussen vakken en typen taken kunnen niet los worden gezien van de context van de scholen en de keuzes die docenten daarbinnen maken om hun beoordelingspraktijk vorm te geven. Ook het maken van taken voor bepaalde vakken door leerlingen vindt plaats binnen de specifieke context van de beoordelingspraktijk voor dat vak op die betreffende school. Vanwege de verwevenheid tussen deze vier aspecten (vak, taak, school, docent) is het zinvol om behalve naar verschillen tussen vakken en tussen typen taken, ook te kijken naar verschillen tussen de scholen en tussen de docenten.
Er zijn *verschillen tussen de vakken*. Deze kunnen in verband worden gebracht met de verschillen in de aard van de vakken. Het vak Engels heeft als taal geen heel duidelijke structuur en vormt behalve een kennisdomein met betrekking tot woordenschat en grammatica vooral een vak van vaardigheden (spreken, luisteren, lezen en schrijven) die kunnen worden geoefend en getoetst aan de hand van allerlei kleine halfopen opdrachten rond woorden en zinnen. Aardrijkskunde is een breed vak over kennis van feiten en begrippen met betrekking tot een veelheid van maatschappelijke en natuurwetenschappelijke onderwerpen, en kent daarnaast vakeigen benaderingen (zoals schaalniveaus en perspectieven) en vakspecifieke vaardigheden (zoals kaartvaardigheden). Dit vak leent zich voor complexe opdrachten waarbij leerlingen kunnen samenwerken en waarbij ook verbanden met andere vakgebieden kunnen worden gelegd. Het vak wiskunde heeft een duidelijke eigen structuur en opbouw en wordt meestal geoefend en getoetst aan de hand van vragen naar berekeningen, toepassingen en inzicht, waarbij antwoorden duidelijk goed of fout zijn maar het proces van oplossen vaak belangrijker wordt gevonden dan de uitkomst.

In de taken die de docenten hebben ingebracht in het onderzoek zijn deze verschillen tussen de vakken terug te vinden, maar is ook de achtergrond zichtbaar van de scholen waarin de docenten werken, zoals dat op de scholen A en B wordt gewerkt in leergebieden en op school C in vakken. Dit heeft erin geresulteerd dat in de verdeling van de taken tussen toetsen en opdrachten de taken bij de docenten Engels en aardrijkskunde op school A alle de vorm hadden van opdrachten, die bij de docenten Engels en aardrijkskunde op school B deels de vorm van opdrachten en deels die van toetsen, en de taken van de docent wiskunde op school B en de docenten van alle drie vakken op school C de vorm van toetsen.
Ook bij de leerlingen vinden we verschillen tussen de vakken. Volgens de antwoorden van de leerlingen op de klassikale vragenlijst (n=184) bespreken de docenten wiskunde in hogere mate vooraf de toetsen en opdrachten die ze geven dan de docenten van beide andere vakken, terwijl de docenten aardrijkskunde (die hun leerlingen minder vaak per week zien) in mindere mate taken, criteria en standaarden vooraf bespreken, proces-gerichte en resultaatgerichte feedback geven, leerlingen bij het beoordelen betrekken en een gemaakte toets of opdracht nabespreken dan de leraren van beide andere vakken.
Er blijkt duidelijk sprake van *verschillen tussen (typen) taken*. Zoals gezegd zijn bij toetsen voor de leerlingen de inhoud en de beoordelingen gemiddeld duidelijker dan bij opdrachten en zijn ze bij toetsen achteraf actiever bezig om verkregen beoordelingen te duiden en kunnen ze beter aangeven wat ze hebben geleerd. Dit wil niet zeggen dat de leerlingen van het maken van opdrachten minder hebben geleerd. Bij opdrachten, en dan vooral de meer uitgebreide opdrachten waarbij een aantal taken na elkaar moet worden uitgevoerd, de leerling zelf de voortgang moet bewaken en het geheel moet blijven overzien, en soms moet worden samengewerkt en in overleg keuzes moeten worden gemaakt en werk verdeeld, doen leerlingen waarschijnlijk allerlei leerervaringen op. Deze leerervaringen zijn echter minder goed te voorzien en ook niet altijd beoogd, er kunnen amper vooraf standaarden voor worden geformuleerd, en het is ook moeilijk om ze mee te nemen in de beoordeling en er feedback op te geven. In een aantal gevallen hebben leerlingen bij zulke opdrachten wel achteraf zelf aangeven dat zij hebben geleerd op het vlak van samenwerken of van plannen, zoals na afloop van toetsen leerlingen hebben aangegeven te hebben geleerd wat ze beter kunnen doen in hun voorbereiding daarop. Zulke bijdragen aan leren betreffen echter niet de vakinhoud. Een mogelijke conclusie is dat leerlingen na ontvangst van beoordelingen van gemaakte toetsen daarover actiever nadenken *doordat* (in de nu onderzochte praktijk) de inhoud en beoordeling bij toetsen gemiddeld voor de leerlingen duidelijker zijn dan bij opdrachten.
Er zijn ook *verschillen tussen de scholen*. In hun opvattingen, intenties en aanpakken rond het geven van feedback en beoordelen leggen de docenten van school A (een school die een paar jaar geleden is gestart met een sterk vernieuwend concept) een accent op het belang dat beoordelingen de leerlingen stimuleren en motiveren en dat ze bijdragen aan verbetering en het verdere leerproces. De docenten zijn nog aan het uitproberen welke taken meer en minder geschikt zijn en aan het ontdekken hoe de leerlingen in de nieuwe leeromgeving op taken en beoordelingen reageren. School B is een sterk vernieuwende

school die al een paar jaar langer bezig is en de eerste examenresultaten heeft geboekt, en die mede naar aanleiding daarvan bepaalde zaken nu iets meer structureert. De docenten zien beoordelen duidelijk mede in functie van het krijgen van zicht op het niveau van de leerlingen (havo/vwo). Ze vinden het belangrijk dat leerlingen hun beoordelingen begrijpen en een goed beeld krijgen van hun kennis en vaardigheden, verwachten van de leerlingen veel eigen initiatief, en volgen hun ontwikkeling. Op beide scholen worden beoordelingen uitgedrukt in woordcategorieën (op school B iets meer gedifferentieerd dan op school A). Op school C wordt meer traditioneel klassikaal onderwijs verzorgd in vakken, klassen en een rooster met lesuren, en worden beoordelingen uitgedrukt in cijfers. De docenten op deze school noemen relatief vaak het belang van een goede voorbereiding van de leerlingen op de toetsen die ze moeten gaan maken. Beoordelingen gaan vaak gepaard met het aangeven wat goed was en wat fout en leerlingen moeten daar inzicht in krijgen, zodat zij daarvan kunnen leren. De docenten op school C beperken hun beoordelingen meer tot de getoetste vakinhoud zelf dan de docenten op beide andere scholen, die vaker ook de aanpak, het proces en de vooruitgang in hun beoordelingen meenemen. Anderzijds is bij de docenten op school C gemiddeld minder sprake van feedback en interactie naar aanleiding van beoordelingen.

Verschillen tussen de scholen zijn er behalve bij de docenten ook bij de leerlingen. In hun antwoorden op de klassikale vragenlijst (n=184) tonen de leerlingen van school B (n=76) zich gemiddeld meer intrinsiek gemotiveerd voor het maken van taken bij de betreffende vakken en docenten dan de leerlingen van de beide andere scholen, en zij zouden ook graag in hogere mate hun docenten de verschillende activiteiten op het vlak van feedback geven en beoordelen zien doen dan de leerlingen op beide andere scholen. In hun antwoorden op de vragenlijsten na elke taak (n=32) geven de leerlingen van school C (n=12) minder dan de leerlingen van beide andere scholen aan dat zij van hun docenten feedback hebben gekregen; dit komt overeen met wat hiervoor werd vastgesteld op basis van de antwoorden van de docenten.

Verschillen tussen de docenten zijn zichtbaar in de antwoorden van de leerlingen op de klassikale vragenlijst (n=184) en de antwoorden op de vragenlijsten na elke taak (n=32). In de antwoorden van de leerlingen op de klassikale vragenlijst komen de verschillen tussen de docenten overeen met hierboven al beschreven verschillen tussen de vakken in combinatie met verschillen tussen de scholen: de meeste feedback is verkregen van de docenten Engels op de scholen A en B en de docenten wiskunde op de scholen B en C (de docent wiskunde op school A deed aan dit deel van het onderzoek niet meer mee). In de antwoorden op de vragenlijsten na elke taak doen zich verschillen tussen docenten voor bij de mate waarin de leerlingen de verkregen feedback als positief en als leerzaam ervaren, en bij de mate waarin de leerlingen aangeven te leren van taken die duidelijk zijn, uitdagen respectievelijk vragen om nadenken. Deze verschillen hangen niet samen met de school of het vak en variëren per genoemd aspect, en zijn daarmee kennelijk echt individuele verschillen.

Indien we de uitkomsten en conclusies van dit onderzoek vergelijken met de bevindingen in de literatuur, dan blijkt dat de beschreven praktijken van docenten en ervaringen van leerlingen deze bevindingen grotendeels weerspiegelen. In de onderzochte beoordelingssituaties wordt een heel scala zichtbaar. In sommige situaties "zat alles mee", zo lijkt het: de docent heeft de te maken taak vooraf goed uitgelegd en toegelicht, in het voorafgaande onderwijs zijn de betreffende kennis en vaardigheden goed ontwikkeld en geoefend, de manier waarop de taak zal worden beoordeeld is vooraf duidelijk, de taak is niet onverwacht moeilijk en is voldoende afgebakend om voor de leerlingen uitvoerbaar te zijn, de beoordeling wordt verduidelijkt door feedback en/of nabespreking, de leerlingen begrijpen de beoordeling en ze kunnen er iets mee.
In de meeste situaties leek aan een of meer van deze voorwaarden niet of niet voldoende voldaan, waardoor sommige leerlingen tijdens het maken van de taak tegen onverwachte problemen aanliepen en/of de ontvangen beoordeling niet goed begrepen of daaruit minder juiste of minder vruchtbare conclusies trokken. Zoals zoveel processen kan ook het beoordelingsproces op allerlei manieren minder positief verlopen, bijvoorbeeld: de leerlingen waren door het voorafgaande onderwijs onvoldoende voorbereid op de taak, de taak was te moeilijk gezien het voorafgaande onderwijs en de voorbereiding van de leerlingen, de taak was onvoldoende duidelijk voor de leerlingen (of mogelijk vooraf niet voldoende doordacht door de docent), de taak vergde meer tijd dan beschikbaar was, de leerlingen moesten tijdens het werken aan de taak beslissingen nemen waarvan zij niet wisten of die aan het resultaat c.q. aan de beoordeling zouden bijdragen of juist niet, bij de beoordeling werden criteria betrokken die de leerlingen niet kenden en/of die los staan van de kennis en vaardigheden waar het om ging, de ontvangen beoordeling maakte onvoldoende duidelijk wat er nu wel en niet goed was gedaan en waarom (standaarden) en waardoor (oorzaken), er werd geen of onvoldoende feedback gegeven of in een nabespreking voorzien, er was geen gelegenheid om iets met het geleerde te doen.
Uit de literatuur kon een aantal richtlijnen worden gehaald voor beoordeling die kan bijdragen aan leren, betreffende het belang van de taken die worden beoordeeld (Black & Wiliam, 1998; Hodgen & Marshall, 2005), beoordelen als onderdeel van het onderwijsleerproces (Elshout-Mohr et al.. 1995; Natriello, 1987), de kwaliteit van beoordelingen (Evans & Engelberg, 1988; Hounsell et al., 2008), beoordeling in relatie tot communicatie en betekenisgeving (Higgins et al., 2001; Ilgen & Davis, 2000), en de relatie tussen beoordeling en motivatie (Butler, 1987; Harlen & Deakin Crick, 2003). De beoordelingspraktijk van de docenten in dit onderzoek sluit goed aan bij de in de literatuur gevonden richtlijnen. Binnen het in het onderzoek onderscheiden repertoire van docentactiviteiten bij beoordelen geven de docenten echter desgevraagd in relatief beperkte mate aan dat zij leerlingen bij de beoordeling betrekken en beoordelingen achteraf mondeling met de leerlingen bespreken, terwijl ook dit in de literatuur wordt aanbevolen (bijvoorbeeld bij Harlen & Deakin Crick, 2003; Sadler, 1989). Een mogelijke interpretatie voor de bevinding dat de docenten deze twee richtlijnen minder navolgen

dan de hiervoor genoemde is dat de docent aan de oordeelkundigheid of objectiviteit van de leerlingen twijfelt (leerlingen bij de beoordeling betrekken) en/of de betreffende richtlijn extra tijd kost (mondeling nabespreken) (zie voor dit laatste bijvoorbeeld Dochy, Segers & Sluijsmans, 1999; Sung, Chang, Chang & Yu, 2010).

Terugkijkend naar de literatuur over beoordeling kan ook worden geconstateerd dat, anders dan in hoofdstuk 5 ten aanzien van feedback, bij beoordeling relatief duidelijke verschillen tussen de scholen en tussen de vakken kunnen worden geconstateerd. Op school A, kortgeleden gestart en sterk vernieuwend, willen de docenten beoordeling zien als onderdeel van het leerproces en gaat daartoe de aandacht uit naar de keuze van geschikte taken in relatie tot hoe de leerlingen daarop reageren. Op school B, ook sterk vernieuwend maar al iets langer bezig, gaat de aandacht uit naar het adequaat communiceren over de beoordeling en verduidelijken hoe de beoordeling tot stand is gekomen. Op school C, een reeds lang bestaande traditionele school, concentreren de docenten zich op de voorbereiding van de leerlingen op de te maken taak en beoordelen ze deze op de beheersing, op dat moment, van de kennis en vaardigheden die aan de orde zijn, waar op beide andere scholen ook de aanpak, het proces en de vooruitgang in de beoordeling wordt meegenomen. De typen taken die de docenten de leerlingen ter beoordeling laten maken verschillen tussen de vakken, maar ook tussen de scholen, zodat het in dit onderzoek op school C steeds ging om toetsen en bij aardrijkskunde vaak om relatief complexe opdrachten die de leerlingen in samenwerking moesten maken. Dat in het onderzoek naar feedback geen duidelijke verschillen tussen taken naar voren kwamen, kan worden verklaard doordat vrijwel alle taken in dat onderzoek een opdrachtkarakter hadden; het ging immers om het geven van feedback terwijl leerlingen zelfstandig aan het werk zijn. In dit onderzoek naar beoordeling bleek dat de leerlingen gemiddeld bij toetsen achteraf gemiddeld meer nadenken over de verkregen beoordeling en beter kunnen aangeven wat ze ervan hebben geleerd, waarschijnlijk doordat ze toetsen als duidelijker ervaren, in termen van criteria en standaarden, dan opdrachten.

Net als in hoofdstuk 5 blijken relaties aantoonbaar tussen de motivatie van de leerlingen en het door hen waargenomen feedbackgedrag van de docent, waarbij nadere analyses, mede in relatie met de verkregen beoordelingen, differentiële effecten laten zien al naar gelang de aard van de motivatie en de wijze waarop de feedback door de leerlingen is ervaren. In dit onderzoek naar beoordelen werden ook relaties zichtbaar van motivatie en feedback met de geleverde prestatie, in termen van het van de docent verkregen cijfer (vergelijk Duijnhouwer, 2010, die relaties tussen feedback, motivatie en prestatie c.q. cijfer onderzocht bij schrijftaken bij studenten).

Een beperking van dit onderzoek was dat door de gekozen opzet geen systematische analyse kon worden gemaakt van de mogelijke effecten van bepaalde kenmerken van taken (objectief en/of zoals door de leerlingen ervaren). Hier ligt een belangrijk terrein voor vervolgonderzoek. Ook zijn verdere analyses wenselijk van de relaties tussen aspecten van motivatie (zoals self-efficacy en intrinsieke motivatie), formatieve beoordeling (feedback) en prestatie c.q. verkregen cijfer (inclusief de mate waarin cijfers kunnen gelden als valide summatieve beoordelingen van prestaties).

7 Fundering van gewenst docenthandelen

7.1 Inleiding

Regelmatig wordt gepleit voor 'evidence-based' onderwijs, gebaseerd op harde wetenschappelijke kennis. Er dient zich echter nog geen consensus aan over de aanwijzingen voor onderwijs die op dergelijk 'bewijs' kunnen worden gebaseerd, en er is zelfs discussie over de vraag of zulk bewijs wel kan worden geleverd. Een bijkomende vraag is waarop zulke aanwijzingen betrekking zouden moeten hebben. Bovendien zijn er naast de wetenschap ook nog andere bronnen waaruit kan worden geput, hebben scholen veel beleidsruimte gekregen, en brengen *stakeholders* hun eigen perspectieven in.
Scholen die werken aan een vernieuwend onderwijsconcept zien zich geconfronteerd met de vraag naar het gewenste pedagogisch-didactisch handelen van hun docenten. De roep om 'evidence-based' onderwijs en de noodzaak tot afleggen van verantwoording doet schoolleiders vragen naar wat de wetenschap te bieden heeft. Voor de wetenschap is het een uitdaging daarop antwoorden te geven, maar schoolleiders moeten deze antwoorden wel combineren met verwachtingen vanuit landelijk beleid en opvattingen en voorkeuren van docenten, leerlingen en externe stakeholders.
De doelstelling van dit onderzoek is het maken van een vergelijking van aanwijzingen voor gewenst docenthandelen vanuit vier te onderscheiden bronnen: wetenschappelijke literatuur, landelijke beleidsdocumenten, onderwijsconcepten van scholen, en de betrokkenen op en rond de scholen.
De vraagstelling is: wat zijn de overeenkomsten en verschillen tussen aanwijzingen voor gewenst handelen van docenten en de gronden waarop die zijn gebaseerd vanuit wetenschappelijke literatuur, landelijke beleidsdocumenten, onderwijsconcepten van scholen en opvattingen van betrokkenen?
De deelvragen zijn:

1. Welke aanwijzingen biedt de wetenschappelijke literatuur voor gewenst docenthandelen?
2. Wat zeggen beleidsgerichte publicaties over gewenst docenthandelen?
3. Wat zeggen de onderwijsconcepten van de scholen over gewenst docenthandelen?
4. Welk profiel van pedagogisch-didactisch handelen wordt op scholen met een vernieuwend onderwijsconcept wenselijk geacht door de schoolleiding, teamleiders, docenten en leerlingen, en externe stakeholders, en welke visies, opvattingen en verwachtingen liggen daaraan ten grondslag?
5. Hoe verhouden de antwoorden op de voorgaande vier vragen zich tot elkaar?

7.2 Opzet en uitvoering van het onderzoek

Hierna wordt de opzet van het onderzoek per bron beschreven.

De procedure bij de wetenschappelijke literatuur, beleidsgerichte publicaties en onderwijsconcepten

In *wetenschappelijke literatuur* is gezocht naar aanwijzingen voor gewenst handelen van docenten. Om dit gericht te kunnen doen, moesten keuzen worden gemaakt. De eerste vraag was, gezien de vele publicaties die jaarlijks verschijnen over leren, onderwijs en leraren: *welke* literatuur?
Het domein is afgebakend tot publicaties vanaf 1985. Vanaf die tijd begonnen internationaal opvattingen over leren en onderwijs te veranderen en ontstond een levendige discussie over oude en nieuwe theorieën, onderzoeken en visies, waarvan de diversiteit door het kiezen van deze periode goed kan worden gedekt. Verder lag het voor de hand publicaties te kiezen van bekende wetenschappers, publicaties waarnaar vaak wordt verwezen, en zowel tijdschriftartikelen als handboeken. Voor het onderzoek was een afgebakende hoeveelheid tijd beschikbaar en het ging niet om een generaliseerbare representatie van alle relevante literatuur (zo dat al zou kunnen) maar om een voor het doel voldoende dekking van te onderscheiden domeinen en invalshoeken. Daarom zijn de al bij de onderzoekers bekende publicaties die aan de hiervoor genoemde criteria voldoen als uitgangspunt genomen, en is via de daarin opgenomen literatuurlijsten de verzameling mogelijk nuttige publicaties uitgebreid. Toen op een gegeven moment de hierna te noemen vijf rubrieken naar voren kwamen, is binnen enkele daarvan nog gericht gezocht naar aanvullingen. Deze werkwijze leidde (in de beschikbare tijd) tot een lijst met 65 publicaties (die kan worden aangeduid als "gemakssteekproef").
Tijdens het doornemen van deze publicaties bleek dat aanwijzingen voor gewenst handelen van docenten deels expliciet worden beschreven zodat zij direct uit de tekst kunnen worden gehaald en deels impliciet blijven en daaruit moeten worden afgeleid. Verder stoelen ze deels op uitkomsten van empirisch onderzoek en deels op theorie die het belang van bepaald handelen inzichtelijk en aannemelijk maakt. Publicaties die uitmonden in aanwijzingen voor handelen van docenten bevatten vaak de vermelding dat het criterium daarbij is dat het handelen bijdraagt aan het leren van de leerlingen, meer direct (bijvoorbeeld door een bepaalde instructieaanpak) dan wel meer indirect (bijvoorbeeld via bevordering van de motivatie). Zulke aanwijzingen werden aangetroffen in de helft van de publicaties. Deze bestrijken vijf rubrieken van theorie en onderzoek: 1) leerprocessen en leeromgevingen, 2) de dagelijkse interactie tussen een leraar en een groep leerlingen, 3) factoren uit meta-analyses van onderzoeken naar effectief onderwijs, 4) standaarden voor opleiding en beoordeling van leraren, 5) opvattingen en keuzes van leraren.

Het laatste aspect waarbij keuzen moesten worden gemaakt, betrof de aard van het handelen zelf. Gelet op de behoeften van de schoolleiders en de vernieuwende onderwijsconcepten op hun scholen, ging het vooral om het pedagogisch-didactisch handelen van docenten in hun interacties met leerlingen. Dit betekende dat het overige handelen van docenten, zoals materiaalontwikkeling, keuzebegeleiding en coördinatie, niet is meegenomen. Wel is gelet op eventuele aanwijzingen die betrekking hebben op samenwerking tussen en professionele ontwikkeling van docenten, gezien het belang daarvan voor het vormgeven van een vernieuwend onderwijsconcept.

In *beleidsgerichte publicaties* wordt vanaf de eerste helft van de jaren '90 in onderzoeken, adviezen van commissies en raden, wetgeving en inspectiekaders aandacht besteed aan rollen, taken en bekwaamheden van docenten. De context waarin docenten werken, is al een aantal jaren in beweging. Er zijn ontwikkelingen gaande, in de samenleving, onder de jeugd, en dus ook in scholen, die maken dat docenten worden geconfronteerd met meer en andere verwachtingen, vereisten en mogelijkheden. Om daaraan tegemoet te komen, werken schoolleiders en docenten aan innovaties. Deze innovaties worden gestimuleerd en gevolgd vanuit het ministerie van OC&W en de Inspectie van het onderwijs en onderzoekers dragen inzichten aan waarop keuzen zouden kunnen worden gebaseerd.
Rapporten, adviezen en beleidskaders bevatten aanwijzingen voor gewenst docent-handelen vanuit meerdere invalshoeken en concepten. Om deze diversiteit te kunnen verwerken was nadere selectie en ordening nodig. Daartoe is aangesloten bij de manieren waarop docenthandelen in handboeken wordt benoemd (Driscoll, 2005; Verloop en Lowyck, 2003; Woolfolk, Hughes & Walkup, 2008), in termen van:

- door de docent te vervullen *rollen*, bijvoorbeeld: kennisoverdrager, begeleider; dit gebeurt vaak in publicaties over onderwijsvernieuwing, en blijft vaak globaal (zoals: docenten moeten minder de ene rol en meer de andere gaan vervullen);
- door de docent te creëren *leeromgevingen*; dit gebeurt vaak in publicaties vanuit bepaalde opvattingen over leren en leerprocessen, en wordt dan uitgewerkt in termen van instructiemodellen, didactische werkvormen, typen opdrachten);
- door de docent te vervullen *taken*, met name op (algemeen) pedagogisch-didactisch vlak in de groep en bij het begeleiden van leerlingen, en als professional in de school;
- door de docent te hanteren *instructiestrategieën*; dit betreft vaak vakdidactische publicaties waarin gedetailleerd wordt ingegaan op de precieze interactie met leerlingen rond een specifiek begrip of probleem in het betreffende vak;
- in het handelen van een docent te onderscheiden *dimensies*; deze insteek is zichtbaar in de SBL-competenties: interpersoonlijk, pedagogisch, didactisch, organisatorisch. Dit betreft, anders dan bij de bovengenoemde manieren om docenthandelen te beschrijven, geen opties waaruit kan worden gekozen, maar facetten van docenthandelen die altijd en tegelijkertijd aan de orde zijn (vergelijk Brekelmans et al., 2000; Doyle, 1986; Shuell, 1996; Volman, 2006).

In publicaties waarin rollen, leeromgevingen, taken of instructiestrategieën worden beschreven, gebeurt dit vaak in termen van wat wenselijk zou zijn, en daarbij komt ook regelmatig aan de orde wat docenten zouden moeten weten en kunnen om het gewenste handelen te kunnen uitvoeren.

Wat betreft de *onderwijsconcepten* van de scholen zijn om te beginnen de daarop betrekking hebbende documenten geanalyseerd die op verzoek door de schoolleiders zijn aangeleverd. Daarnaast is aanvullend op de websites en in de schoolgidsen en beleidsplannen van de scholen gezocht naar eventuele extra of andere informatie over de onderwijsconcepten. Ook andere documenten waarin aanwijzingen voor gewenst handelen van docenten worden beschreven, zijn opgevraagd en geanalyseerd. Dit betrof informatie voor nieuwe medewerkers, informatie over de zorgstructuur, de opbrengsten van professionaliseringsdagen, schoolspecifieke overzichten van gewenste competenties van docenten, en beoordelingscriteria en uitkomsten van tevredenheids-onderzoeken.
De auteur heeft alle documenten geanalyseerd en de belangrijkste aspecten daarin genoteerd: kenmerken van de school en het onderwijsconcept (de visie), taken en activiteiten van docenten, leeractiviteiten en leerprocessen, en leerresultaten (doelen). Daarna zijn de uitkomsten per school in een aantal categorieën samengevat. Een andere onderzoeker heeft deze exercitie nagevolgd, waarna de selectie en indeling van de informatie voor 89% van de cellen (scholen x aspecten) overeen bleek te komen. Over de overige 11 % is in overleg consensus bereikt. Het bleek mogelijk de informatie van alle onderzochte scholen te ordenen in dezelfde acht categorieën: zorgen voor een goed pedagogisch leerklimaat en bijdragen aan persoonlijke vorming; curriculumaanbod; instructie geven, werkvormen en opdrachten aanbieden; activeren, stimuleren van leeractiviteiten en/of zelfsturing; begeleiden van leerproces en/of zelfstandigheid, zelfsturing, coachen; inspelen op verschillen; volgen van het leerproces, feedback geven en beoordelen; reflecteren en leren en samenwerken met collega's. Deze categorieën lijken op die in hoofdstuk 3 in kolom D van figuur 3.2, en dat is niet verwonderlijk gezien de overlap tussen de aan beide onderzoeken deelnemende scholen. Voor onderhavig onderzoek zijn echter opnieuw gegevens verzameld en geanalyseerd en bij dit onderzoek zijn ook andere scholen betrokken, maar dit leidde dus toch tot vergelijkbare categorieën.
Naast de analyse van documenten zijn op elke school met de schoolleiders, teamleiders, docenten en leerlingen en met een aantal stakeholders van buiten de school gesprekken gevoerd over het onderwijsconcept van de school en de daaruit voortvloeiende aanwijzingen voor gewenst handelen van docenten (zie hieronder). Voor zover daaruit aanvullende informatie over de onderwijsconcepten van de scholen naar voren kwam, zijn deze toegevoegd aan de gemaakte overzichten per school.

Procedure, respondenten en instrumenten eerste ronde: belangrijke taken en hun context

In een eerste ronde (mei-juni 2009) zijn per school groepsinterviews gehouden met de schoolleiding, teamleiders en docenten. De schoolleiders kozen welke teamleiders (meestal: alle), docenten (meestal: gespreid over afdelingen en vakken) en leerlingen (gespreid over afdelingen en leerjaren) meededen. Deze gesprekken zijn gevoerd aan de hand van vooraf opgestelde gerichte open vragen. De interviews werden integraal opgenomen, na afloop uitgetypt en vervolgens geanalyseerd. De interviewvragen voor de schoolleiders en teamleiders (een gezamenlijk groepsinterview per school) en een aantal docenten (twee groepsinterviews per school) zijn zoveel mogelijk vergelijkbaar gehouden. Zie figuur 7.1. Ook zijn groepsinterviews gehouden met twee groepen leerlingen per school.

Bij de interviews met de schoolleiding en teamleiders werd (na de introductie van het onderzoek) gestart met de vraag: ”Wat staat centraal in het onderwijsconcept van deze school?” en werd vervolgens gevraagd waarom hiervoor gekozen was. Daarna werd gevraagd wat belangrijk is dat docenten doen gezien het onderwijsconcept van de school en waarom dat zo is. In de groepsinterviews met docenten werd na de introductie direct concreet gevraagd wat volgens hen de belangrijkste taken zijn van een docent op deze school. Daarna werd dezelfde vraag opnieuw gesteld vanuit een meer algemeen perspectief, dus los van de betreffende school.

Nadat deze vragen waren beantwoord, werd een korte gesprekspauze ingelast waarin iedere deelnemer individueel een opdracht uitvoerde. Deze bestond uit het kiezen van de volgens hem of haar belangrijkste taken van docenten (minimaal drie, maximaal zeven) uit een set van 22 kaarten met taken (de taken van de PDH, zie de hoofdstukken 3 en 4). Direct daarna beantwoordden de schoolleiders en teamleiders, en docenten, individueel een aantal schriftelijke vragen over de gemaakte keuze. Het groeps-interview werd daarna vervolgd met aanvullende vragen over het beoogde handelen van docenten in relatie tot het onderwijsconcept en de leerdoelen (onder meer betreffende de mate van realisatie en de realiseerbaarheid ervan). Zie figuur 7.1.

In de interviews met leerlingen (in groepen van vijf tot acht leerlingen uit verschillende leerjaren) werd ingegaan op de volgende topics: wat docenten doen; wat docenten (volgens de leerlingen) belangrijk vinden, waaraan zij dat merken en of zij dat zelf ook belangrijk vinden; wat zij zouden willen dat docenten doen; hoe docenten zich opstellen naar de leerlingen en wat ze daarvan vinden; in hoeverre docenten onderling verschillen en wat ze daarvan vinden; waarop ze zouden letten als ze de kwaliteit van docenten zouden moeten beoordelen; en wat volgens hen de kenmerken zijn van een ideale docent. In veel gevallen werd doorgevraagd naar het waarom van hun meningen.

Startvragen:
Schoolleiding en teamleiders
Wat staat centraal in het onderwijsconcept van deze school? (leerdoelen, klimaat, leeromgeving)? Waarom is hiervoor gekozen? Wat is belangrijk dat docenten doen, gezien het onderwijsconcept van de school? Waarom?
Docenten:
Wat zijn volgens u de belangrijkste taken van een docent op deze school?
Wat vindt u in het algemeen (los van deze school) belangrijk dat docenten doen?

Opdracht A (individueel):
Allen: Kaartjes met taken: minimaal 3 en maximaal 7 taken uitkiezen waarvan u het vooral belangrijk vindt dat docenten doen. Geselecteerde taken noteren op het antwoordblad.

Schriftelijke vragen (individueel te beantwoorden):
Allen:
Kunt u motiveren waarom u juist deze taken heeft uitgekozen? Wat moet een goede docent bij deze taken doen? Waarom vindt u dat een goede docent deze dingen moet doen? Waar baseert u dit op?
Schoolleiders en teamleiders:
Kunt u aangeven hoe goed of in welke mate een goede docent deze dingen moet doen? Zo ja, waarop zijn deze standaarden gebaseerd?
Hangen de antwoorden op de vorige vragen nog samen met hoe u aankijkt tegen het leraarschap in deze tijd (ook gelet op de huidige positie van de school in de samenleving en de omgang met de leerlingen)? Wilt u uw antwoord toelichten?
Docenten:
Hangen de antwoorden op de vorige vragen nog samen met hoe u aankijkt tegen het leraarschap in deze tijd (ook gelet op de huidige positie van de school in de samenleving en de omgang met de leerlingen)? Wilt u uw antwoord toelichten?
Komt u bij het uitvoeren van deze taken, als u deze goed wilt doen, nog knelpunten of dilemma's tegen? Zo ja, welke? Welke keuzes maakt u daarbij, en waarom?
Krijgt u feedback van leerlingen op de keuzes die u maakt, en kunt u die ook gebruiken?

Vervolg groepsinterview:
Allen:
Ontbreekt op de (22) kaartjes met taken nog iets dat u belangrijk vindt als taak van een docent? (bijvoorbeeld iets inhoudelijks of organisatorisch). Zo ja, wat, en waarom vindt u dat belangrijk?
Schoolleiders en teamleiders:
Komen docenten bij de door u gekozen taken naar u weet nog knelpunten of dilemma's tegen? Welke keuzes maken docenten daarbij volgens u, en waarom? Wat vindt u van de keuzes die zij maken?
Moeten docenten op alle fronten inzetbaar zijn of is specialisatie wenselijk en mogelijk?
Zijn de docenten op uw school voldoende geëquipeerd om de gewenste taken te vervullen? Indien nee: waar schort het dan aan en waardoor komt dat?
In welke mate staan de docenten naar uw idee achter het onderwijsconcept van de school? Op basis waarvan denkt u dat? In welke mate is het onderwijsconcept op dit moment volgens u feitelijk gerealiseerd?
Docenten:
Moeten docenten op alle fronten inzetbaar zijn of is specialisatie wenselijk en mogelijk?
Wat staat centraal in het onderwijsconcept van deze school? (leerdoelen, klimaat, leeromgeving)?
Wat is belangrijk dat docenten doen, gezien het onderwijsconcept van deze school? Is voor u duidelijk wat van u wordt verwacht? In hoeverre staat u achter het onderwijsconcept? Kunt u toelichten waarom?
In welke mate wordt het onderwijsconcept volgens u door docenten op deze school gerealiseerd?
In hoeverre is het onderwijsconcept en zijn de leerdoelen volgens u realiseerbaar? Wat wel/wat niet? Kunt u uw antwoord toelichten?
In hoeverre voelt u zich als docent voldoende geëquipeerd om de gewenste taken te vervullen?
Zijn er taken in het contact met de leerlingen die u als docent graag zou willen doen, of beter zou willen doen, maar waar u op dit moment onvoldoende aan toekomt? Indien ja, welke, en waarom?

Figuur 7.1: Interviewvragen eerste ronde groepsinterviews schoolleiders en teamleiders, en docenten

Procedure, respondenten en instrumenten tweede ronde: paarsgewijze vergelijkingen en gronden

In een tweede ronde (september-oktober 2009) zijn gegevens verzameld bij externe stakeholders, en vervolgens opnieuw bij de schoolleiders, teamleiders en docenten. Elke schoolleider heeft op verzoek van de onderzoekers drie of vier stakeholders gekozen en gevraagd aan het onderzoek mee te werken. Dit betrof ouders, directeuren van toeleverende basisscholen, leden van de medezeggenschapsraad of ouderraad, bestuursleden en zorgcoördinatoren. In groepsinterviews met deze stakeholders per school is in kaart gebracht welke concrete docenttaken door hen van belang worden geacht, in relatie tot het onderwijsconcept van de school. Deze interviews hadden eenzelfde opzet als de hiervoor beschreven groepsinterviews in de eerste ronde: enkele open startvragen aan de groep, een individueel te maken opdracht en beantwoording van schriftelijke vragen, en voortzetting van het groepsinterview met nogmaals een aantal open vragen.
De startvragen bij de stakeholders waren: Hoe typeert u uw eigen relatie en/of positie ten opzichte van de school? Welke belanghebbenden vertegenwoordigt u? Wat staat centraal in het onderwijsconcept van deze school? (leerdoelen, klimaat, leeromgeving)? Wat zijn volgens u de belangrijkste taken van een docent op deze school? Kunt u toelichten waarom?
Daarna kregen de stakeholders individueel een vragenlijst voorgelegd met een reeks paarsgewijs aangeboden taken van docenten waaruit zij steeds een (gedwongen) keuze moesten maken. Het ging om zeven taken die in (alle) 21 verschillende combinaties werden aangeboden. Steeds werd gevraagd welke taak men de belangrijkste vindt voor een docent op deze school. Per keuzepaar werd tevens gevraagd naar de grond(en) voor de gemaakte keuze, waarbij een of meer voorgestructureerde antwoorden konden worden aangekruist en/of een zelf geformuleerd antwoord toegevoegd.
Na het individueel beantwoorden van de vragen in de vragenlijst werd het groeps-interview met de stakeholders hervat en zijn de volgende vragen gesteld:

- Hangen de antwoorden op de vorige vragen nog samen met hoe u aankijkt tegen het leraarschap in deze tijd (ook gelet op de huidige positie van de school in de samenleving en de omgang met de leerlingen)? Wilt u uw antwoord toelichten?
- Moeten docenten op alle fronten inzetbaar zijn of is specialisatie wenselijk en mogelijk?
- In hoeverre is het onderwijsconcept en zijn de leerdoelen van deze school volgens u realiseerbaar? Wat wel/wat niet?
- In welke mate wordt het onderwijsconcept volgens u door docenten op deze school gerealiseerd?
- Wat zijn volgens u de knelpunten en dilemma's?

Een vergelijkbare vragenlijst met paarsgewijs aangeboden taken met per taak een vijftal gronden als de stakeholders hebben ingevuld, werd ook voorgelegd aan de

schoolleiders, teamleiders en docenten die hadden meegedaan aan de eerste ronde, maar nu met acht (in plaats van zeven) paarsgewijs aangeboden taken (zie onder), in (alle) 28 verschillende combinaties. Hen werd eveneens gevraagd per paar een keuze te maken en deze keuze te beargumenteren, aan de hand van dezelfde categorieën gronden als in de vragenlijst bij de stakeholders. Deze vragenlijst is individueel verspreid via de postvakjes op de scholen (dus niet in een bijeenkomst).
We beschrijven nu de selectie van de taken (uit de 22 van de PDH) en de samenstelling van de voorgestructureerde antwoorden op de vraag naar de gronden voor de keuzen.

De keuze van de taken
De keuze van taken is gebaseerd op de uitkomsten van de eerste ronde en de volgende overwegingen:

- De selectie is primair gericht op de taken die door docenten, schoolleiders en teamleiders in de eerste ronde het meest zijn gekozen. De verwachting was dat de minder gekozen taken ten opzichte van de meer gekozen taken bij de paarsgelijke vergelijkingen meteen al niet of weinig zouden worden gekozen (en zo geen evenwichtige vergelijking zouden opleveren). Dit leidde in eerste instantie tot de keuze voor negen taken: 1, 2, 3, 7, 8, 11, 15, 17 en 22 (nummering PDH).
- Uitgangpunt was ook het streven naar een zo goed mogelijke verdeling van de te selecteren taken over voor dit onderzoek te onderscheiden clusters van taken, waarbij de taken op het vlak van samenwerking en professionele ontwikkeling niet zijn meegenomen. De focus lag daarmee op clusters van taken die zijn gericht op de interactie van de docent met de leerlingen. Gekozen werd voor de volgende indeling: pedagogisch leerklimaat; leeromgeving en leerdoelen; differentiatie; leeractiviteiten en zelfstandig leren; volgen van het leerproces, feedback en beoordeling.
- Ook is gelet op een evenwichtige verdeling van taken tussen de clusters met de doelen en inhouden van het onderwijsconcept enerzijds (de eerste twee zojuist genoemde clusters) en de clusters met de didactische invulling anderzijds (de laatste drie genoemde clusters).
- Tot slot is meegewogen of de formuleringen van de taken wel voldoende uitgesproken was. Op basis hiervan is de meer algemene taak 7 (leerlingen stimuleren en activeren) ingeruild voor de meer zeggende taak 16 (volgen wat leerlingen leren en hoe ze zich ontwikkelen).

Dit resulteerde in de volgende negen taken: 1, 2, 3, 8, 9, 11, 15, 16, en 17. Voor paarsgewijze vergelijkingen zou dit 36 combinaties van taken vergen en dit werd uiteindelijk toch te veel geacht. Besloten werd taak 1 (zorgen voor een goed leerklimaat) te schrappen omdat deze taak algemeen is en door velen wordt onderschreven. Bij de stakeholders is, om hen niet te overvragen, het aantal combinaties beperkt tot 21, door het schrappen van taak 16 (volgen van het leerproces), omdat zij mogelijk weinig zicht hebben op hoe deze taak op de school wordt vormgegeven. De aan hen voorgelegde taken waren daarmee: bijdragen aan de

persoonlijke vorming van leerlingen (2), geven van les, uitleg en instructie (3), leerlingen vaardigheden aanleren (8), inspelen op verschillen tussen leerlingen (9), leerlingen laten samenwerken en hen daarbij ondersteunen (11), leerlingen ondersteunen bij het sturen van hun leerproces (15), en geven van feedback aan leerlingen (17).

De gronden voor de keuzes

In de eerste ronde zijn door elke schoolleider, teamleider en docent drie tot zeven taken gekozen die belangrijk zijn voor een docent om goed te kunnen functioneren. Deze taakperceptie maakt deel uit van een persoonlijk normatief raamwerk dat vaak slechts in geringe mate met andere leden binnen een schoolcultuur wordt gedeeld (Kelchtermans, 2007). Ook in de tweede ronde, bij de opdracht met paarsgewijs aangeboden taken, moesten steeds keuzes worden gemaakt. Bij het maken van zulke keuzes spelen onderliggende visies, (taak)opvattingen en verwachtingen vaak een rol, maar deze blijven vaak onbewust en impliciet. De vraag daarbij is *waarom* men de gekozen taak de belangrijkste vindt. De basis waarop keuzes worden gemaakt, kan verschillen, en kan bijvoorbeeld een meer pragmatische, ethische of moreel georiënteerde invalshoek of grondslag hebben (Luttenberg, 2000). Deze oriëntaties kunnen al dan niet in lijn liggen met de beoogde onderwijsvernieuwing en ook onbewust van invloed zijn op de keuzes die een docent maakt.

Op basis van Kelchtermans (2007), Van Maanen (1977), Luttenberg (2000) en Eraut (1994) hebben we aanvankelijk negen typen gronden onderscheiden die een rol kunnen spelen bij het kiezen voor bepaalde (typen) taken als belangrijkste:

1. een technisch-rationele invalshoek, op basis van wetenschappelijke kennis (theorie, onderzoek) over oorzaak-gevolg relaties (instrumenteel voor middel-doel)
2. een economische invalshoek, in termen van efficiëncy
3. een praktische invalshoek, in termen van praktijkkennis en ervaring, in de context van deze school en leerlingen, gedeeld met collega's, middelen en doelen afwegend
4. een traditionele invalshoek, in termen van tradities en gewoonten (zo doen we het nu eenmaal of altijd)
5. een persoonlijke invalshoek, in termen van eigen centrale waarden en overtuigingen (vaak affectief geladen)
6. een sociaal normatieve invalshoek, in termen van sociale normen of “het behoort" (van collega's, de school, de overheid)
7. een morele invalshoek, in termen van plichten en verantwoordelijkheden
8. een ethische invalshoek, in termen van universeel geldig geachte waarden
9. een politieke invalshoek, in termen van macht, sociale condities en emancipatie

Omdat het niet haalbaar leek vanwege de belasting van de respondenten om bij elk paar taken negen antwoordcategorieën aan te bieden, is dit aantal beperkt tot vijf. De typen 7 en 8 zijn inhoudelijk moeilijk te onderscheiden en zijn daarom samengenomen. Type 4 is weggelaten omdat het inhoudelijk de meest zwakke grond is. De typen 2 en 9 zullen naar verwachting niet zo snel naar voren worden gebracht door stakeholders en zijn op grond daarvan geschrapt.

Dit heeft geleid tot het aanbieden per keuzepaar van de volgende vijf antwoord-modaliteiten, bij de vraag: "Ik maak deze keuze, omdat de gekozen taak: (meerdere antwoorden mogelijk)":

1) het meeste bijdraagt aan het leren en de ontwikkeling van leerlingen volgens wat daarover bekend is
2) in de praktijk vaak het belangrijkste voor een leraar blijkt te zijn om aandacht aan te besteden
3) volgens mijn persoonlijke overtuiging een essentiële taak is van een leraar op deze school
4) in de visie van de school tot de verantwoordelijkheid van de leraar behoort en dus van hem wordt verwacht
5) het meest een appel doet op wat de leraar voor zijn leerlingen zou moeten betekenen

Besloten werd een extra open antwoordcategorie toe te voegen om de respondenten de mogelijkheid te bieden zelf een afwijkend of aanvullend antwoord te formuleren.

Analyses

De gegevens van de eerste ronde (individuele keuze van drie tot zeven taken uit 22 taken op kaartjes, en schriftelijke open vragen daarover (schoolleiders, teamleiders en docenten), en vragen in groepsinterviews (schoolleiding en teamleiders, docenten, leerlingen) zijn als volgt geanalyseerd. De keuzen van taken en de antwoorden op de vragen daarover zijn kwantitatief verwerkt (geteld). De antwoorden op de vragen in de groepsinterviews zijn gecategoriseerd en geturfd.

De gegevens van de tweede ronde die zijn verkregen met de individuele schriftelijke vragenlijst met paarsgewijze vergelijkingen (met gedwongen keuze) en gesloten vragen naar de gronden per keuze (stakeholders, schoolleiders, teamleiders en docenten) zijn als volgt kwantitatief geanalyseerd.

Bij de analyse van de gemaakte keuzen is gebruik gemaakt van een multidimensionele schaaltechniek (Alscal), waarbij de gegevens zijn verwerkt als ordinale gegevens. Hierbij is het mogelijk de proporties te berekenen waarmee aangeboden items, in dit geval taken van docenten, zijn gekozen en vervolgens de onderlinge verhoudingen tussen de items (taken) visueel zichtbaar te maken.

De gemaakte berekeningen zijn gebaseerd op de gedachte dat als in een keuzepaar een taak wordt gekozen door de helft van de respondenten, de conceptuele afstand tussen beide taken in hun ogen minimaal is. In dat geval worden beide taken immers even vaak gekozen (beide 50%), en in termen van "populariteit" zijn beide taken dan even populair. Hoe meer de proportie waarmee in een paar een bepaalde taak is gekozen afwijkt van 50%, hoe groter de conceptuele afstand is tussen de twee taken, want de ene taak is dan vaker gekozen en dus meer populair dan de andere taak. Per aangeboden paar is het verschil berekend tussen de proportie waarmee de taak is gekozen en 50% (vergelijk de aanpak van Davison & Woods, 1983). Deze verschillen

zijn met behulp van Alscal omgezet in coördinaten op dimensies. Bij de keuze van het aantal dimensies is gestreefd naar een aantal dat enerzijds aan de onderlinge verschillen in populariteit tussen de aangeboden taken zoveel mogelijk recht doet (in termen van proportie verklaarde variantie) en waarbij anderzijds de uitkomsten ook inhoudelijk goed interpreteerbaar zijn.

7.3 Resultaten

7.3.1 Gewenst docenthandelen in de wetenschappelijke literatuur

Leren en leeromgevingen

In de afgelopen vijftig jaar is vanuit drie opeenvolgende paradigma's gekeken naar, nagedacht over, en onderzoek gedaan naar leren en leeromgevingen: behaviorisme, cognitieve psychologie en sociaal constructivisme. Elk daarvan heeft aanwijzingen opgeleverd voor het handelen van docenten, en alle drie zijn nog steeds relevant, afhankelijk van de beoogde leerdoelen en leerprocessen. Auteurs die een overzicht willen bieden van theorie en onderzoek op dit terrein beginnen dan ook regelmatig met het opnieuw kort beschrijven van deze drie stromingen (zoals: Greeno, Collins & Resnick, 1996; Seidel & Shavelson, 2007; Shuell, 1993; Veenman, 1992).
Vanuit het behaviorisme gezien is leren het opnemen van informatie en oefenen van vaardigheden en zijn de taken van de leraar het overdragen en trainen daarvan, in een gestructureerde omgeving. Het gaat hierbij om leerstof die in duidelijke onderdelen kan worden verdeeld en waarbij antwoorden goed of fout zijn (zoals bij feitenkennis) en om vaardigheden die stapsgewijs kunnen worden ingeslepen tot een efficiënte routine. In deze traditie zijn effectief docentgedrag en te behalen leerresultaten observeerbaar en wetmatig, en zijn goed klassenmanagement en voldoende effectieve leertijd belangrijke voorwaarden. Toepassing in nieuwe situaties (transfer) wordt niet gezien als problematisch.
In de cognitieve psychologie verschuift de aandacht naar de ontwikkeling van meer complexe kennis en het leren oplossen van problemen met meer mogelijke aanpakken en meer goede oplossingen, en het denken van de leerling dat hiervoor nodig is. Onderkend wordt dat de leerling zelf in het leerproces een actieve rol speelt. Theorieën hierover baseren zich op aannames, redeneringen en afleidingen, Leerlingen moeten leren hun eigen denken en leren te sturen (monitoren, evalueren, bijstellen), en docenten moeten de gewenste cognitieve en metacognitieve denkprocessen voordoen en leerlingen ondersteunen voor zover en zolang dat nodig is (*cognitive apprenticeship*; leren van afkijken en van verwoorden in interactie en reflectie). Over de mogelijkheid van transfer wordt verschillend gedacht, toepassing van het geleerde op nieuwe problemen of in nieuwe situaties blijkt vaak lastig.
De gedachte dat aangeboden kennis niet kant en klaar wordt opgenomen en opgeslagen maar door lerenden wordt verbonden met wat zij reeds weten en denken en zodoende

wordt ge(re)construeerd, bekend als "constructivisme", valt nog onder de cognitieve psychologie. In het *sociaal* constructivisme verschuift de focus naar het leren om competent deel te nemen aan praktijken die kenmerkend zijn voor een bepaalde sociaal en cultureel gevormde gemeenschap (*community of practice*), zoals een schoolklas, vakgebied (discipline) of beroepsgroep. Docenten en leerlingen vormen in die optiek een *community of learning*, waarin wordt gediscussieerd en onderhandeld over betekenissen, bedoelingen en verworvenheden.

Afhankelijk van de leerlingen en de beoogde leerprocessen en leerdoelen, zijn al deze denkkaders (nog) actueel. Ook in theorievorming en onderzoek zijn deze drie tradities nog actief, respectievelijk gericht op het leggen van (empirische) relaties tussen input (proces) en output (product), op leren als individueel (en individueel verschillend) cognitief proces, en op leren als gemeenschappelijk proces van participatie en acculturatie (zie Brown, Collins & Duguid, 1989; Duffy & Cunningham, 1996; Seidel & Shavelson, 2007; Veenman, 1992; Wilson & Cole, 1996).

Volgens Brown en Campione (1996) zijn post-behavioristische leerprincipes niet langer te vangen in vaste procedures. Docenten die vanuit moderne leerprincipes willen werken, kunnen daardoor geen houvast ontlenen aan bepaalde welomschreven aanpakken. Ze moeten het hebben van hun begrip van de principes, en deze flexibel en op maat toepassen, in onderlinge samenhang, en gericht op het doel om leerlingen te leren denken en redeneren (in hun hoofd, op papier, in een gesprek) in een bepaald kennisdomein. Voor docenten liggen hierbij volgens Brown en Campione drie typen opgaven:

a) *een leeromgeving creëren* waarin leerlingen werken aan betekenisvolle taken waarbij zij moeten zoeken en verzamelen, delen en bespreken, en presteren en beoordelen (en beoordeeld worden);
b) in interactie met de leerlingen *een aantal taken vervullen*: uitleggen, instrueren, voordoen (model zijn), discussie bewerkstelligen, coachen (ondersteunen, op maat hulp bieden, feedback geven), beoordelen;
c) een repertoire gebruiken van *specifieke acties en reacties in interactiesituaties*: verschillende typen vragen stellen, hints geven, etc.

Ad a) Nadere uitwerkingen van het wenselijk geachte type leeromgeving zijn beschreven door onder meer Greeno et al. (1996), De Corte (2000) en Bolhuis (2003).

Greeno et al. stellen dat een leeromgeving ook betrekking heeft op een bepaalde inhoud (kennisdomein) en is ingebed in een curriculum. Daarnaast geven ze aan dat opdrachten aan leerlingen op zich al een leeromgeving oproepen, waarin op bepaalde kennis en vaardigheden een beroep wordt gedaan, bepaalde (dezelfde of andere) kennis en vaardigheden (verder) kunnen worden ontwikkeld, en op verschillende manieren en in verschillende mate (ook variërend tussen leerlingen) materiaal, steun en interactie wenselijk kan zijn, en dus ook een verschillend beroep op de docent kan worden gedaan.

De Corte werkt het begrip "krachtige leeromgeving" uit, met een aantal kenmerken: realiseren van doelgerichte en cumulatieve leerprocessen, in (voor leerlingen)

betekenisvolle contexten en met voldoende ruimte voor samenwerking, met een goede balans tussen systematische instructie en begeleiding en ontdekkend en explorerend leren, bevordering van zelfregulatie van het leerproces en ontwikkeling van (meta-) cognitieve (meer of minder) domeinspecifieke vaardigheden, en rekening houden met individuele verschillen tussen leerlingen.
Bolhuis beschrijft het model van "procesgericht onderwijzen", waarin eveneens het geleidelijk bevorderen van zelfregulatie bij de ontwikkeling van domeinspecifieke kennis centraal staat en het behandelen van leerprocessen en leerresultaten als sociale verschijnselen, met daarnaast speciale aandacht voor motivationele en affectieve (emotionele) aspecten van leren.
Ad b) Beschrijvingen van gewenst geachte leeromgevingen bevatten of impliceren ook door docenten te vervullen taken, zoals het geven van instructie, het aan leerlingen geven van leerzame opdrachten, het begeleiden van leerlingen (als zij aan opdrachten werken), het leerlingen laten discussiëren en samenwerken, en het rekening houden met affectieve aspecten en met individuele verschillen.
Een centrale rol in gewenst docenthandelen wordt vervuld door de begeleiding, die het karakter zou moeten hebben van coaching, dat wil zeggen: op maat (individueel verschillend) en in de tijd afnemend (toewerkend naar zelfregulatie). Collins, Brown en Newman (1989) hebben dit uitgewerkt in een "cognitive apprenticeship model", waarin een aantal van de eerder genoemde docenttaken is opgenomen. Zij laten ook zien hoe de in dit model beschreven reeks van pedagogisch-didactisch handelen door een docent altijd is ingebed in een kennisdomein. Bovendien maken zij in hun model duidelijk dat in onderwijs dat is gericht op bevordering van de zelfregulatie van leerlingen, de docent en de leerlingen samen het onderwijsleerproces vormgeven en de leerlingen daarin een eigen aandeel hebben: vanuit de geleidelijk afnemende coaching moeten zij, via articulatie, reflectie en exploratie, hun vermogen tot zelfstandig denken en handelen in het betreffende domein ontwikkelen.
Ad c) Specifieke typen acties en reacties van moment tot moment tijdens interactiesituaties behoren ook tot de voor een docent gewenste vaardigheden, maar deze zijn te gedetailleerd voor het doel van het onderhavige onderzoek.

Fredriksen en White (1997), Mayer (2004a) en Kirschner, Sweller en Clark (2006) benadrukken dat de hedendaagse aandacht voor de gewenste actieve rol van de leerlingen zelf, niet betekent dat het er vooral om zou gaan dat leerlingen zichtbaar actief bezig zijn (met discussies, opdrachten, materialen, samenwerking, projecten etc.). Waar het om gaat is dat ze *cognitief* actief zijn, met begrijpen, redeneren, problemen oplossen. Dit geldt ook voor discussies en voor samenwerking: ook dan gaat het er om dat leerlingen hun inzichten, problemen, suggesties, argumenten en oplossingen expliciet verwoorden, uitwisselen en productief maken (ervan leren). De ondersteuning en begeleiding door de docent moet er vooral op zijn gericht deze cognitieve activiteit op gang te houden, te voeden en zo nodig bij te sturen. Dit vergt een voortdurende afweging van de juiste balans, dosering en timing.

Constructivistische ideeën en idealen betekenen volgens deze en andere auteurs ook niet dat docenten niet meer "gewoon" zouden kunnen lesgeven, uitleggen en instrueren. Dat belemmert namelijk geenszins de gewenste ontwikkeling van kennis bij de leerlingen. Ook hoeft het niet vooral te gaan over het gebruiken en toepassen van kennis, want het ontwikkelen van kennis op zich, van feiten, begrippen en principes, is ook een belangrijk leerdoel. Wel is het zaak de steun geleidelijk te verminderen, gelijk op met de groei in kennis (op het betreffende domein) en de mogelijkheden tot zelfsturing van de leerlingen (eveneens op het betreffende domein, mede door die kennis). Dit impliceert dat ook het volgen van het leerproces een belangrijke docenttaak is.
Elshout-Mohr, Van Hout-Wolters en Broekkamp (1999) laten gedetailleerd zien hoe het gewenste didactisch handelen van de docent afhankelijk is van het type leerdoel en type leertaak waar het om gaat (reproductief of productief, kennis of vaardigheid, cognitief of metacognitief, nabije transfer of verre transfer). Het in de praktijk brengen en effectief maken van dergelijk nauwkeurig didactisch handelen vergt vaak een leerproces bij zowel de leerlingen als de docent zelf.
Webb (2009), tot slot, wijst erop dat docenten met hun leerlingen soms een lange aanloop moeten nemen voordat een bepaalde soort leersituatie vruchtbaar kan worden. Voor het leerzaam doen zijn en productief maken van, in dit geval, samenwerking tussen leerlingen, moeten leerlingen eerst leren om een groep te vormen en daarin te overleggen en samen te werken, moeten taken beschikbaar zijn die zodanig zijn gestructureerd dat leerlingen deze doelgericht samenwerkend zelfstandig moeten en kunnen uitvoeren, en moeten docent en leerlingen leren hoe de docent tussentijds de interactie en de gedachteontwikkeling in de groepen kan ondersteunen en bevorderen. Hieruit kunnen we leren dat het vormgeven van een vernieuwend onderwijsconcept vraagt om continuïteit (in de relatie tussen docent en leerlingen) en cumulatieve ontwikkeling (via leerlijnen), en dus goed moet worden doordacht en over een langere periode planmatig en in onderlinge samenwerking moet worden aangepakt.

De dagelijkse interactie tussen docent en leerlingen

Veel van het onderzoek dat ten grondslag ligt aan de hiervoor beschreven literatuur is uitgevoerd bij specifiek uitgekozen zeer capabele docenten en/of in speciaal voor het onderzoek geconstrueerde situaties. Daardoor kunnen theoretische principes en modellen duidelijker zichtbaar worden en enigszins zuiver op hun verloop en effecten worden onderzocht.
Uit observaties in "echte" schoolklassen komt een dagelijkse praktijk naar voren die zeer veelvormig is en heel complex, en wel des te meer naarmate leerlingen daadwerkelijk eigen keuzes kunnen maken en zelfstandig en in samenwerking aan het werk en aan het leren zijn, zoals in nieuwe onderwijsconcepten wordt beoogd (Doyle, 1986; Glaser, 1990; Shuell, 1993, 1996; Shulman, 1986). De docent (of, bij team-teaching of teamcoaching: docenten) en leerlingen reageren op elkaar, en hun handelen

en gedrag staat daardoor in voortdurende wisselwerking. Er vinden allerlei transacties, uitwisselingen en onderhandelingen plaats, zowel op het sociale en relationele vlak als in termen van leerdoelen, leerstof, taken, feedback en beoordeling. De onderlinge interacties worden beïnvloed door denkbeelden, emoties, percepties en verwachtingen over en weer, en zijn daar op hun beurt weer op van invloed. Dit werkt door op de leerprocessen bij de leerlingen, bedoeld en onbedoeld, en ook bij de docent kan sprake zijn van leren. Waar docent en leerlingen gedurende langere tijd met elkaar te maken hebben, zal ook sprake zijn van ontwikkeling (van de betrokkenen, en van hun onderlinge relatie en interactie), waardoor het hele patroon van processen zelf ook nog eens kan verschuiven.
Gezien deze complexe dagelijkse realiteit kan men zich afvragen of onderzoek van die praktijk dan wel inzicht kan geven in wat wenselijk handelen van docenten is, en hoe inzichten daarin die voortkomen uit theorie en onderzoek in de praktijk vruchtbaar kunnen worden gemaakt.
Waar het handelen van docenten zo veelvormig is dat het moeilijk valt te typeren in termen van een duidelijk te omschrijven theorie of model over leren, didactiek en instructie, ligt het voor de hand dat het vooral persoonlijke eigenschappen van de docent zijn waarvan eventuele relaties met leeractiviteiten, –processen en –resultaten bij leerlingen zullen kunnen worden gelegd. Dat is ook wat er gebeurde, toen in de jaren ’60 en ’70 van de vorige eeuw onderzoek via observaties in schoolklassen in zwang kwam (zie Shuell, 1996). Daaruit kwam naar voren dat docenten die duidelijk, enthousiast, flexibel en taakgericht waren, hun leerlingen veel gelegenheid boden om zich de stof eigen te maken en zorgden voor een hoog cognitief niveau van interactie, gemiddeld de beste resultaten behaalden.
In een ander type onderzoek dat in dezelfde periode op gang kwam, het proces-product onderzoek, was men eveneens gericht op bepalende kenmerken van effectief onderwijs. In dit geval ging men echter uit van geconstrueerde modellen met clusters van factoren waarin men uitkomsten van reeds uitgevoerde onderzoeken probeerde te plaatsen om zo tot conclusies te komen. Ook dit type onderzoek heeft kenmerken van docenten opgeleverd (zie Scheerens, 2007). Naast gedragskenmerken die vergelijkbaar zijn met de hierboven genoemde, zoals betrouwbaarheid, inzet en flexibiliteit, zijn dit aspecten van kennis, vaardigheid en ambitie: beheersing van de leerstof, verbale vaardigheid, hoge verwachtingen, steeds streven naar verbetering.

Factoren voor effectief onderwijs

Uitspraken over gewenst docenthandelen kan men, in plaats van op theorieën over leren en leeromgevingen en onderzoek van de dagelijkse onderwijspraktijk, ook willen baseren op “harde” gegevens en “evidence”, en willen beschrijven in overzichten van “wat werkt”. Zulke uitkomsten worden meestal gebaseerd op het eerdere genoemde proces-product onderzoek. Dit is vaak theorie-arm en minder gericht op begrijpen en begrijpelijk maken *hoe* en *waarom* of *waardoor* iets werkt.

Studies in deze traditie zijn vaak gebaseerd op meta-analyses, statistische analyses van de (gemiddelde) uitkomsten van grote aantallen (eerder en door anderen uitgevoerde) onderzoeken. Deze werkwijze voegt aan het theorie-arme karakter nog een tweede nadeel toe, namelijk dat in die onderzoeken uiteenlopende variabelen zijn onderzocht die slechts in eenzelfde model zijn onder te brengen door dat wat enigszins op elkaar lijkt samen te nemen. De categorieën in zo'n model bevatten daardoor zulke uiteenlopende zaken dat het lastig is om er een duidelijke betekenis aan toe te kennen. Categorieën met labels als "Leeromgeving", "Onderwijsstrategie (gestructureerd, direct, beheersing, etc.)", "Onderwijsstrategie (constructivistisch)", "Ondersteuning" en "Feedback/controle/evaluatie/tests" zijn zo breed en vaag, dat ze vrijwel hun nuttige waarde verliezen.

Wie zulke studies bestudeert en vergelijkt (zie Campbell et al., 2004; Scheerens, 2007; Seidel & Shavelson, 2007; Shuell, 1996), constateert enerzijds dat de uitkomsten variëren (niet verwonderlijk als men zich baseert op verschillende selecties van uitgevoerde onderzoeken en verschillende modellen en categorieën hanteert om tot gemiddelde effecten te komen) maar ziet anderzijds ook dat bijna alles in principe een steentje kan bijdragen (in principe, want er zijn geen methoden die altijd werken, effecten zijn altijd meervoudig, en leerlingen kunnen uiteenlopend reageren, zie Shuell, 1996). Anders gezegd: resultaten van onderwijsleerprocessen kunnen door veel factoren worden beïnvloed en bevorderd. Schoolleiders, curriculumontwerpers en docenten kunnen dus inderdaad, zoals soms wel gebeurt, worden vergeleken met dirigenten: het gaat om het geheel, de balans, de afstemming, de wisselwerking. De boodschap kan dan zelfs worden: "het [goed samengestelde] geheel is meer dan de som der delen" (zie ook Bosker, Creemers & Stringfield, 1999). Dit wil overigens niet zeggen dat dit goed samengestelde geheel ook vooraf moet en kan worden ontworpen. Heel goede docenten laten vaak geen bepaald patroon zien (handelen niet volgens een zorgvuldig opgesteld model van handelen) (Shulman, 1987), maar spelen alert en flexibel in op het onderwijsleerproces zoals zich dat ontwikkelt (zie hiervoor: processen in schoolklassen zijn veelvorming en complex en voor een docent zijn daarbij vooral ook bepaalde persoonlijke eigenschappen van belang).

Een tweede en meer specifieke conclusie is dat het er vooral om gaat te bevorderen dat leerlingen veel, intensief en gevarieerd met de leerstof bezig zijn (cognitief, affectief, metacognitief). Dit is een samenvattende interpretatie van de gebleken effectiviteit van uiteenlopende factoren als: veilig leerklimaat, ordelijke omgeving, goed klassen-management, taakgerichtheid, structuur, duidelijkheid, beschikbare en bestede tijd, stellen van denkvragen, activerende werkvormen, betekenisvolle en uitdagende opdrachten, op leerinhoud (kennis, vaardigheid) gerichte sociale interactie (tussen docent en leerlingen, tussen leerlingen onderling), feedback. Anders gezegd: hoe dichter op de kern van het leerproces, hoe beter (zie Seidel & Shavelson, 2007). Hierbij horen wel twee nuanceringen (zie ook Campbell et al., 2004). De eerste is dat effecten van docenthandelen zullen verschillen al naar gelang de schoolcontext, de leerlingen en

het schoolvak. De tweede is dat docenten, zeker in deze tijd met diverse taken en hoge verwachtingen, niet steeds en op alle fronten even succesvol zullen kunnen zijn.

Standaarden voor opleiding en beoordeling van leraren

In het voorgaande is een aantal belangrijke taken, aspecten en onderdelen van gewenst docenthandelen naar voren gekomen. Zulke aanwijzingen voor het handelen van docenten hebben we niet uitgebreid teruggevonden in wetenschappelijke literatuur over standaarden voor de opleiding en beoordeling van leraren. Deze literatuur betreft vooral Amerikaanse publicaties en die gaan vooral over de gewenste kennisbasis van docenten. Die kennisbasis wordt wel breed opgevat, inclusief de zogenoemde "procedurele kennis". Vanuit de hiervoor beschreven complexiteit van de omgeving waarin docenten en leerlingen werken en leren en waarin docenten flexibel en adequaat moeten kunnen opereren, is het begrijpelijk dat wordt gepleit voor veel bagage (kennis) en een breed repertoire (vaardigheden).
In een veel geciteerde publicatie benoemde Shulman (1987) een aantal categorieën van kennis waarover docenten zouden moeten beschikken: vakinhoudelijke kennis, algemene pedagogische kennis over klassenmanagement en groepsprocessen, kennis over curriculum, leerdoelen, programma's en leermiddelen, kennis over leerlingen, en kennis over manieren waarop leerlingen vakinhoudelijke kennis en vaardigheden kunnen ontwikkelen, de typische problemen die daarbij kunnen optreden en hoe je daar als docent op kunt reageren (door Shulman "pedagogical content knowledge" genoemd, in Nederlands van oudsher bekend als "vakdidactische kennis").
Daarnaast constateerde Shulman (in 1987) dat er steeds meer "general principles of effective teaching" bekend worden. Volgens hem is de zwakte daarvan, namelijk dat die voorbijgaan aan het feit dat gewenst docenthandelen vaak sterk (vak)inhoud-specifiek is, tegelijk ook hun kracht, omdat algemene principes de toch al enorm complexe taak van een docent enigszins kunnen vereenvoudigen. Anderen echter (Grossman & Schoenfeld, 2005; Mayer, 2004a, 2004b) bepleiten juist indringend het belang van specifieke vakinhoudelijke en vakdidactische kennis en vaardigheden.
In een recente, eveneens gezaghebbende publicatie (Bransford, Derry, Berliner & Hammerness, 2005) wordt expliciet gesteld dat docenten tussen allerlei aspecten en gezichtspunten moeten kunnen pendelen en deze moeten kunnen combineren, en daarom een kader nodig hebben, een overzicht of totaalplaatje ('conceptual map'), en de vaardigheid om met behulp daarvan gebeurtenissen en processen te interpreteren en hun keuzes en handelen daarop af te stemmen. Het kader dat de auteurs hiertoe aanreiken (en dat overigens al werd voorgesteld door De Corte, 2000). bevat vier componenten: kennis over a) wat er moet worden geleerd, b) hoe leerlingen dit kunnen leren, c) welke leeromgeving dit leren kan bevorderen, en d) hoe kan worden vastgesteld dat en wat er is geleerd. Effectief onderwijs vergt dat deze vier componenten goed op elkaar worden afgestemd.

Tot slot is er de nodige literatuur over visies, ervaringen en praktijkkennis van leraren zelf, in relatie tot gewenst handelen volgens nieuwe onderwijsconcepten. Uit het vele dat hierover is geschreven, beperken we ons tot de volgende vaststellingen (zie Doyle, 1986; Fox, 1983; Larsson, 1983; Lyons, 1990; Pratte en Rury, 1991; Prawat, 1992; Windschitl, 2002).

Leraren hebben eigen opvattingen over onderwijs, leren en de rollen van en relatie tussen henzelf en de leerlingen. Zo kunnen leraren de vakinhoud als uitgangspunt nemen en als hun taak zien die zo goed mogelijk over te dragen, of vertrekken vanuit de leerlingen en hun leerproces en zich primair daarop richten. Zij kunnen hun vak zien als een vast omlijnd geheel of als een domein in ontwikkeling, en analoog de mogelijkheden en voorkeuren van de leerlingen als gegeven of als onderhevig aan verandering en vatbaar voor beïnvloeding. Zij kunnen zichzelf onmisbaar achten en in hun handelen vasthouden aan bepaalde patronen, of leerlingen veel ruimte geven en open staan voor nieuwe ervaringen, inzichten en aanpakken.

Opvattingen en handelen van leraren hangen vaak samen en voor verandering in handelen is vaak eerst of ook verandering in opvattingen nodig. Als leraren niet achter een voorstel tot vernieuwing kunnen staan, zal hun handelen waarschijnlijk niet wezenlijk veranderen. (Zo kan bijvoorbeeld een docent in een keuzewerktijduur, waarin de bedoeling is dat leerlingen zelf kiezen waaraan ze werken en daarbij op maat worden begeleid, van de leerlingen eisen dat ze aan 'zijn eigen vak' werken, of het hele uur gebruiken om les te geven; zie ook 8.3). Als leraren een nieuwe visie niet goed begrijpen, zullen ze wellicht wel hun handelen veranderen maar is de kans reëel dat de beoogde vernieuwing schipbreuk lijdt. (Zo kan een docent die heeft begrepen dat hij de leerlingen meer zelfstandig moet laten werken en leren, min of meer stoppen met lesgeven en de leerlingen alles zelf laten uitzoeken).

Als leraren hun onderwijs willen veranderen volgens een meer constructivistische visie zien zij zich voor lastige dilemma's geplaatst. Ze moeten steeds kiezen voor een bepaalde balans, tussen individueel leren en samenwerkend leren, tussen aansluiten bij kennisontwikkeling zoals die zich bij leerlingen voordoet en het nemen van de kennis van experts als norm, tussen vooraf uitleggen en instrueren en leerlingen zelf laten ontdekken, tussen diepgaande verwerking van geselecteerde stof en de breedte die het programma vereist, en tussen opvattingen en verwachtingen van hemzelf, de schoolleiding, de leerlingen, de ouders en de toezichthouders.

Ook leerlingen hebben opvattingen over onderwijs, leren, en de rollen van de docent en henzelf. Leraren zullen hun onderwijspraktijk alleen samen met de leerlingen kunnen veranderen. Veel leerlingen hebben een voorkeur voor een meer traditionele rolverdeling (de docent geeft klassikaal en gestructureerd les, de leerlingen luisteren en werken daarna aan gesloten opdrachten) en als een leraar daarin verandering wil aanbrengen (de leerlingen werken zelfstandig aan open opdrachten), zullen leraar en leerlingen beiden hun invloed gaan aanwenden om de ander de gewenste rol te laten

nemen. Het omgekeerde is ook mogelijk: leerlingen die meer ruimte willen voor eigen inbreng en samenwerken en leraren die aan de bestaande praktijk willen vasthouden (omdat ze vrezen hun grip op de gang van zaken te verliezen), of de verandering niet aandurven (omdat ze denken dat de leerlingen het niet aankunnen). Onderwijs wordt door leraar en leerlingen samen gemaakt.
Naarmate de keuzeruimte en zelfstandigheid van de leerlingen toenemen, neemt het belang van het traditionele ambacht van de leraar op het vlak van vakkennis, pedagogische kwaliteiten en klassenmanagement eerder toe dan af. In een meer constructivistische leeromgeving moet de leraar (nog) meer met de leerlingen in gesprek en (nog) meer inspelen op het verloop van leeractiviteiten en leerprocessen. Op zijn ambachtelijke professionaliteit zal een (nog) groter beroep worden gedaan om intensief en flexibel bruggen te bouwen tussen zijn kennis van en inzicht in de belangrijke inhouden en vaardigheden in zijn vakgebied, zijn waarneming en taxatie van wat en hoe leerlingen denken (begrippen, redeneringen, misconcepties, lacunes), en de processen die zich in de leeromgeving tijdens het werken in de groep en bij individuele leerlingen afspelen.

7.3.2 Gewenst docenthandelen in beleidsgerichte publicaties

Bekwaamheden en bekwaamheidseisen
In beleidsgerichte publicaties over gewenst docenthandelen, in termen van bekwaamheden of bekwaamheidseisen, wordt vaak onderscheid gemaakt tussen enerzijds vakkennis (vakinhoudelijk en vakdidactisch), anderzijds pedagogisch-didactische, interpersoonlijke en organisatorische kwaliteiten en handelingsrepertoire (Van Gennip en Vrieze, 2008; Ministerie van OC&W, 1993; Onderwijsraad, 2005a). In meer concrete uitwerkingen in termen van competenties worden meestal nog twee andere competenties toegevoegd: bekwaamheden in samenwerken met anderen (in en buiten de school), en werken aan voortgaande persoonlijke ontwikkeling, met het oog op onderwijsverbetering en in de context van schoolontwikkeling (Stuurgroep Beroepskwaliteit Leraarschap, 1997; Stichting Beroepskwaliteit Leraren, 2004).
Met enige regelmaat is geschreven over te stellen (minimum) bekwaamheidseisen, die zouden moeten kunnen worden getoetst aan duidelijke normen (Onderwijsraad, 2005a; Stuurgroep Beroepskwaliteit Leraarschap, 1997). Gaandeweg is dit streven echter genuanceerd. Zo vroeg de Onderwijsraad in 2005 om aandacht voor het feit dat bekwaamheidseisen onderling samenhangen en niet allemaal even belangrijk zullen zijn waar het gaat om het bevorderen van het leren van leerlingen. De Commissie Leraren (2007), die eveneens stelde dat een goede leraar primair zorgt dat zijn leerlingen zoveel mogelijk leren, ziet de ontwikkelde SBL-bekwaamheden als basis maar wijst erop dat verdere ontwikkeling onontbeerlijk is, en dat leraren daartoe moeten beschikken over onderzoekende vaardigheden en een gerichtheid op werken op basis van dat wat effectief blijkt.

In beleidsgerichte publicaties gaat de aandacht vaak uit naar wat kan worden opgemerkt over gewenst docenthandelen zonder te verbijzonderen naar schoolvak of onderwijssector. Het (algemeen) pedagogisch-didactisch handelen komt daarmee centraal te staan. Daarnaast wordt ook wel aandacht gevraagd voor de persoon van de leraar (Commissie Leraren, 2007; Van Gennip en Vrieze, 2008), waarbij persoonskenmerken (zoals passie, betrokkenheid, humor) echter vooral worden geacht tot uitdrukking te komen in het pedagogisch handelen.
De genoemde publicaties benoemen amper specifieke aspecten of onderdelen van gewenst docenthandelen. De enige meer concrete uitwerking is te vinden in de SBL-competenties (2004). Het beeld dat uit het voorgaande naar voren komt is dat een goede leraar beschikt over een breed repertoire van handelingsmogelijkheden waaruit deze van moment tot moment en rekening houdend met de context een adequate keuze maakt. Voor nadere richtlijnen wordt vooral verwezen naar het belang van kennisnemen en gebruiken van resultaten uit onderzoek.

Onderzoek dat evidence-based resultaten beoogt op te leveren
Gravemeijer en Kirschner (2007) maken duidelijk dat de roep om verbetering en verantwoording niet kan steunen op zogenoemde harde bewijzen uit onderzoek. De daarvoor gewenste echte experimenten zijn volgens hen namelijk amper uitvoerbaar. Bovendien is zulk onderzoek gericht op generaliserende conclusies (terwijl tegenwoordig juist wordt gevraagd rekening te houden met verschillen tussen leerlingen en omstandigheden, en maatwerk te bieden), laat het voornamelijk zien wat werkt en niet hoe het werkt (terwijl docenten juist behoefte hebben aan inzichten waarop zij hun keuzes voor bepaalde aanpakken kunnen baseren), en werkt het vertragend (want gegronde kennis ontwikkelen duurt lang) en conserverend (omdat het zich richt op wat goed uitvoerbaar en goed meetbaar is).
Momenteel komt er steeds meer aandacht voor onderzoek in de school en door docenten zelf. Docenten moeten daartoe dan wel worden geëquipeerd, en dat is opgenomen in de kwaliteitsagenda voor het opleiden van leraren (OC&W, 2008b). Daarin wordt gepleit voor het duidelijk vastleggen van het eindniveau van de lerarenopleidingen, in termen van een gezamenlijke kennisbasis, eindtermen en examens. De opleidingen dienen dit te doen in samenwerking met de beroepsgroep en het afnemende scholenveld. Onderdeel van het te behalen eindniveau zou moeten zijn dat leraren opbrengstgericht en evidence-based kunnen en willen werken: ze moeten leeropbrengsten kunnen analyseren en daarover in gesprek kunnen gaan, en conclusies kunnen trekken voor bijstelling van het onderwijsproces.
Deze aandacht voor onderzoek door docenten zelf wil niet zeggen dat het niet belangrijk blijft om kennis te nemen van onderzoek dat door onderzoekers is gedaan. Een aantal vernieuwende scholen heeft hun leeromgeving en aanpak gericht op onderwijsvormen die in ons land bekend zijn geworden onder de noemer “nieuw leren”. De laatste jaren zijn op verzoek van het Ministerie van OC&W enkele daarop gerichte studies uitgevoerd. We vatten deze kort samen.

Teurlings, Van Wolput en Vermeulen (2006) stellen vast dat uitwerkingen van nieuw leren vooral betrekking hebben op drie aspecten: een activerende leeromgeving met accent op zelfstandig leren, werken met betekenisvolle contexten en opdrachten en integratie van vakken, en samenwerking tussen leerlingen. Ten aanzien van actief en zelfstandig leren wijst het beschikbare onderzoek volgens hen op positieve effecten op motivatie, leergedrag en leerresultaten, maar het zelfstandig leren moet wel geleidelijk worden opgebouwd, met de docent als begeleider en coach. Wat betreft het werken met betekenisvolle contexten en integratie van vakken, laat onderzoek meer betrokkenheid, motivatie en begrip bij leerlingen zien, maar ook hierbij is goede begeleiding door de docent onontbeerlijk. Van samenwerking tussen leerlingen kunnen volgens onderzoek effecten worden verwacht op motivatie, sociale vaardigheden en manier van verwerken van de leerstof en transfer, maar de docent moet deze samenwerking wel aanleren en begeleiden. De algemene conclusie is dat de genoemde effecten pas te verwachten zijn op termijn, en ook alleen als de nieuwe manier van werken en leren consequent is ingevoerd. In het algemeen hangt het te verwachten effect van dergelijke nieuwe onderwijsvormen af van de kwaliteit van de begeleiding door de docent, en dus zijn of haar pedagogisch-didactisch handelen. Het is daarbij ook van belang vanuit welke opvatting over leren de docent handelt, en hoe coherent de school de vernieuwing uitwerkt.

Oostdam, Peetsma en Blok (2007) constateren dat de doelen van nieuw leren niet heel duidelijk zijn, en dat er ook nog weinig onderzoek naar is gedaan. Ook zij verwijzen naar de drie hiervoor genoemde aspecten, en zij zien vooral evidentie voor de effectiviteit van instructie die mede is gericht op zelfregulatie en van begeleiding die inspeelt op verschillen tussen de leerlingen (het eerste aspect), en veel minder duidelijk voor beide overige aspecten (authentiek leren en samenwerkend leren). Naar hun mening biedt de wetenschap op dit moment voornamelijk een constellatie van inzichten in leerprocessen en leerarrangementen. Die inzichten wijzen in de richting van de wenselijkheid van een balans tussen een instructiegerichte en een leerlinggerichte aanpak. De specifieke omschrijvingen van gewenst docenthandelen die de auteurs noemen, betreffen voordoen, volgen en op maat ondersteunen van gewenste denkprocessen (*modeling, monitoring, feedback, scaffolding*), en het geven van betekenisvolle taken waarbij leerlingen kunnen samenwerken.

Ontwikkelingen, problemen en vernieuwingsprocessen

Een aantal veranderende omstandigheden die scholen nopen tot vernieuwing is expliciet benoemd in een advies van de Onderwijsraad (1998), die stelde dat het werk van docenten door een reeks van ontwikkelingen zwaarder is geworden: een meer divers wordende samenleving, meer individualisering, mondiger leerlingen en ouders, verbreding van het onderwijs- en vormingsaanbod, introductie van eindtermen en kerndoelen, deregulering en vergroting van autonomie van scholen, schaalvergroting en concurrentie tussen scholen, de opkomst van ICT, en didactische veranderingen.

Tegelijk is de druk om de kwaliteit van het onderwijs te verbeteren en daarover verantwoording af te leggen, toegenomen. De Onderwijsraad (2007) stelt echter vast dat de kwaliteit van het primaire proces, het pedagogisch-didactisch handelen van docenten, moeilijk valt te borgen en verantwoorden, want het is lastig om op dat handelen zicht te krijgen en te bepalen wat het aandeel en de verantwoordelijkheid van individuele docenten is voor de onderwijskwaliteit in de school als geheel.
Schoolleiders en docenten handelen in een context waarin onder meer sprake is van motivatieproblemen bij leerlingen en toenemende verschillen tussen leerlingen. Als gevolg daarvan ontstaat behoefte aan de ontwikkeling van nieuwe leerarrangementen die leerlingen meer aanspreken en zijn veel scholen bezig met processen van vernieuwing (Busman, Klein en Oomen, 2006; Klein et al., 2008; Ministerie van OC&W, 2004, 2008a; Oostdam et al., 2007). De scholen bepalen zelf hoe het onderwijs wordt vormgegeven en welke pedagogisch-didactische aanpak wordt gehanteerd, en scholen ondernemen dan ook allerlei initiatieven en laten een grote variëteit zien aan nieuwe vormen van leren (Inspectie van het Onderwijs, 2006; Onderwijsraad, 2005b). Op bijna alle scholen neemt de schoolleiding het initiatief tot innovatie, maar ook docenten doen dit. Schoolleiders, teamleiders en docenten kunnen daarbij in hun perceptie van de ontwikkelingen in hun school sterk verschillen (Ministerie van OC&W, 2008a).
Een beperkte groep scholen (schattingen variëren van 5% tot 15% van alle scholen voor VO) voert echt ingrijpende vernieuwingen door, waarin leerlingen veel zelfstandig werken en leren in grotere ruimten en worden begeleid door meer docenten (met aanpassingen in gebouw en rooster). Gemeenschappelijke ontwikkelingen op veel scholen betreffen een toenemende aandacht voor activerende didactiek en een interactieve leeromgeving, zelfstandig werken en leren, samenwerking tussen leerlingen, verschillen tussen leerlingen, en hun persoonlijke ontwikkeling (Hofman, Hofman, Dijkstra, De Boon & Meeuwisse, 2007; Ministerie van OC&W, 2008a; Oostdam et al., 2007). Belangrijke aandachtspunten zijn momenteel het vinden van een goede balans tussen docentgestuurde en leerlinggestuurde aanpakken en het signaleren van en begeleiden bij individuele problemen.

Bronnen met wettelijke status: SBL en Inspectie
Hiervoor bleek dat beleidsgerichte publicaties slechts in beperkte mate aanwijzingen bevatten voor gewenst docenthandelen. Deze blijven beperkt tot enkele algemene suggesties voor de gewenste leeromgeving (aanbieden van betekenisvolle taken en van de mogelijkheid tot samenwerken), en incidentele verwijzingen naar elementen uit het cognitive apprenticeship model (modeling, feedback, scaffolding) die liggen op het snijvlak van algemeen pedagogisch-didactisch handelen en vakspecifieke instructie-strategieën. In de meeste bestudeerde documenten staat het pedagogisch-didactisch handelen van docenten centraal, maar wordt dit niet verder geconcretiseerd.

Het “pedagogisch-didactisch handelen” kan het beste worden benaderd in termen van taken van docenten. Dit past ook bij de enige meer uitgewerkte beschrijving van gewenst docenthandelen, namelijk zoals geformuleerd in de SBL-competenties (2004).
Daarnaast is er het Toezichtkader voor het voortgezet onderwijs van de Inspectie van het onderwijs (2009). Dit onderscheidt negen kwaliteitsaspecten, geclusterd in vier rubrieken: opbrengsten, onderwijsleerproces, zorg en begeleiding, en kwaliteitszorg, elk verdeeld in subaspecten (in totaal 41). Daarin wordt vooral de school als organisatie aangesproken. Negen subaspecten verwijzen naar gewenst docenthandelen: zorgen dat de leerlingen op een respectvolle manier met elkaar en anderen omgaan; stimuleren van het welbevinden en de motivatie van de leerlingen; geven van duidelijke uitleg van de leerstof; realiseren van een taakgerichte werksfeer; leerlingen actief betrekken bij de onderwijsactiviteiten; de instructie afstemmen op verschillen in ontwikkeling tussen leerlingen; de verwerkingsopdrachten afstemmen (idem); de onderwijstijd afstemmen (idem); systematisch volgen en analyseren van de voortgang in de ontwikkeling van de leerlingen. Deze aspecten passen bij de hiervoor samengevatte kennis uit onderzoek, maar vormen daaruit wel een selectie. Het toezichtkader bestrijkt daarmee slechts een deel van het pedagogisch-didactisch handelen van docenten.
Gelet op de schaarste aan meer systematische indelingen van het pedagogisch-didactisch handelen van docenten is in onderhavig onderzoek gebruik gemaakt van een in eerder onderzoek ontwikkelde en gevalideerde vragenlijst waarmee dit handelen in brede zin in kaart kan worden gebracht (Sol & Stokking, 2008b; zie hoofdstuk 3 van dit proefschrift). De daarin onderscheiden (22) taken van docenten zijn voorgelegd aan de schoolleiders, teamleiders, docenten en externe stakeholders op de scholen waar onderhavig onderzoek is uitgevoerd. Hiervan wordt verslag gedaan in par. 7.3.4.
De aanwijzingen voor gewenst algemeen pedagogisch-didactisch docenthandelen die kunnen worden gedestilleerd uit de SBL-competenties en het Toezichtkader van de Inspectie kunnen worden geplaatst naast de 22 taken uit het genoemde eerdere onderzoek. Het resultaat daarvan staat in figuur 7.2.

Indicatoren SBL (2004) *[(x) = nummer competentie]*	*Toezichtkader Inspectie (2009)*	*Taken in Vragenlijst PDH (2008)*
(1) Schept klimaat voor samenwerking, (1) Maakt leerlingen medeverantwoordelijk (2) Biedt veilige leer- en werkomgeving (2) Bespreekt de sfeer en de omgang (4) Zorgt voor taakgerichte omgeving	4.7 Respectvolle omgang 4.8 Stimuleren welbevinden 5.2 Taakgericht werken	1 Zorgen voor een goed leerklimaat
(2) (3) Bevordert de zelfstandigheid van leerlingen		2 Bijdragen aan persoonlijke vorming
	5.1 Geven van duidelijke uitleg	3 Geven van les, uitleg en instructie
(3) Brengt leerlingen in contact met toepassingen in beroep en samenleving		4 Leerlingen het nut van de leerstof laten zien
		5 Leerlingen wijzen op verbanden tussen vakken
		6 Leerlingen laten werken aan opdrachten
(2) Is nieuwsgierig naar ideeën van leerlingen (3) Motiveert en daagt uit (3) Zorgt voor actieve betrokkenheid	5.3 Leerlingen actief erbij betrekken	7 Leerlingen stimuleren en activeren
		8 Leerlingen vaardigheden aanleren
(2) (3) Houdt rekening met verschillen tussen leerlingen	6.2 Instructie afstemmen op verschillen	9 Inspelen op verschillen tussen leerlingen
(3) Biedt gevarieerde leeractiviteiten aan (3) Biedt keuzemogelijkheden	6.3 Opdrachten afstemmen op verschillen 6.4 Onderwijstijd idem	10 Inspelen op verschillen door middel van opdrachten
(1) Schept klimaat voor samenwerking (1) Stimuleert samenwerken (2) Leert leerlingen samenwerken		11 Leerlingen laten samenwerken en hen daarbij ondersteunen
(3) Biedt keuzemogelijkheden aan (4) Ook wat betreft de planning		12 Leerlingen eigen keuzes laten maken
(2) Zorgt dat leerlingen zelf keuzes kunnen maken		13 Leerlingen ondersteunen in het maken van keuzes
(2) Laat leerlingen zelfstandig werken		14 Leerlingen eigen leerproces laten sturen
(1) Biedt kader en hulp bij het eigen leerproces (2) Stimuleert reflectie op eigen leerproces (3) Begeleidt leerlingen in hun leerproces (3) Leert leerlingen leren en werken		15 Leerlingen ondersteunen bij het sturen van hun eigen leerproces
(2) (3) Signaleert problemen bij leerlingen (3) Vormt zich een beeld van wat leerlingen kunnen (3) Weet hoe leerlingen zich ontwikkelen en welke problemen zich (kunnen) voordoen		16 Volgen van het leerproces
(3) Geeft opbouwend commentaar op het werk en werken van leerlingen		17 Geven van feedback
		18 Zorgen dat feedback wordt gebruikt
(3) Doet op basis van evaluatie en toetsing betrouwbare uitspraken over kennis en vaardigheid		19 Beoordelen van het werk en de prestaties van leerlingen
(2) Bespreekt met collega's de sfeer in de groepen en problemen bij individuele leerlingen (4) Organiseert samenwerking met collega's (5) Werkt goed samen met collega's (7) Gebruikt evaluatie, reflectie en feedback om zich verder te ontwikkelen		20 Open staan voor leren en samenwerken
		21 Samen met collega's een groep leerlingen begeleiden
(7) Formuleert leervragen, werkt daar aan (7) Benut kansen om zichzelf verder te ontwikkelen zoals vakbladen. trainingen, projecten, onderzoek		22 Werken aan de eigen professionele ontwikkeling

Figuur 7.2 Aanwijzingen voor pedagogisch-didactisch docenthandelen vanuit de SBL-competenties, het Toezichtkader van de Inspectie, en de 22 taken uit de Vragenlijst PDH

7.3.3 Gewenst docenthandelen in de onderwijsconcepten van de scholen

De resultaten van de analyses van de onderwijsconcepten van de vier scholen die aan dit onderzoek meewerkten, worden hieronder samengevat per onderscheiden categorie (zie 7.2).

Zorgen voor een goed leerklimaat en bijdragen aan persoonlijke vorming
In de leeromgeving die op de scholen wordt nagestreefd, staan aspecten centraal die te maken hebben met veiligheid, orde en het aanbieden van structuur en met het hanteren van duidelijke regels en grenzen. Op alle vier scholen vindt men het belangrijk leerlingen vertrouwen te geven en op drie van de vier scholen wordt het aangaan en opbouwen van een relatie benadrukt. Men streeft ernaar leerlingen gelijkwaardig te behandelen, met daarbij soms nog expliciet een nadruk op openheid, wederzijds respect en elkaar serieus nemen. Op twee scholen wordt een aantal gewenste persoonskenmerken van medewerkers benoemd: open, flexibel, maar ook betrouwbaar, duidelijk en streng. De scholen benoemen verschillende pedagogische doelen voor de leerlingen: stimuleren van zelfstandigheid en eigen verantwoordelijkheid (meer in het algemeen (2x) of toegespitst op het leren (2x)); stimuleren van eigen initiatief (2x); leren omgaan met vrijheid en grenzen (2x) en de nadruk leggen op het bijbrengen van normen en waarden (3x), gemeenschapszin (2x) en leren als sociaal proces (1x). Op drie scholen benadrukt men het oog hebben voor de eigenheid van leerlingen.

Het aanbod in het curriculum
Op drie scholen wordt gespecificeerd dat herkenbare, contextrijke leerstof wordt aangeboden, in samenhang en doorlopende leerlijnen. Tevens worden vakoverstijgende vaardigheden aangeboden. Bij één school gaat het vooral om het aanbieden van extra opdrachten.

Instructie geven, werkvormen en opdrachten aanbieden
In de onderwijsconcepten staan vooral wensen met betrekking tot activerende, inspirerende instructie, waarbij de docent op meer manieren kan uitleggen (2x), herkenbare voorbeelden kan aandragen (1x) en een goed overzicht heeft over de leerstof en de leerlijn (2x). Ook wordt variatie belangrijk gevonden, tot uitdrukking komend in het aanbieden van verschillende werkvormen (2x) en gevarieerde didactiek (1x). Naast individuele opdrachten moeten ook veel samenwerkingsopdrachten worden aangeboden (4x). Leerlingen moeten de mogelijkheid krijgen bij het uitvoeren van opdrachten keuzes te maken (4x) en succeservaringen op kunnen doen (1x).

Activeren, stimuleren van leeractiviteiten en/of zelfsturing
Op alle scholen is het de bedoeling dat docenten leerlingen motiveren, activeren, stimuleren en uitdagen. Zij kunnen dat doen door te luisteren, vragen te stellen, eisen te stellen en verwachtingen te expliciteren (2x). Zij moeten de leerlingen zelfstandig laten

werken, plannen, samenwerken, reflecteren (3x), elkaar laten begeleiden (1x) en feedback laten geven (2x) en ze moeten leerlingen stimuleren zelf meer verantwoording te nemen voor hun eigen leerproces. Op één school is een accent zichtbaar op het benutten van inzichten uit de cognitieve psychologie, gericht op de ontwikkeling van kennis en van hogere orde cognitieve vaardigheden.

Begeleiden van het leerproces en/of de zelfstandigheid en zelfsturing; coachen
Op alle scholen wordt het belang benadrukt van het begeleiden van de ontwikkeling van kennis en vaardigheden (zelfstandig werken, samenwerken, keuzes maken, feedback geven, reflecteren). Op twee scholen wordt aangegeven dat docenten de sturing moeten "durven loslaten." Op een van deze twee scholen legt men ook nadruk op het aan leerlingen geven van inzicht in de ontwikkeling van eigen talenten en op het stimuleren van sociaal bewustzijn.

Inspelen op verschillen
Op alle vier scholen wil men maatwerk leveren, toegespitst op niveau (4x), leerstijl (3x), typen intelligentie (2x), wensen (2x), talenten (1x) en zorgbehoeften (1x) van leerlingen. Docenten moeten daartoe differentiëren in aanbod en aanpak en de leerlingen keuzes bieden; ze moeten aansluiten bij het niveau van leerlingen en eruit halen wat erin zit (3x).

Volgen, feedback geven en beoordelen
Op alle vier de scholen wordt het volgen, signaleren, feedback geven en in gesprek gaan over resultaten, leerproces en/of gedrag benadrukt. Op drie scholen worden ook nog andere aspecten onder de aandacht gebracht: communiceren van de beoordelingscriteria en bespreken van toetsen; hanteren van gelijke normen; zorgen voor de aansluiting tussen leerstof en toetsen (elk op één school).

Reflecteren en leren en samenwerken met collega's
In deze laatste categorie worden belangrijk geacht: een pro-actieve houding aannemen (2x), experimenteren (1x), elkaar informeren (1x), elkaar aanspreken (1x), onderling afstemmen (2x), de dialoog aangaan (3x), samenwerken (2x), feedback ontvangen en benutten (2x), reflecteren op het eigen handelen (2x), blijven leren en zich blijven ontwikkelen (1x), het uiterste uit jezelf halen (1x), begeleiden van collega's (1x).

Al met al blijkt dat in de onderwijsconcepten van de scholen en bij de geïnterviewde schoolleiders, teamleiders, docenten, leerlingen en stakeholders een aantal docenttaken centraal staan. Belangrijk is dat docenten zorgen voor een goed leerklimaat en bijdragen aan persoonlijke vorming van leerlingen. Zij moeten inspirerende instructie kunnen geven, die aansluit bij het niveau van verschillende (groepen) leerlingen. Ze moeten leerlingen activeren en stimuleren en daarbij inspelen op verschillen tussen leerlingen en leerlingen keuzes bieden. Zij moeten het leerproces van leerlingen volgen

en (liefst zoveel mogelijk op maat) leerlingen begeleiden en coachen bij het zelfstandig werken en het samenwerken. Zij moeten leerlingen feedback geven en de dialoog aangaan met leerlingen. Ze moeten zelf bereid zijn tot reflecteren en tot het ontvangen van feedback. Ze moeten zichzelf blijven ontwikkelen en samenwerken.

7.3.4 Gewenst docenthandelen volgens de betrokkenen in en rond de scholen

Eerste ronde: schoolleiders, teamleiders en docenten

De werving door de schoolleiders van deelnemers aan deze eerste ronde resulteerde in een totaal aantal van 47 respondenten. Tabel 7.1 toont de verdeling per school en per functiecategorie.

Tabel 7.1: Aantal respondenten per school per functiecategorie

School	*Schoolleider*	*Teamleider*	*Docent*	*Totaal*
A	1	9	5	15
B	3	4	2	9
C	2	5	4	11
D	1	7	4	12
Totaal	7	25	15	47

De keuze van drie tot zeven belangrijke taken van docenten
De deelnemers is gevraagd drie tot zeven taken te kiezen uit een totaal van 22 taken (aangeboden op gekleurde kaartjes), namelijk "taken waarvan u het vooral belangrijk vindt dat docenten dit doen" (de formulering bij de schoolleiders en teamleiders). Tabel 7.2 laat de percentages zien waarmee de verschillende aantallen taken zijn gekozen. De meeste respondenten hebben zeven taken gekozen (het gestelde maximum). Het minimum van drie taken werd geen enkele keer gekozen.

Tabel 7.2: Percentages waarmee de verschillende aantallen taken zijn gekozen (n=47)

School	*7 taken*	*6 taken*	*5 taken*	*4 taken*
A	73	-	20	7
B	44	22	11	22
C	73	27	-	-
D	66	16	8	8
Totaal	66	15	11	9

Tabel 7.3 biedt een overzicht van de 22 taken die zijn aangeboden en per taak het percentage van de deelnemers dat voor die taak heeft gekozen, per functie, per school, en in totaal.

Over alle respondenten samen zijn de volgende taken het meest gekozen: taak 1 (zorgen voor een goed leerklimaat: 77%), taak 7 (leerlingen stimuleren en activeren: 72 %) en taak 3 (geven van les, uitleg en instructie: 53 %). Vervolgens zijn het meest gekozen: taak 17 (het geven van feedback), taak 15 (leerlingen ondersteunen bij het sturen van het leerproces), taak 22 (professionele ontwikkeling) en taak 2 (bijdragen aan persoonlijke ontwikkeling).

Er zijn verschillen tussen de scholen. Op de scholen A en C werd het geven van feedback aan leerlingen (taak 17) vaker gekozen dan op de andere scholen. Op school D werd vaker gekozen voor taak 13 (leerlingen ondersteunen bij het zelf maken van keuzes) dan op de andere scholen. Op school B werd minder vaak gekozen voor taak 1 (zorgen voor een goed leerklimaat) en taak 3 (geven van les, uitleg en instructie) dan op de andere scholen. Op de scholen C en D is minder vaak gekozen voor taak 7 (leerlingen stimuleren en activeren) dan op de scholen A en B, maar juist vaker voor taak 11 (leerlingen laten samenwerken en hen daarbij ondersteunen). Op de scholen B en C is minder vaak gekozen voor taak 8 (leerlingen vaardigheden aanbieden) en taak 22 (professionele ontwikkeling) dan op de andere scholen. Taak 15 (leerlingen ondersteunen bij het sturen van het leerproces) is op school A minder vaak gekozen dan op de andere drie scholen.

Bij drie taken waren de verschillen tussen de scholen significant ($p<.05$): taak 1: zorgen voor een goed leerklimaat ($F(3,43)=4,88$); taak 14: leerlingen hun eigen leerproces laten sturen ($F(3,43)=3,54$); taak 19: beoordelen van het werk en de prestaties van leerlingen ($F(3,43)=5,96$)

Tabel 7.3 bevat ook de splitsing naar functie: schoolleiders, teamleiders en docenten. Alle drie groepen hebben in hoge mate gekozen voor taak 1 (zorgen voor een goed leerklimaat), taak 3 (geven van les, uitleg en instructie) en taak 17 (geven van feedback aan leerlingen). Drie andere taken worden door zowel teamleiders als docenten vaker gekozen dan door schoolleiders: taak 7 (leerlingen stimuleren en activeren), taak 8 (leerlingen vaardigheden aanleren) en taak 15 (leerlingen ondersteunen bij het sturen van hun leerproces). Taak 22 (professionele ontwikkeling) wordt vaker door schoolleiders en teamleiders gekozen dan door docenten.

De verschillen tussen de functiegroepen in de mate waarin taken zijn gekozen, zijn voor drie taken statistisch significant ($p<.05$), namelijk taak 2: bijdragen aan persoonlijke vorming van leerlingen ($F(2,44)=1,76$); taak 7: leerlingen stimuleren en activeren ($F(2,44)=5,95$); en taak 20: openstaan voor leren en samenwerken ($F(2,44)=3,51$). Taak 2 is het meest gekozen door de schoolleiders en het minst door de teamleiders. Taak 7 is het meest gekozen door de docenten en het minst door de teamleiders. Taak 20 is door de schoolleiders vaker gekozen dan door de teamleiders en docenten.

Tabel 7.3: Percentages waarmee de taken zijn gekozen, per functie, per school, en totaal (de per kolom meest gekozen taken zijn vet gemarkeerd)

Taken van docenten	Functie***			School				tot
	SL	*TL*	*DO*	*A*	*B*	*C*	*D*	
1. Zorgen voor een goed leerklimaat **	**71**	**68**	**93**	**87**	**33**	**82**	**92**	**77**
2. Bijdragen aan de persoonlijke vorming van leerlingen *	**71**	16	**53**	**40**	**44**	**36**	25	36
3. Geven van les, uitleg en instructie	**43**	**52**	**60**	**53**	**33**	**55**	**67**	**53**
4. Leerlingen het nut van de leerstof laten zien	14	16	20	13	22	27	8	17
5. Leerlingen wijzen op verbanden tussen vakken	-	8	7	13	-	-	8	6
6. Leerlingen leren werken aan opdrachten	14	-	-	6	-	-	-	2
7. Leerlingen stimuleren en activeren *	29	**72**	**93**	**93**	**78**	**55**	**58**	**72**
8. Leerlingen vaardigheden aanleren	14	**36**	**53**	**53**	22	**27**	**42**	**38**
9. Inspelen op verschillen tussen leerlingen	29	24	7	13	11	27	25	19
10. Inspelen op verschillen door middel van opdrachten	14	8	13	27	11	-	-	11
11. Leerlingen laten samenwerken en hen daarbij ondersteunen	**57**	32	40	20	22	**64**	**50**	38
12. Leerlingen eigen keuzes laten maken	14	16	7	6	33	18	-	13
13. Leerlingen ondersteunen bij het zelf maken van keuzes	29	24	33	27	22	18	**42**	28
14. Leerlingen hun eigen leerproces laten sturen **	14	8	7	6	**33**	-	-	9
15. Leerlingen ondersteunen bij het sturen van hun leerproces	29	**48**	**47**	20	**56**	**64**	**50**	**45**
16. Volgen wat leerlingen leren en hoe ze zich ontwikkelen	29	24	7	13	**33**	**36**	-	19
17. Geven van feedback aan leerlingen	**43**	**44**	**53**	**67**	22	**55**	33	**47**
18. Bevorderen dat feedback wordt gebruikt	29	24	13	6	22	**36**	25	21
19. Beoordelen van het werk en de prestaties van leerlingen **	14	20	20	-	**44**	-	42	19
20. Open staan voor leren en samenwerken *	**43**	24	-	20	11	27	17	19
21. Samen met andere leraren een groep leerlingen begeleiden	-	8	13	-	11	18	8	9
22. Professionele ontwikkeling	**43**	**52**	20	**53**	22	27	**50**	**40**

** significant verschillend tussen functiecategorieën (p<.05)*
*** significant verschillend tussen de scholen (p<.05)*
**** SL=Schoolleider, TL=Teamleider, DO=Docent*

In de vragenlijst met open vragen die tijdens het groepsinterview na het maken van de keuze voor bepaalde taken is ingevuld, werd daarna gevraagd wat een goede docent bij de gekozen taken zou moeten doen. Hierop zijn heel diverse antwoorden gegeven, die vaak op meerdere taken tegelijk betrekking hadden. Het duidelijkst gerelateerd aan een bepaalde taak zijn de antwoorden die betrekking hebben op taak 1 (zorgen voor een goed leerklimaat) en taak 3 (geven van les, uitleg en instructie). Bij taak 1 geeft men aan dat een goede docent rustig, duidelijk en voorspelbaar moet zijn. Hij of zij moet zorgen voor veiligheid en moet bijdragen aan het ontwikkelen van zelfvertrouwen bij de leerling(en). Hij of zij moet oog hebben voor de relatie met de leerling en de relaties tussen leerlingen onderling. Bij taak 3 geeft men aan dat een goede docent enthousiast moet kunnen uitleggen en daarbij activerend, boeiend en geduldig moet zijn. Hij of zij moet de lessen goed voorbereiden, de voorkennis van de leerlingen activeren en nadenken over mogelijkheden om leerlingen zelfstandig te laten werken, diverse didactische werkvormen gebruiken, en de interactie aangaan met leerlingen.

De motieven voor de keuze

Gevraagd naar de motieven voor de gemaakte keuze zijn antwoorden gegeven die kunnen worden ingedeeld in twee categorieën: gebaseerd op de gekozen taken zelf en gebaseerd op een doel of kader buiten de gekozen taken. Motieven die zijn gebaseerd op de taken zelf konden worden onderscheiden in twee typen:

a. Maakt onderscheid in voorwaardelijke taken en kerntaken of primaire taken. De respondenten die voor deze onderscheiding hebben gekozen, relateren dit onderscheid vaak aan taak 1 (zorgen voor een goed leerklimaat) en taak 7 (leerlingen stimuleren en activeren) en iets minder vaak aan taak 3 (geven van les, uitleg en instructie) en taak 8 (leerlingen vaardigheden aanleren). Daarnaast worden de taken genoemd die betrekking hebben op het begeleiden en coachen van leerlingen in dit verband (taak 11, 13, 15, 17), maar duidelijk minder vaak.
b. Maakt onderscheid in taken die meer algemeen zijn of andere taken omvatten, en meer specifieke taken. De taken 1 (zorgen voor een goed leerklimaat) en 15 (leerlingen ondersteunen bij het sturen van hun leerproces) worden in dit verband het meest genoemd als algemeen of omvattend.

Ook motieven gebaseerd op een doel of kader buiten de gekozen taken waren er in twee typen:

a. Sluit aan bij een centraal uitgangspunt van de respondent zelf (thema, inhoudelijk idee, persoonlijke theorie, opvatting): Deze taken maken deel uit van het leerproces; Ze zijn allemaal actiegericht; Je hebt als expert de rol van kennisloket; Hierbij staat de zelfsturing centraal; Hiermee leren leerlingen veel; Leren ontstaat als er meer gebeurt dan kennisoverdracht van A naar B, er moet worden gewerkt d.m.v. denkvragen, samenwerken, complexe opdrachten etc.; Het accent ligt voor mij op de coachende rol van de docent; Hiermee worden leerlingen uitgedaagd.
b. Verwijst naar het onderwijsconcept van de school.

De redenen waarom een goede docent deze dingen moet doen

De vraag waarom een goede docent deze dingen moet doen, leidde vooral tot drie typen antwoorden:

- Om voorwaarden voor leren te creëren (zorgen voor een veilige omgeving, duidelijk zijn in wat je verwacht; leerlingen motiveren; het leerrendement maximaliseren; voor leerlingen een “leermeester” zijn die het eerst voordoet) (23x)
- Om bepaalde leerdoelen te stimuleren (naast kennis ook vaardigheden aanbieden; leerstof aanbieden, maar ook aandacht voor bredere ontwikkeling) (13x)
- Om resultaten te bereiken op de langere termijn (om leerlingen optimaal te laten deelnemen aan de maatschappij, om kansen te creëren voor de toekomst, om een attitude te ontwikkelen voor de rest van het leven) (8x)

Enkele andere antwoorden werden elk een paar maal gegeven: om bij leerlingen activiteit en zelfsturing te bevorderen; om in te kunnen spelen op verschillen tussen leerlingen; om tegemoet te komen aan de visie van de school; om recht te doen aan eigen functioneren en professionaliteit.

De argumentatie waarop de uitspraken worden gebaseerd
Bij de argumentatie voor de gemaakte keuzen noemt men vaak de eigen ervaring (32x). Ook worden genoemd: gesprekken met andere betrokkenen (bijvoorbeeld intervisie en feedback, 11x) en bepaalde uitgangspunten of theorieën en vakliteratuur (10x). Het schoolconcept en de eindexamennormen worden beide slechts door een enkeling genoemd als argument voor de gemaakte keuzen.

Dilemma's die docenten zelf tegen komen bij het uitvoeren van de gekozen taken
Aan de docenten (N=16) is gevraagd of zij bij het uitvoeren van de door hen gekozen taken nog knelpunten of dilemma's tegenkomen. De antwoorden waren divers maar zijn als volgt in te delen:

- voorwaarden (gebrek aan tijd, gebrek aan contacttijd met leerlingen; weinig tijd om leerproces te kunnen volgen; de grootte en samenstelling van klassen (diversiteit in typen leerlingen, aantal rugzakleerlingen))
- elkaar beconcurrerende doelen (aandacht voor vakkennis versus vaardigheden; bieden van vrijheid versus voorkomen van vrijblijvendheid; eisen stellen versus motiveren en interesse wekken; inspelen op verschillen versus nadruk op samenwerking)
- het aanbod aan leerstof in methodes (veel "onzinnige" informatie)
- een verschil in opvattingen over te voeren beleid met de directie

De oplossingen waarvoor deze docenten kiezen als zij dilemma's willen hanteren, zijn ook divers van aard (mede samenhangend met het type dilemma): zelf extra tijd investeren, eigen welbevinden als uitgangspunt nemen en de gewenste aanpak aanpassen, keuzes maken en prioriteiten stellen (gericht op bepaalde groepen of leerlingen), nadruk leggen op bepaalde leerdoelen (juist op het sociale aspect of juist op het bereiken van een optimaal eindexamenresultaat), zich conformeren aan het beleid.

De standaard bij de uitvoering van de gekozen taken volgens schoolleiders en teamleiders
Aan de schoolleiders en teamleiders (N=31) is gevraagd in welke mate een goede docent de gekozen taken moet uitvoeren (de standaard). Zij gaven daarop de volgende antwoorden: de docent moet

- achter de waarden van de school staan en bereid zijn zich te ontwikkelen (willen leren, reflecteren, samenwerken, open staan voor feedback, gedreven zijn, zijn best doen) (14x)
- een bepaalde breedte in het repertoire beheersen (alle taken, een groot deel) (2x)
- een bepaald gewenst niveau van handelen beheersen (voor 80 %; in hoge mate; streven naar excellentie; hoogst mogelijke niveau; heel goed) (6x)
- niet alle taken beheersen (mag tussen docenten verschillen; ieder brengt eigen kwaliteiten in) (2x)
- bepaalde leerresultaten bij leerlingen behalen (5x)
- positieve reacties van leerlingen nastreven (3x)
- bepaalde denkactiviteiten bij leerlingen stimuleren (3x)

De gronden waarop de standaarden zijn gebaseerd volgens schoolleiders en teamleiders

Aan de schoolleiders en teamleiders is ook gevraagd op welke gronden zij deze standaarden baseren. De leverde vooral de volgende antwoorden op: mijn visie of mijn persoonlijke overtuiging (11x), ervaring, intuïtie (8x). Daarnaast werden vakliteratuur, onderzoek, bijscholing (4x) en discussies met andere betrokkenen en consensus (4x) genoemd. Enkelen noemden nog observaties van lessen (2x).

Het leraarschap in deze tijd

Op de (aan alle respondenten gestelde) vraag of de gegeven antwoorden samenhangen met de manier waarop men aankijkt tegen het leraarschap in deze tijd, antwoordt 77% van de schoolleiders, teamleiders en docenten bevestigend (36 van de in totaal 47 respondenten). In de toelichting hierop werd aangegeven dat het takenpakket van de docent is veranderd en uitgebreid (meer begeleiden (individueel en in groepen); meer maatwerk leveren; bij meer vakken kunnen begeleiden; leerlingen kunnen motiveren; normen en waarden aanleren) (11x); dat de school meer opvoedingstaken heeft gekregen dan vroeger (verschuiving van opvoedingstaken van ouders naar school) (3x); dat de samenleving andere eisen stelt aan toekomstige werknemers en het onderwijs de leerlingen daarop moet voorbereiden (5x); en dat de leerlingen van nu anders zijn dan de leerlingen van vroeger (meer autonoom, soms vluchtiger, willen gezien worden; ze hebben veel andere dingen aan hun hoofd) (7x).

Eerste ronde: leerlingen

In de analyse van de onderwijsconcepten van de scholen (zie 7.3.3) zijn ook de uitkomsten meegenomen van de interviews met groepjes leerlingen. Hun inbreng betrof vooral de taken 1 (zorgen voor een goed leerklimaat), 3 (lesgeven) en 20 (onderling samenwerken). Wat docenten volgens hen moeten doen, valt als volgt te ordenen. Docenten moeten a) leerlingen respectvol en onderling gelijk behandelen; b) contact met hen leggen, een band met hen opbouwen, aardig zijn; c) duidelijke kaders en regels stellen en consequent zijn; d) goed voorbereid, enthousiast en afwisselend lesgeven, iets toevoegen aan het boek, duidelijk en met voorbeelden uitleggen, niet teveel in één les proppen, begeleiding geven, feedback geven, toetsen nabespreken, uitleggen waarom bepaalde zaken nodig zijn; e) samenwerken en afstemmen met andere docenten over planning en eisen. Deze aspecten aan leerklimaat (a t/m c), lesgeven (d) en samenwerken (e) zijn op alle vier scholen genoemd. Leerlingen op één school (de school waar in het beschreven onderwijsconcept expliciet inzichten uit de cognitieve psychologie zijn verwerkt) noemden meer specifieke aspecten van het lesgeven dan leerlingen op de andere scholen.

Tweede ronde: schoolleiders, teamleiders, docenten en stakeholders

De individuele keuzen bij de paarsgewijs aangeboden taken

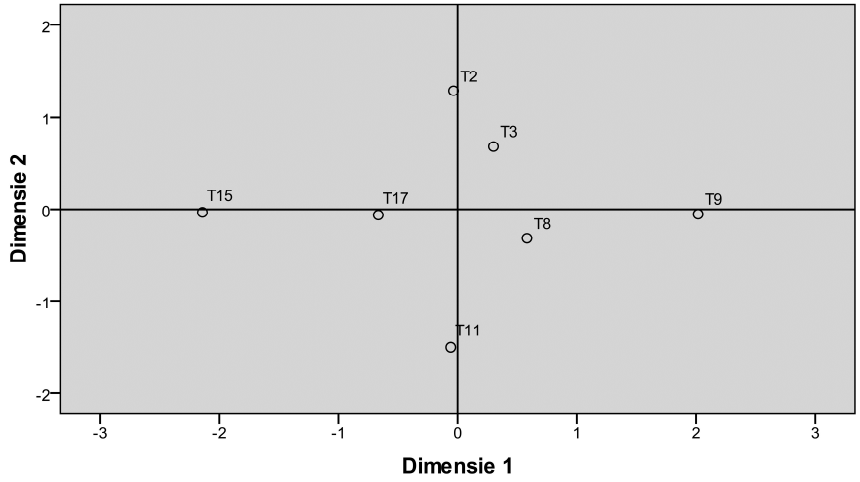
Bij de analyse van de data van de paarsgewijze vergelijkingen met behulp van Alscal is gekozen voor een tweedimensionale oplossing. Hiermee werd in de groep respondenten als geheel (N=57) een percentage verklaarde variantie bereikt van 88% en deze oplossing was ook inhoudelijk goed interpreteerbaar. De dimensies bleken bovendien redelijk stabiel, zowel bij vergelijking tussen de stakeholders enerzijds en de schoolleiders, teamleiders en docenten anderzijds, als bij vergelijking bij de laatste groep tussen de analyse op zeven taken (dezelfde als bij de stakeholders) en op alle acht aangeboden taken (met taak 16 als extra aangeboden taak).
Tabel 7.4 bevat de coördinaten van de taken op de twee dimensies bij de stakeholders (over 7 taken) en bij de schoolleiders, teamleiders en docenten (over 7 en over 8 taken).

Tabel 7.4: Uitkomsten multidimensionele schaalanalyses (Alscal) in twee dimensies

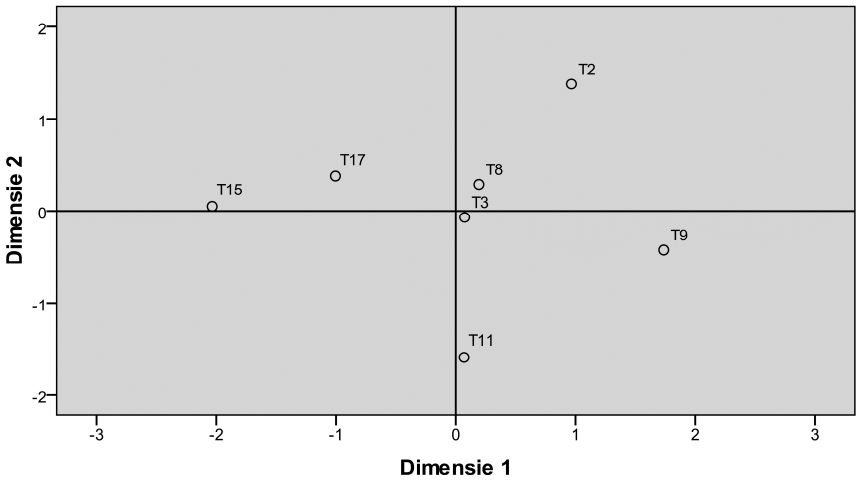
	Stakeholders (N=18)		Schoolleiders, teamleiders, docenten (N=39)		Schoolleiders, teamleiders, docenten (N=39)	
Aantal taken	*7*		*7*		*8*	
RSQ	.94		.88		.85	
Dimensies	*1*	*2*	*1*	*2*	*1*	*2*
T2	-.03	1.28	.97	1.38	1.17	1.04
T3	.30	.68	.07	-.07	-.31	.64
T8	.58	-.31	.19	.29	-.17	.33
T9	2.02	-.05	1.74	-.43	1.31	-.27
T11	-.06	-1.50	0.07	-1.59	-.39	-1.16
T15	-2.14	-.03	-2.03	.05	-2.00	-.28
T16	-	-	-	-	1.51	-1.15
T17	-.67	-.06	-1.01	.37	-1.11	.85

De gegevens in tabel 7.4 worden grafisch weergegeven in de figuren 7.3 t/m 7.5. Voor een juiste interpretatie hiervan moet men zich realiseren dat het bij de getallen op X- en Y-as niet gaat om hoe vaak de taken zijn gekozen, maar om *de posities van de taken op de twee dimensies* die door de analysetechniek zijn gegenereerd. Deze dimensies moeten inhoudelijk worden geïnterpreteerd. Bij *de afstanden tussen de taken* in de grafiek kan worden opgemerkt dat als taken dicht bij elkaar liggen, dit kan worden opgevat als een aanwijzing dat deze taken in het referentiekader van de respondenten ook dicht bij elkaar liggen, en dus op een of andere manier in hun denken op elkaar lijken of met elkaar te maken hebben. Het omgekeerde geldt natuurlijk ook: verder van elkaar liggende taken worden kennelijk in het denken van de docenten als verschillend gezien.

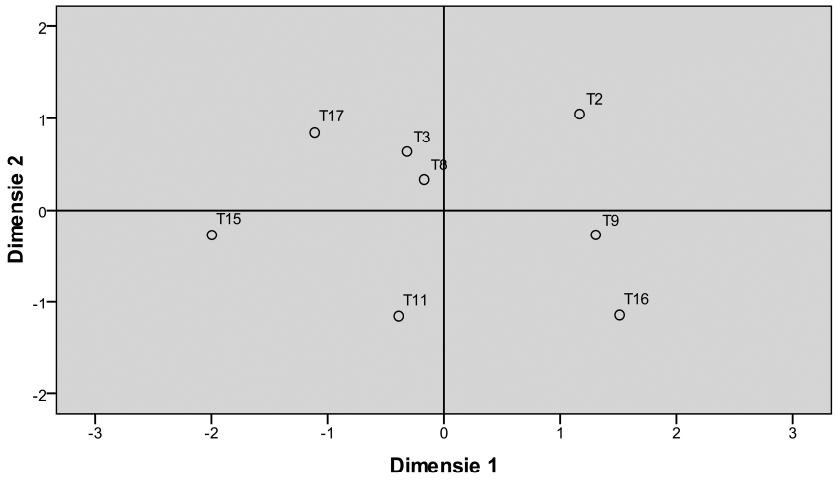
De figuren 7.3 en 7.4 bevatten de resultaten over zeven taken bij de stakeholders en bij de schoolleiders, teamleiders en docenten. Beide figuren lijken vrij sterk op elkaar. Kennelijk variëren de populariteiten van de zeven geselecteerde taken bij de stakeholders ongeveer op dezelfde wijze als bij de schoolleiders, teamleiders en docenten, waardoor de taken in beide grafieken ongeveer dezelfde posities op beide dimensies innemen. Dit is een aanwijzing dat de referentiekaders van beide groepen in behoorlijke mate overeenkomen.



Figuur 7.3: Weergave van 7 taken op twee dimensies bij de stakeholders (N=18)



Figuur 7.4: Weergave van 7 taken op twee dimensies bij schoolleiders, teamleiders, docenten (N=39)



Figuur 7.5: Weergave van 8 taken op twee dimensies bij schoolleiders, teamleiders, docenten (N=39)

Op de eerste dimensie in de figuren 7.3 en 7.4 (horizontale as) zien we aan de linkerkant de taken 15 (leerlingen ondersteunen bij het sturen van hun leerproces) en 17 (geven van feedback aan leerlingen), in het midden de taken 3 (geven van les, uitleg en instructie) en 8 (leerlingen vaardigheden aanleren) en aan de rechterkant taak 9 (inspelen op verschillen tussen leerlingen).
Op de tweede dimensie (verticale as) zien we onderin taak 11 (leerlingen laten samenwerken en hen daarbij ondersteunen); daarboven opnieuw de taken 3 (geven van les, uitleg en instructie) en 8 (leerlingen vaardigheden aanleren) en bovenin taak 2 (bijdragen aan de persoonlijke vorming van leerlingen).
De taken 3 en 8 blijken op beide dimensies op een middenpositie te liggen, dus min of meer neutraal ten opzichte van de uitersten. Deze taken behoren tot de traditionele basistaken van een docent.
Vanuit het midden geredeneerd ligt op de verticale dimensie taak 2 (bijdragen aan persoonlijke ontwikkeling) als uiterste aan de ene kant en taak 11 (leerlingen laten samenwerken en hen daarbij ondersteunen) aan de andere kant. Daarmee lijkt het alsof men vanuit de basistaken aan de ene kant zich meer kan richten op het bijdragen aan de persoonlijke vorming van een leerling en aan de andere kant meer aandacht kan hebben voor de sociale vorming (het samenwerken). Gezien de vrij grote afstand tussen deze taken op de verticale as lijken deze taken voor de respondenten nogal verschillende keuzes te vertegenwoordigen. De verticale as lijkt inhoudelijk te kunnen worden getypeerd als een vormingsdimensie, waarop de docent zich vanuit zijn basistaken meer kan richten op bijdragen aan de persoonlijke vorming van leerlingen aan de ene kant of op het samenwerken tussen leerlingen aan de andere kant.
Op de eerste dimensie (horizontale as) zien we aan de linkerkant de taken 15 (leerlingen ondersteunen bij het leren sturen van hun leerproces) en 17 (geven van

feedback). Deze taken hebben beide betrekking op het begeleiden van het cognitieve leerproces bij leerlingen, In het midden van de dimensie zien we opnieuw de taken 3 en 8 en aan de rechterkant zien we taak 9 (inspelen op verschillen tussen leerlingen). Kennelijk staat voor de respondenten het begeleiden van het leerproces van leerlingen min of meer los van het inspelen op verschillen bij leerlingen. Dit zou kunnen betekenen dat het ondersteunen bij het leren sturen van het eigen leerproces wordt opgevat als iets dat min of meer algemeen opgaat voor leerlingen en dat het leveren van maatwerk daarbij niet op de voorgrond staat. Deze eerste dimensie is dan zo op te vatten dat de docent vanuit zijn primaire taken meer nadruk kan leggen op enerzijds het begeleiden van het leerproces of anderzijds het inspelen op verschillen. Mogelijk wordt het leerproces vooral geassocieerd met de te leren vakinhoud en het inspelen op verschillen vooral met de leerlingen als persoon.
Bij vergelijking van de figuren 7.3 en 7.4 zien we als duidelijkste verschil dat bij de stakeholders de taken 3, 8 en 17 rondom het midden liggen op ongeveer even grote afstanden van elkaar. Bij de schoolleiders, teamleiders en docenten ligt taak 17 (het geven van feedback aan leerlingen) duidelijk op wat een grotere afstand van de taken 3 en 8, die juist samen veel meer een cluster vormen. Kennelijk is het zo dat voor deze laatste groep de taken 3 en 8 dicht bij elkaar liggen en dat het geven van feedback daar niet direct bij hoort, maar meer bij het ondersteunen van het leerproces.
In figuur 7.5 wordt zichtbaar wat de onderlinge posities van de taken worden als een andere taak wordt toegevoegd. De toegevoegde taak 16 (volgen wat leerlingen leren en hoe ze zich ontwikkelen) krijgt op de horizontale dimensie een positie aan de rechterkant en lijkt daarmee te passen bij het inspelen op verschillen tussen leerlingen en minder bij het cognitieve leerproces dat zich aan de linkerkant bevindt. Op de verticale dimensie ligt taak 16 aan de onderkant, zodat volgen wat leerlingen leren meer lijkt te passen bij de samenwerking tussen leerlingen dan bij de persoonlijke vorming van leerlingen die zich aan de bovenkant van deze dimensie bevindt.
Hoewel de posities van de andere taken door toevoegen van taak 16 iets verschuiven, is deze verschuiving beperkt en blijven de posities van de taken ten opzichte van elkaar min of meer dezelfde. Dat de weergave van de taken op deze twee dimensies zo stabiel is, ondersteunt de waarde van de afbeelding van de taken in deze twee dimensies en de daarop gebaseerde interpretaties.

Overeenkomsten en verschillen tussen de scholen

De Alscal-analyses kunnen ook worden uitgesplitst per school, zie tabel 7.5. Bij het vergelijken van clusters van taken die dicht bij elkaar liggen, worden de volgende verschillen tussen scholen zichtbaar.

- Op school A ligt het leren samenwerken (taak 11) dicht bij persoonlijke vorming (taak 2), het volgen van het leerproces (taak 16) bij het ondersteunen bij het leren sturen van het leerproces (taak 15), en het vaardigheden aanleren (taak 8) bij het inspelen op verschillen (taak 9).

- Op school B liggen het geven van les, uitleg en instructie (taak 3), het aanleren van vaardigheden (taak 8), en het leren samenwerken (taak 11) dicht bij elkaar. Het geven van feedback (taak 17) ligt vrij dicht bij bijdragen aan persoonlijke vorming (taak 2).
- Op school C liggen het bijdragen aan persoonlijke vorming (taak 2) en het leren van vaardigheden (taak 8) dicht bij elkaar, evenals het volgen van het leerproces (taak 16) bij het inspelen op verschillen (taak 9), en het leren samenwerken (taak 17) bij het lesgeven (taak 3).
- Op school D liggen vaardigheden aanleren (taak 8) en leerlingen ondersteunen bij het sturen van hun leerproces (taak 15) vrij dicht bij elkaar.

Tabel 7.5: Uitkomsten multidimensionele schaalanalyses over twee dimensies per school

	School A (N=8)		School B (N=9)		School C (N=10)		School D (N=12)	
Aantal taken	*8*		*8*		*8*		*8*	
RSQ	.52		.65		.65		.71	
Dimensies	*1*	*2*	*1*	*2*	*1*	*2*	*1*	*2*
T2	.20	1.46	.04	1.09	.15	1.35	1.34	.98
T3	-.19	-1.16	-.12	-.33	-.89	.31	-1.47	.93
T8	.61	-.69	-.12	-.33	.25	.86	-1.12	-.36
T9	.56	-.67	1.56	-.1.03	1.30	.22	.55	.44
T11	-.03	1.11	-.58	-.50	.17	-1.20	1.66	-.60
T15	-1.62	-.02	-2.17	-.66	-1.13	-1.49	-1.21	-.75
T16	-1.62	-.17	1.69	.23	1.75	-.25	.17	-1.54
T17	2.08	.15	-.30	1.53	-1.60	.21	.08	.90

De beide dimensies komen op de scholen A, B en C ongeveer overeen en lijken op het algemene beeld (zie de figuren 7.3-7.5). De 2^{e} dimensie op de scholen A-C vormt op school D de 1^{e} dimensie. De 2^{e} dimensie op school D kan mogelijk worden gezien als lopend van docent zelf actief (lesgeven, feedback geven) naar docent meer observerend (volgen leerproces).

Uit de per school verschillende manieren waarop de taken vanuit de keuzen van de respondenten gezien dicht bij elkaar liggen, kan mogelijk worden geconcludeerd dat de leerdoelen die de afgelopen 10-15 jaar meer worden benadrukt, zoals vaardigheden en samenwerking, per school een iets andere betekenis kunnen krijgen. Vaardigheden aanleren kan samenhangen met inspelen op verschillen (verschillen in vaardigheid), persoonlijke vorming (vaardigheid als vormingsaspect), leren samenwerken (samenwerken als vaardigheid), geven van les, uitleg en instructie (instructie in vaardigheden) en leren sturen van het leerproces (zelfsturing als vaardigheid). Samenwerking kan samenhangen met persoonlijke vorming (het sociale als vormingsaspect), vaardigheden aanleren (samenwerking als vaardigheid) en geven van les, uitleg en instructie (instructie in samenwerken).

Verschillen tussen scholen worden ook zichtbaar als we kijken naar wat het meest centraal ligt. Op school B liggen de taken 3 (geven van les, uitleg en instructie), 8 (leerlingen vaardigheden aanleren) en 11 (leren samenwerken) duidelijk in het midden. Dit zou kunnen betekenen dat voor de respondenten op deze school dit ook de meest centrale taken zijn. Op school D ligt taak 9 (inspelen op verschillen) centraal. Dit zou kunnen betekenen dat voor de respondenten op deze school het inspelen op de verschillen tussen leerlingen een centrale betekenis heeft, van waaruit de andere taken worden gekozen en ingevuld. Op beide andere scholen is een centraal punt minder zichtbaar.

Gronden voor de gemaakte keuzes bij de paarsgewijze vergelijkingen

Na elke paarsgewijze vergelijking werd gevraagd naar de grond(en) voor de gemaakte keuze. Daarbij werden steeds de volgende vijf gronden aangeboden: (omdat deze taak)

- het meeste bijdraagt aan het leren en de ontwikkeling van leerlingen volgens wat daarover bekend is
- in de praktijk vaak het belangrijkste voor een leraar blijkt te zijn om aandacht aan te besteden
- volgens mijn persoonlijke overtuiging een essentiële taak is van een leraar op deze school
- in de visie van de school tot de verantwoordelijkheid van de leraar behoort en dus van hem wordt verwacht
- het meest een appel doet op wat de leraar voor zijn leerlingen zou moeten betekenen

De stakeholders kruisten gemiddeld 1,4 van deze gronden aan en de schoolleiders, teamleiders en docenten gemiddeld 1,9 (gemiddeld over alle keuzeparen).
Uit vergelijking van de totaalpercentages bij de stakeholders met de subtotaal-percentages over 21 keuzeparen bij de andere respondenten blijkt dat de mate waarin voor de verschillende gronden is gekozen tussen beide groepen niet veel verschilt. Beide groepen hebben grond 3 het meest gekozen ("omdat de gekozen taak volgens mijn persoonlijke overtuiging een essentiële taak is van een leraar op deze school"). Bij de stakeholders gaat het hierbij om 45 % en bij de schoolleiders, teamleiders en docenten om 49 % van de betrokkenen. De stakeholders hebben iets vaker ook voor de gronden 4 en 5 gekozen (44% en 23%; de andere respondenten 36% en 15 %). Grond 4 luidde als volgt: "(omdat de gekozen taak) in de visie van deze school tot de verantwoordelijkheid van de leraar behoort en dus van hem wordt verwacht". Grond 5 luidde: "(omdat de gekozen taak) het meest een appel doet op wat de leraar voor zijn leerlingen zou moeten betekenen". Deze verschillen tussen beide groepen in de mate waarin de gronden zijn aangekruist zijn niet groot en niet significant.
Per keuzepaar zijn er wel vrij duidelijke verschillen in gekozen gronden tussen beide groepen respondenten. Bij de stakeholders is relatief vaak (11 van de 21 keer) grond 3 de meest gekozen grond ("volgens mijn persoonlijke overtuiging een essentiële taak is van een leraar op deze school"). Verder kozen zij vrij vaak grond 4 (9 keer) ("in de

visie van deze school tot de verantwoordelijkheid van de leraar behoort en dus van hem wordt verwacht"). De schoolleiders, teamleiders en docenten kozen bij 16 van deze 21 keuzeparen voor grond 3 en slechts weinig voor de andere gronden, zodat het interessant is stil te staan bij de keuzeparen waarbij dit het geval was.
Bij de paren waarbij het vaakst werd gekozen voor grond 4 (gerelateerd aan de visie van de school) (de paren 3, 11 en 21) werd taak 15 (ondersteunen bij het leerproces) gemiddeld belangrijker gevonden dan taak 11 (laten samenwerken en daarbij ondersteunen), taak 3 (geven van les) belangrijker dan taak 11 (laten samenwerken en daarbij ondersteunen) en taak 8 (vaardigheden aanleren) belangrijker dan taak 11 (laten samenwerken en daarbij ondersteunen). Mogelijk neemt het samenwerken tussen leerlingen in de visie van de scholen toch niet zo'n centrale plaats in. Bij de paren waarbij ze het vaakst kozen voor grond 1 (gerelateerd aan dat wat bekend is over leren en de ontwikkeling van leerlingen) (de paren 8 en 17) vonden ze taak 8 (vaardigheden aanleren) belangrijker dan taak 9 (inspelen op verschillen) en ook taak 17 (geven van feedback) belangrijker dan taak 9 (inspelen op verschillen). Mogelijk zijn vaardigheden en feedback voor de respondenten meer gerelateerd aan kennis over leren en ontwikkeling dan inspelen op verschillen. Bij het paar waarbij het vaakst werd gekozen voor grond 2 (gerelateerd aan wat in de praktijk de belangrijkste taak is) (keuzepaar 12) werd taak 3 (geven van les, uitleg en instructie) belangrijker gevonden dan taak 8 (leerlingen vaardigheden aanleren).

Overeenkomsten en verschillen tussen scholen in de gronden voor de gemaakte keuzes

In de tabellen 7.6 en 7.7 wordt een overzicht gegeven van de proporties waarmee de vijf aangeboden gronden zijn gekozen per school, bij de stakeholders en bij de schoolleiders, teamleiders en docenten. Per regel zijn de proporties die het hoogste zijn en/of tenminste 0.50 zijn, vet gezet.
De mate waarin de stakeholders de diverse gronden hebben aangekruist, verschilt tussen de scholen, vooral bij grond 4, waarbij de stakeholders bij de scholen B en D deze grond duidelijk vaker aankruisten dan die bij school C en vooral school A. Deze verschillen bij grond 4 zijn statistisch significant ($F(3,14) = 3{,}93$ bij $p<.05$). De stakeholders bij de scholen B en D refereerden vrij vaak aan de visie van de school, de stakeholders bij school A deden dat amper.
Door de stakeholders is bij de helft van de keuzeparen een zelf geformuleerde grond toegevoegd. Deze antwoorden hebben vooral betrekking op het feit dat de ene taak als voorwaarde wordt gezien voor de andere of dat men een bepaalde taak als een kerntaak van een docent ziet.
Uit tabel 7.7 blijkt dat op alle scholen door de schoolleiders, teamleiders en docenten grond 3 het meest wordt gebruikt om de gemaakte keuze van de belangrijkste taak per keuzepaar te funderen. De percentages per grond verschillen niet significant tussen de scholen. Ook verschillen de antwoorden niet significant tussen schoolleiders, teamleiders en docenten onderling.

Tabel 7.6: Proporties en sd per school van gekozen gronden, stakeholders

Over 21 keuzeparen		*Grond 1*	*Grond 2*	*Grond 3*	*Grond 4*	*Grond 5*
School A	P	0.29	0.26	**0.33**	0.06*	0.32
	sd.	0.31	0.30	0.37	0.07	0.33
School B	P	0.48	0.16	**0.56**	**0.67***	0.27
	sd	0.34	0.17	0.28	0.27	0.33
School C	P	0.30	0.04	**0.45**	0.39*	0.07
	sd	0.16	0.46	0.40	0.41	0.09
School D	P	0.38	0.24	0.42	**0.54***	0.28
	sd.	0.20	0.15	0.35	0.27	0.20

* Percentages zijn statistisch significant verschillend tussen de scholen (p<.05)

Tabel 7.7: Proporties en sd per school van gekozen gronden, schoolleiders, teamleiders en docenten

Over 21 keuzeparen		*Grond 1*	*Grond 2*	*Grond 3*	*Grond 4*	*Grond 5*
School A	P	0.27	0.25	**0.46**	0.26	0.14
	sd	0.23	0.26	0.31	0.22	0.22
School B	P	0.40	0.24	**0.44**	0.37	0.10
	sd	0.36	0.19	0.34	0.37	0.10
School C	P	0.34	0.11	**0.46**	0.36	0.19
	sd	0.18	0.08	0.22	0.18	0.21
School D	P	0.37	0.28	**0.55**	0.43	0.21
	sd	0.31	0.27	0.35	0.32	0.16

Van de schoolleiders, teamleiders en docenten heeft 30 % (N=12) eenmaal of enkele malen een eigen grond toegevoegd. Dit gebeurde bij 12 % van alle (1092) keuzen, als aanvulling (naast een of meer aangekruiste aangeboden gronden). De betreffende antwoorden hielden in dat de als belangrijkste gekozen taak een voorwaarde is voor of een onderdeel is van de andere taak in het paar, of dat je als je aan de ene (gekozen) taak werkt ook aan de andere taak werkt, of dat de gekozen taak een persoonlijke uitdaging vormt.

7.4 Conclusies en discussie

Aanleiding en vraagstelling

De vraagstelling was: wat zijn de overeenkomsten en verschillen tussen aanwijzingen voor gewenst handelen van docenten en de gronden waarop die zijn gebaseerd vanuit wetenschappelijke literatuur, landelijke beleidsdocumenten, onderwijsconcepten van scholen en opvattingen van betrokkenen?

Aanwijzingen per bron

Uit de *wetenschappelijke literatuur* (deelvraag 1) blijkt dat van docenten handelen wordt verwacht op drie niveaus: creëren van een goede leeromgeving, vervullen van een aantal pedagogisch-didactische taken, en tijdens interacties met leerlingen een gedifferentieerd gespreksrepertoire hanteren (vragen stellen, hints geven, etc.).
Aanwijzingen voor wenselijke kenmerken van de leeromgeving en de te vervullen taken zijn vooral te vinden in theorie en onderzoek in de cognitieve psychologie. Centraal staat het bevorderen van cognitieve en metacognitieve denkprocessen, gericht op analyseren, redeneren en problemen oplossen binnen een bepaald domein van kennis en vaardigheid. De gewenste taken zijn het duidelijkste beschreven in het *cognitive apprenticeship model*: uitleggen, instrueren, voordoen, bevorderen van discussie, laten oefenen en daarbij ondersteunen, op maat hulp bieden en feedback geven, en beoordelen aan een standaard, leerlingen denkbeelden en leerervaringen laten articuleren en hen daarop laten reflecteren, en hen aanmoedigen tot verdere exploratie. Docenten coachen met dit repertoire van handelen hun leerlingen naar steeds meer zelfstandig denken en handelen in het betreffende domein, mede op basis van de kennis die zij (de leerlingen) in dat domein ontwikkelen.
Aanvullend blijft vanuit de behavioristische traditie goed klassenmanagement van belang en gestructureerd aanbieden van informatie en laten oefenen en automatiseren van vaardigheden, en wordt vanuit sociaal-constructivistische opvattingen het belang benadrukt van sociale interacties tussen leerlingen onderling (samenwerkend leren) en van werken aan betekenisvolle taken (geplaatst in een herkenbare context). Ook is aandacht wenselijk voor affectief-motivationele aspecten.
Deze principes zijn geen vaste procedures die docenten kunnen volgen, maar principes die zij, vanuit een goed begrip ervan, flexibel, op maat en in onderlinge samenhang kunnen toepassen.
Een eerdere onderzoekstraditie, gericht op wenselijke persoonlijke kenmerken van docenten, leidde tot de conclusie dat docenten deskundig, duidelijk, enthousiast, flexibel en taakgericht moeten zijn en hoge verwachtingen moeten hebben van de leerlingen en van zichzelf. Deze kenmerken hebben kennelijk hun geldigheid nog niet verloren want ze werden ook in dit onderzoek door de leerlingen genoemd. Momenteel worden hieraan als kenmerken toegevoegd dat docenten adaptief moeten kunnen werken vanuit een professioneel denkkader, waarin inzichten in wat leerlingen moeten leren, hoe zij dat kunnen leren, welke leeromgeving en welk docenthandelen daaraan kunnen bijdragen en hoe kan worden vastgesteld wat er is en wordt geleerd, in onderlinge samenhang worden ingezet.
In *beleidsgerichte publicaties* (deelvraag 2) wordt meer aandacht besteed aan wat leraren zouden moeten weten en kunnen dan aan hoe ze zouden moeten handelen. Traditioneel verwacht men van docenten vakkennis en pedagogisch-didactische bekwaamheden. Tegenwoordig worden ook bekwaamheden voor samenwerken en verdere professionele ontwikkeling verwacht, en recent ook voor doen van onderzoek.

Vaak wordt opgemerkt dat bij een goede leraar de leerlingen zoveel mogelijk leren. De meest uitgewerkte beschrijving van daartoe gewenste pedagogisch-didactische bekwaamheden is te vinden in de SBL-competenties. Deze worden in ruime mate gedekt door de (22) taken in een instrument dat in een eerder onderzoek op verzoek van een aantal vernieuwende scholen is ontwikkeld (Sol & Stokking, 2008b; zie hoofdstuk 3) en die in dit onderzoek ook zijn gebruikt bij de gegevensverzameling.
Verder wordt verwacht dat een leraar beschikt over een breed handelingsrepertoire waaruit hij in elke specifieke situatie een adequate keuze maakt. Er is een trend om te verwachten dat de leraar zich daarbij mede baseert op kennis uit onderzoek. In antwoord op veranderingen en problemen werken scholen aan vormen van "nieuw leren": ontwikkeling van meer zelfstandig leren, aanbieden van betekenisvolle contexten en opdrachten (in leergebieden), en samenwerking door leerlingen. Deze vernieuwingen kunnen bijdragen aan leren maar goede begeleiding door een docent blijft daarbij wel nodig. Momenteel wordt ook meer aandacht gevraagd voor het signaleren en begeleiden van individuele leerlingen die met problemen kampen. Hoe dit zich verhoudt tot de voorgenoemde vernieuwingen is nog niet uitgekristalliseerd.
In *de onderwijsconcepten van de scholen* (deelvraag 3) staat een aantal docenttaken centraal. Belangrijk is dat docenten zorgen voor een goed leerklimaat en bijdragen aan persoonlijke vorming van leerlingen. Zij moeten inspirerende instructie geven, die aansluit bij het niveau van verschillende (groepen) leerlingen. Ze moeten leerlingen activeren en stimuleren en daarbij inspelen op verschillen tussen leerlingen en leerlingen keuzes bieden. Zij moeten het leerproces van leerlingen volgen en (liefst zoveel mogelijk op maat) leerlingen begeleiden en coachen bij het zelfstandig werken en het samenwerken. Zij moeten leerlingen feedback geven en de dialoog aangaan met leerlingen. Docenten moeten ook zelf bereid zijn tot reflecteren en tot het ontvangen van feedback. Ze moeten zich blijven ontwikkelen en samenwerken.
Dat wat de *leerlingen* op de scholen noemen als wenselijke kenmerken van docenten (deelvraag 4) past goed bij de onderwijsconcepten van de scholen en heeft vooral betrekking op drie taakgebieden van docenten: het leerklimaat, het lesgeven, en de onderlinge samenwerking.
De schoolleiders, teamleiders en docenten op de vier scholen vinden als groep de volgende zeven taken van docenten (uit de 22 taken die werden voorgelegd) de belangrijkste: zorgen voor een goed leerklimaat; geven van les, uitleg en instructie; leerlingen stimuleren en activeren; leerlingen ondersteunen bij het sturen van hun leerproces; geven van feedback aan leerlingen; bijdragen aan persoonlijke vorming van leerlingen; werken aan eigen verdere professionele ontwikkeling.
De verschillen in de mate waarin beide taken van elk aangeboden paar door de respondenten werden gekozen als belangrijkste taak, bleken te kunnen worden geordend op twee dimensies. Over alle scholen samen liggen daarbij de taken 'geven van les, uitleg en instructie' en 'leerlingen vaardigheden aanleren' in het midden. Deze zou men kunnen zien als traditionele basistaken van een docent. Op de ene dimensie die vervolgens kan worden onderscheiden, ligt de taak 'bijdragen aan de persoonlijke

vorming van leerlingen' aan de ene kant en de taak 'laten en leren samenwerken' aan de andere. Mogelijk beschouwt men deze vormingsaspecten enigszins als tegenpolen (het individuele en het sociale). Op de andere dimensie staan enerzijds de taak 'ondersteunen van het leerproces' (taak 15) en anderzijds de taak 'inspelen op verschillen tussen de leerlingen' (taak 9). Mogelijk wordt het ondersteunen van het leerproces primair gerelateerd aan de leerinhouden en niet direct in verband gebracht met inspelen op verschillen tussen leerlingen. Kennelijk kunnen taken die vanuit onderwijskundig oogpunt duidelijk met elkaar samenhangen op de onderscheiden dimensies op grote afstand van elkaar worden geplaatst en worden deze mogelijk niet direct op elkaar betrokken.

De betrokkenen op de scholen zijn ook bevraagd op hun achterliggende opvattingen en motieven. De gronden waarop men zich baseert bij het kiezen van de (maximaal zeven) belangrijkste taken van een docent op de betreffende school (eerste ronde), zijn vooral terug te voeren op de eigen ervaring en opvattingen, en minder op gesprekken met anderen, bepaalde theorieën of vakliteratuur, of het onderwijsconcept van de school. Veel respondenten geven aan dat de gegeven antwoorden samenhangen met het leraarschap in deze tijd, waarin volgens hen het takenpakket van de docent is veranderd en de leerlingen anders zijn dan vroeger, de school meer opvoedingstaken heeft gekregen en de voorbereiding van de leerlingen op de samenleving andere leerdoelen met zich meebrengt (zoals meer zelfstandigheid en het vermogen tot samenwerken en blijven leren).

De grond die de respondenten het vaakste hebben aangekruist voor hun keuze bij de paarsgewijze vergelijkingen van taken (tweede ronde), was dat de gekozen taak naar hun persoonlijke overtuiging een essentiële taak is van een leraar op de betreffende school. Over alle gemaakte keuzen samen, dus over alle aangeboden paren van steeds twee verschillende taken van docenten, waren hierbij geen duidelijke verschillen tussen de schoolleiders, teamleiders en docenten van de vier scholen enerzijds en de externe stakeholders van deze scholen anderzijds. Per keuzepaar bezien kwam wel een duidelijk verschil naar voren. De stakeholders blijken gemiddeld vaker dan de schoolleiders, teamleiders en docenten te hebben gekozen voor de grond dat de gekozen taak in de visie van de school tot de verantwoordelijkheid van de leraar behoort en dus van hem of haar wordt verwacht. Dit gemiddelde verschil werd echter vooral veroorzaakt doordat op enkele van de vier scholen het onderwijsconcept voor de externe stakeholders kennelijk opvallend duidelijk was.

Vergelijking tussen de bronnen

De uitspraken over gewenst handelen van docenten die uit deze bronnen zijn verkregen, vertonen zowel overeenkomsten als verschillen (deelvraag 5). De set van 22 taken die eerder is opgesteld en nu is gebruikt bij de bevraging van betrokkenen, kan dienen als kader voor de vergelijking, omdat die set het domein van pedagogisch-didactisch handelen breed dekt.

In de geselecteerde wetenschappelijke literatuur ontbreekt de persoonlijke vorming van de leerlingen, maar ook over pedagogische doelen is wetenschappelijke literatuur beschikbaar (bijvoorbeeld Veugelers & Volman, 2007). De bestudeerde literatuur besteedt meer dan de andere bronnen aandacht aan op leren gerichte interacties met leerlingen (tijdens het coachen). De reeks taken van docenten in het *cognitive apprenticeship* model is terug te vinden in de set van 22 taken.
In de beleidsgerichte bronnen zijn het vooral de SBL-competenties die een breed scala aan pedagogisch-didactisch handelen beschrijven. De in de wetenschappelijke literatuur genoemde taken komen daarin min of meer allemaal terug. Aanvullend op de bestudeerde wetenschappelijke literatuur wordt in het recente beleid speciale aandacht gevraagd voor het doen van onderzoek door docenten, en voor het signaleren en begeleiden van individuele leerlingen die met problemen kampen.
In de onderwijsconcepten van de scholen wordt eveneens een breed pallet van taken beschreven. Daarin lijkt een zeker accent te liggen op pedagogische doelen en op het zoveel mogelijk aansluiten bij de leerlingen. In de wetenschappelijke literatuur (*cognitive apprenticeship*) en de SBL-competenties (organisatorisch, interpersoonlijk, pedagogisch, didactisch) is sprake van een organiserend kader. Dit komt in de onderwijsconcepten minder duidelijk naar voren. (De ordening die in dit hoofdstuk is gebruikt om de inhouden van de onderwijsconcepten van de scholen te beschrijven, is door de onderzoeker aangebracht).
De betrokkenen op de scholen kiezen, indien zij worden gevraagd in het geheel van taken prioriteiten te stellen, vooral voor taken die tot het meer traditionele repertoire behoren, zoals zorgen voor een goed leerklimaat, lesgeven, leerlingen stimuleren en activeren, bijdragen aan persoonlijke ontwikkeling. De in de wetenschappelijke literatuur, beleidsdocumenten en onderwijsconcepten genoemde meer vernieuwende taken, gericht op zelfstandig leren en samenwerken van leerlingen, en het meer inspelen op verschillen tussen leerlingen, worden duidelijk minder gekozen.

Discussie

De onderwijsconcepten kunnen omvattend worden genoemd: ze bestrijken een heel scala aan leerdoelen, kenmerken van leeromgevingen en van docenten verwachte taken. Ze vertegenwoordigen daardoor meer een verzameling van (deels nogal verschillende) ambities dan een samenhangende richtinggevende visie. Dit biedt mogelijk een verklaring voor de bevinding dat de te onderscheiden taken van docenten door de betrokkenen binnen de scholen nogal verschillend blijken te kunnen worden gepercipieerd, geprioriteerd en geordend.
De onderwijsconcepten kunnen in hun huidige vorm duidelijker worden verbonden met (nieuwe) pedagogisch en maatschappelijk gewenste doelen (die doorklinken in landelijke beleidsdocumenten), dan met beschikbare wetenschappelijke kennis over leren en leeromgevingen. Ook de gronden die betrokkenen aanvoeren voor gemaakte

keuzes tussen taken, hebben, mogelijk mede daardoor, weinig te maken met beschikbare kennis.

De taakperceptie van de betrokkenen op de scholen (Kelchtermans, 2007; Luttenberg, 2000) lijkt, gezien deze uitkomsten, meer te verschillen in hun perceptie, prioritering en ordening van concrete taken dan in de gronden die ze daarvoor aanvoeren, die veelal te maken hebben met hun persoonlijke overtuiging aangaande gewenst handelen op de betreffende school.

Uit de analyses van de paarsgewijze vergelijkingen tussen acht taken bleek dat de ordening van deze taken bij één school (school D) tot een ander resultaat leidde dan op de andere drie scholen, terwijl in eerdere onderzoeken (zie de hoofdstukken 3 en 5) is gebleken dat de perceptie van de leerlingen van deze, meer traditionele school gemiddeld niet afweek van die van de leerlingen op de overige, sterk vernieuwende scholen die aan die onderzoeken deelnamen. Kennelijk kunnen docenten op verschillende scholen verschillend tegen de taken aankijken zonder dat dit tot uitdrukking hoeft te komen in hun handelen zoals door leerlingen wordt waargenomen. Aan de andere kant bleek in dit onderzoek dat docenten van een andere, ook meer traditionele school (school A) op overeenkomstige wijze aankijken tegen de taken als de twee sterk vernieuwende scholen die aan dit onderzoek deelnamen (scholen B en C). Kennelijk kan de manier waarop docenten aankijken tegen de taken samenhangen met de specifieke context van hun school (vergelijk Campbell et al., 2001), nog los van het meer traditioneel of meer vernieuwend onderwijsconcept van de school.

Naast de hierboven genoemde verschillen die konden worden geconstateerd tussen de scholen, bleken de keuzen van docenten ook individueel te verschillen, en de meest genoemde grond voor hun keuzen waarin wordt verwezen naar hun persoonlijke overtuiging kan ook zo worden begrepen dat docenten gezien de complexiteit van hun beroep (al helemaal in de huidige tijd waarin veel in beweging is en diverse veranderingen worden gevraagd) niet anders kunnen dan voor hun handelen op basis van hun eigen opvattingen en ervaringen een persoonlijke vorm vinden (vergelijk Brown & McIntyre, 1993; Onderwijsraad, 1998).

Dat de betrokkenen op de scholen vooral taken prioriteren die tot het traditionele repertoire van de docent kunnen worden gerekend en minder voor vernieuwende taken kiezen zoals die vanuit constructivistische opvattingen naar voren worden geschoven, is begrijpelijk als men in aanmerking neemt dat dergelijke vernieuwingen voor docenten lastige dilemma's met zich meebrengen (Windschitl, 2002).

De mogelijkheden om de onderwijsconcepten te verbinden met wetenschappelijke inzichten en met landelijke beleidsdocumenten worden op dit moment niet volledig benut. Meer specifiek kan wetenschappelijke kennis over belangrijke aangrijpingspunten voor het handelen van docenten wel degelijk worden verbonden met een aantal van de doelen en ambities in de onderwijsconcepten, zoals die met betrekking tot het leerklimaat, de zelfregulatie, de rol van interactie en samenwerking, werken met betekenisvolle taken en aandacht voor feedback en beoordeling.

Meerdere principes of aanpakken, zoals leerlingen betekenisvolle taken geven, leerlingen laten samenwerken, en de zelfregulatie van leerlingen bevorderen (wat onder meer betekent dat ze eigen keuzes moeten kunnen maken), worden in wetenschappelijke literatuur en/of beleidsdocumenten aanbevolen op drie gronden tegelijk. Ze kunnen *het beoogde cognitieve leerproces ondersteunen*, ze kunnen *bijdragen aan de motivatie van leerlingen*, en ze vertegenwoordigen *wenselijke leerdoelen op zich* (bepaalde taken kunnen vervullen en problemen oplossen; kunnen samenwerken; het eigen handelen, denken en leren kunnen reguleren).

De wetenschappelijke literatuur en de beleidsdocumenten geven, tot slot, enkele aanwijzingen voor belangrijke mogelijk *aanvullingen* in de onderwijsconcepten en voor de manier waarop in de onderwijsconcepten en in het gewenste handelen van docenten prioriteiten kunnen worden gesteld. Deze wijzen in de richting van meer systematische coaching, op maat, over langere tijd, op weg naar meer zelfregulatie; meer systematisch signaleren van en inspelen op problemen bij individuele leerlingen; en meer expliciet kiezen en handelen op basis van onderzoek in de eigen praktijk (Collins et al., 1989; Ministerie van OC&W, 2008b).

De poging om uit publicaties over wetenschappelijke theorieën en onderzoeken aanwijzingen te halen voor gewenst handelen van docenten kan gelden als gelukt, want er is een aantal aanwijzingen verkregen, op verschillende niveaus van abstractie en specificatie, die elkaar deels bevestigen en deels aanvullen. Het is echter lastig te zeggen hoe overtuigend en generaliseerbaar deze aanwijzingen zijn. Op basis van wetenschappelijke theorieën kan de onderwijspraktijk niet zomaar worden getypeerd en, omgekeerd, uit onderzoeken naar de praktijk kunnen niet zomaar conclusies worden getrokken, want de praktijk is veelvormig en bevat tegenstrijdigheden, en wordt niet zuiver vormgegeven vanuit een bepaalde theorie of aanpak. Bovendien kunnen vanuit eenzelfde theorie verschillende aanbevelingen worden geformuleerd en omgekeerd verschillende theorieën tot overeenkomstige aanwijzingen komen.

Tot slot is duidelijk dat de aanwijzingen betrekking hebben op “content standards”, in de zin van aandachtspunten, en niet op “performance standards”, dat wil zeggen hoe vaak of intensief of goed een docent iets zou moeten doen (vergelijk Van der Schaaf, Stokking & Verloop, 2003). Handelen van docenten is waarschijnlijk slechts beperkt op die manier protocolleerbaar. Wel mag van leraren worden verwacht dat zij zich bij hun handelen door kennis uit theorie en onderzoek laten voeden. Niet ‘evidence-based’, maar ‘theory-and-research-informed’, waarbij zij aanwijzingen en mogelijkheden die in de literatuur worden beschreven, weloverwogen verbinden met hun specifieke werksituatie, het onderwijsconcept van de school en hun persoonlijke doelstellingen (vergelijk Opfer & Pedder, 2011).

8 Van huidig handelen naar gewenst handelen

8.1 Inleiding

De voorgaande vijf hoofdstukken zijn gebaseerd op onderzoeken die zijn uitgevoerd op verzoek van schoolleiders van een aantal vernieuwende en enkele meer traditionele scholen in het voortgezet onderwijs. De resultaten zijn steeds aan de schoolleiders, teamleiders en docenten teruggekoppeld en aan hen gepresenteerd en toegelicht (zie ook hoofdstuk 2). Het bleek voor hen niet eenvoudig om uitkomsten van onderzoek te interpreteren en te bedenken en bediscussiëren hoe zij daarvan gebruik zouden kunnen maken ten behoeve van verdere ontwikkeling (zie ook 4.4.6). Op basis van deze ervaringen en diverse literatuur (zie onder andere Bennebroek Gravenhorst, 2002, 2006; Runhaar, Sanders & Sleegers, 2009) kan worden geconcludeerd dat ten minste drie condities bij processen van verandering, vernieuwing en verbetering van pedagogisch-didactische aspecten van het onderwijs en benutting van resultaten van onderzoek daarnaar van groot belang zijn (zie ook hoofdstuk 7): het hebben van een visie op de gewenste ontwikkeling, het leiding geven aan het veranderingsproces, en het kunnen omgaan met datafeedback. Hierna werken we elk van deze condities kort nader uit. Aan de uitwerking van de tweede genoemde conditie (leiding geven aan het veranderingsproces, zie 8.3) is een samenvattende beschrijving toegevoegd van een onderzoek dat is uitgevoerd naar een veranderingsproces op een school (het zesde door ons uitgevoerde onderzoek waarvan in dit proefschrift verslag wordt gedaan, zie hoofdstuk 2). Deze beschrijving is toegevoegd omdat deze een aantal van de op basis van de literatuur beschreven factoren pregnant illustreert.

8.2 Hebben van een visie op de gewenste ontwikkeling

Een duidelijke visie op de gewenste ontwikkeling is, afgaande op een aantal onderzoeken en auteurs, een belangrijke vereiste voor het welslagen van een veranderingsproces. Het invoeren van vernieuwingen in het onderwijs blijkt namelijk complex te zijn en een kwestie van lange adem (Waslander, 2007). Docenten en scholen moeten vaak een grote omschakeling maken en daarbij verschillende aspecten op een goede manier met elkaar in verband brengen. De effecten van nieuwe onderwijsvormen, zoals meer zelfstandig leren, op het leergedrag van leerlingen blijken bijvoorbeeld sterk afhankelijk te zijn van een goede begeleiding door de docent, ingebed in de relatie tussen de docent en de leerlingen. Het overdragen van verantwoordelijkheid voor het leerproces van docenten aan leerlingen, met als doel het stimuleren van actief en zelfstandig leren, moet geleidelijk gebeuren en moet worden geïntegreerd in het onderwijsleerproces als geheel en worden opgenomen in de visie van de school (Teurlings et al., 2006).

Hoe belangrijk het is om veranderingen op een school te baseren op een duidelijke visie en doordacht beleid, blijkt ook uit de pogingen die al in de tweede helft van de 90-er jaren zijn ondernomen om te komen tot meer actief en zelfstandig lerende leerlingen. De daarop gerichte maatregelen bleven nogal eens beperkt tot organisatorische veranderingen (bijvoorbeeld door de invoering van keuzewerktijduren, dat zijn lesuren waarin de leerlingen een zekere mate van keuzevrijheid hebben ten aanzien van de inhoud en/of aard van hun leeractiviteiten), terwijl een duidelijke visie op onderwijs en aandacht voor de benodigde verandering in de schoolcultuur en voor het veranderingsproces zelf ontbraken (zie onder andere Tweede Fase Adviespunt, 2001; Sol & Stokking, 2008a).
Uit onderzoek van Anderman, Maehr en Midgley (1999) op meerdere scholen blijkt dat een leeromgeving waarin leerlingen keuzen kunnen maken en kunnen werken aan betekenisvolle taken kan bijdragen aan hun motivatie voor school en leren. Anderman en collega's vonden in dit verband relevante verschillen tussen de onderwijsvisies van de schoolleiders en constateerden dat de school in hun onderzoek die het duidelijkste een dergelijke leeromgeving wist te realiseren zich vooral kenmerkte doordat hierover in het team voortdurend onderling werd gesproken en sprake was van de ontwikkeling van een gemeenschappelijke visie. Dit onderstreept het belang van een duidelijke, gezamenlijke visie op de gewenste verandering.
Het is niet eenvoudig voor schoolleiders om een eenduidige visie te formuleren en doordacht beleid te voeren. Het combineren van de soms heel diverse inspiratiebronnen die zich van buitenaf aandienen, kan leiden tot een onderwijsconcept dat niet heel eenduidig is en daardoor mogelijk ook minder wervend en sturend (Sol & Stokking, 2010; zie hoofdstuk 7). In wetenschappelijke publicaties worden niet zelden doelen en principes aan elkaar gekoppeld of tegenover elkaar geplaatst die mogelijk beter kunnen worden gezien als afzonderlijk te maken en op verschillende manieren te combineren keuzes (vergelijk Bolhuis & Voeten, 2004; De Kock, Sleegers & Voeten, 2004). Wetenschappelijke discussies zijn op deze manier niet altijd behulpzaam bij het ontwikkelen van een visie en kunnen leiden tot verwarring of uitmonden in karikaturen. Het is een uitdaging voor schoolleiders om verschillende vernieuwingen samen te brengen in een coherente visie en leiding te geven aan de uitwerking daarvan. Daarvoor is ook kennis nodig over innovatieprocessen (Runhaar et al., 2009; Waslander, 2007).

8.3 Leiding geven aan het veranderingsproces

Aanwijzingen vanuit de literatuur

Veranderingsprocessen in organisaties: twee benaderingen
Voor organisaties zijn er regelmatig aanleidingen om een veranderingsproces in gang te zetten (Bennebroek Gravenhorst, 2002). Vaak gaat het om nieuwe of hogere eisen vanuit de omgeving, zoals: efficiënt werken, kwaliteit leveren, maatwerk realiseren.

Bennebroek Gravenhorst onderscheidt bij veranderingsprocessen in organisaties twee benaderingen: ontwerpgericht en ontwikkelingsgericht. Uit zijn onderzoek blijkt dat managers vaak neigen naar een ontwerpbenadering, waarbij zij de bedenkers zijn en medewerkers de uitvoerders. Die benadering leidt vaak tot problemen. Veranderen wordt dan te uitsluitend gekoppeld aan sturen en niet verbonden met leren. Een samenhangende, ontwikkelingsgerichte benadering van veranderen, sturen en leren blijkt betere resultaten op te leveren.

Managers die een veranderingsproces benaderen als een gemeenschappelijk proces van leren en ontwikkeling zorgen ervoor zoveel mogelijk alle medewerkers bij de gewenste verandering te betrekken, hen mede probleemeigenaar te laten zijn (zodat een gedeeld gevoel van urgentie kan ontstaan), en een duidelijk doel en een duidelijke weg naar dat doel uit te stippelen. Tijdens het veranderingsproces wordt daarbij op meer manieren geleerd: door het opdoen van ervaring, het conceptualiseren van de verandering, en de onderlinge dialoog en samenwerking.

Voor het verloop en de opbrengst van het proces is het daarbij volgens Bennebroek Gravenhorst heel belangrijk welke verwachtingen de betrokkenen ervan hebben. Deze dienen positief te zijn en tegelijk realistisch (met vertrouwen in de haalbaarheid). Ook de volgorde en samenhang van interventies en activiteiten is van belang, waarbij bijvoorbeeld acute problemen eerst worden opgelost, voordat verdere ontwikkeling kan plaatsvinden. Een dergelijke benadering van verandering in organisaties vraagt van managers specifieke competenties: creëren en managen van cultuur, nemen van besluiten, vormen van coalities, ondernemen van concrete acties, vasthouden van het momentum en de inzet van de betrokkenen, en bewaken van de draagkracht van de organisatie (waken tegen overvragen).

Managers kijken echter vaak positiever en optimistischer tegen de verandering aan dan medewerkers. Door binnen de organisatie systematisch gegevens te verzamelen en de uitkomsten terug te koppelen, kunnen alle partijen kennis nemen van elkaars perspectieven. Dergelijke feedback is dan geen sturingsmiddel voor het management maar een bijdrage aan een gezamenlijk leerproces (Bennebroek Gravenhorst, 2006). Alle betrokkenen, management, middenkader en medewerkers (ook onderling, vooral als sprake is van verschillende afdelingen of groepen binnen de organisatie) kunnen op basis van de gegevens met elkaar in gesprek gaan, besluiten tot vervolgacties, en deze ook uitvoeren.

Deze twee benaderingen van veranderingsprocessen zijn ook te herkennen in onderwijsorganisaties. Vernieuwingsprocessen in scholen kunnen worden beschreven in termen van invoering van een nieuw ontwerp, bijvoorbeeld een onderwijsconcept. Een school die op deze wijze wil vernieuwen, stelt zichzelf de taak het betreffende concept te implementeren. Vernieuwingsprocessen kunnen ook worden beschreven in termen van ontwikkeling: schoolontwikkeling, professionele ontwikkeling. Dan gaat het om ‘samen leren’.

Introduceren van een ontwerp dat moet worden geïmplementeerd, kan worden gezien als een interventie die erop is gericht het veranderingsproces te sturen, Imants (2003)

is, net als Bennebroek Gravenhorst (zie hiervoor), van mening dat een dergelijke vorm van sturing weinig bijdraagt aan reflectie en leren. Dat is volgens hem wel het geval bij sturing door middel van de interventie feedback, waarbij gegevens over het verloop en het (voorlopige) resultaat van het vernieuwingsproces worden verzameld en terug-gekoppeld. Op die manier kunnen schoolontwikkeling en professionele ontwikkeling van docenten goed samengaan en kunnen scholen voor docenten een goede leer-omgeving zijn. Als condities daarvoor noemt Imants: samenwerking, ruimte voor het maken van eigen keuzes, duidelijke doelen, regelmatige feedback, gedeelde normen en waarden, gezamenlijk focus (op het leren van de leerlingen), reflectie, en dialoog (interactie en gedeelde betekenisgeving). In zulke scholen vormen de docenten een professionele gemeenschap en vormt de school een lerende organisatie, waarin verschillende interpretaties van de verkregen feedback worden uitgewisseld en productief gemaakt.
Ook in scholen die specifiek werken aan het bevorderen van meer zelfstandig leren van leerlingen, kan het vernieuwingsproces op beide geschetste manieren verlopen. Solzbacher (2006) deed onderzoek op een aantal scholen die met genoemde vernieuwing bezig zijn en concludeerde dat deze in twee groepen konden worden verdeeld. In de ene groep werkte men expliciet en doelgericht aan het aanleren van leercompetenties (tijdens daarvoor bestemde uren in het rooster). In de andere groep verliep de vernieuwing impliciet en sprak men over een verandering van leercultuur die bijdraagt aan het beoogde zelfstandig werken. In beide groepen ervoer men dat voor het bewerkstelligen van verandering een bepaalde strategie en coördinatie nodig zijn, alsmede voldoende geschikt materiaal, samenwerking in het docententeam, en ontwikkeling bij docenten van een breed didactisch repertoire.
Beide genoemde hoofdvormen van een vernieuwingsproces, ontwerpgericht en ontwikkelingsgericht, worden hieronder nader verkend, alsmede het belang van percepties, opvattingen en ervaringen van docenten in de context van vernieuwing.

Vernieuwen als doelgericht, planmatig en gestuurd implementeren van verandering
Een duidelijke beschrijving van een vernieuwingsproces via het implementeren van een ontwerp is te vinden bij Spillane, Reiser en Reimer (2002). Zij concluderen dat docenten de vernieuwing moeten kunnen plaatsen en benoemen en daarmee betekenis geven, dat verandering van gedrag ook vraagt om verandering van cognities (kennis en opvattingen: een beeld hebben van waar het naar toe moet en de overtuiging hebben dat het zinvol en kansrijk is), en dat omgekeerd ander gedrag ook verandering van cognities teweeg brengt (eigen ervaringen dragen bij aan begrip van wat met de vernieuwing wordt bedoeld). Zij beschrijven een aantal factoren en succesvoorwaarden die bij een vernieuwingsproces een rol kunnen spelen, ten aanzien van de vernieuwing, de docenten, de werkomgeving en het gevoerde beleid. We vatten deze kort samen.
De *vernieuwing* moet een helder doel hebben en een valide (op kennis gebaseerde) strategie, en de verantwoordelijkheden moeten duidelijk zijn. Dit neemt niet weg dat docenten aan de vernieuwing een eigen invulling moeten kunnen geven. Daartoe moet

de vernieuwing niet alleen als concrete maatregel zijn beschreven maar ook in termen van de achterliggende rationale, op een iets abstracter niveau. Juist doelen die iets minder concreet zijn omschreven en meer in termen van een streefrichting, laten ruimte voor eigen invulling en kunnen ook op langere termijn motiverend en richtinggevend zijn.

Bij *docenten* spelen veel cognities een rol: verwachtingen, opvattingen, overtuigingen, voorspellingen, verklaringen, mentale modellen, begrippen en ambiguïteiten. Die kunnen ertoe leiden dat een vernieuwing die te weinig afwijkt van het bestaande oppervlakkig wordt waargenomen en dat primair de overeenkomsten worden gezien ("dat doe ik al"). Verder spelen waarden en emoties een belangrijke rol, in relatie tot het zelfbeeld ("heb ik het dan tot dusver altijd verkeerd gedaan?") en tot de neiging van veel docenten om meningsverschillen en conflicten uit de weg te gaan (Hargreaves, 2001; Lasky, 2005; Runhaar et al., 2009).

In *de werkomgeving* zijn de geldende normen, regels, definities en visies van belang (dat wat wordt beweerd en voor waar aangenomen). Docenten kunnen verschillen wat betreft de kring waarin zij zich bewegen. De ene docent krijgt hierdoor meer impulsen en heeft meer potentiële invloed op anderen, dan de andere. Ook de formele organisatiestructuur kan hier een rol spelen, bijvoorbeeld door het instellen van subteams op basis van groepen leerlingen (in plaats van schoolvak).

Tenslotte is er *het beleid van de school*. De visie op de gewenste vernieuwing moet ruimte laten maar niet te abstract blijven want dan kan iedereen er iets anders onder verstaan en dat is meestal niet wenselijk. Daarnaast heeft de schoolleiding tot taak voldoende leermogelijkheden te creëren voor de docenten (tijd voor reflectie en uitwisseling). Idealiter wordt het vernieuwingsproces voortdurend bijgestuurd (door tussentijdse feedback) en wordt de lat geleidelijker hoger gelegd.

De conclusies van Spillane et al. (2002) sluiten aan bij die van andere auteurs. Bergen en Van Veen (2004) noemen als kenmerken van een "krachtige leeromgeving voor docenten": ruimte en aandacht voor eigen praktijkervaringen, samenwerking, experimenteren, steun, feedback en voldoende tijd. In veel vernieuwingstrajecten wordt hieraan niet voldaan (door ambitieuze en vage doelen, gelijktijdige organisatorische veranderingen, weinig ruimte tot eigen inbreng, chronisch gebrek aan tijd). Bergen en Van Veen stellen dat vaak lijkt te worden onderschat welke implicaties vernieuwingen met zich meebrengen en dat invoering docenten vaak voor problemen stelt waarover onvoldoende is nagedacht. Guskey (1995, 2002b, 2003) noemt soortgelijke kenmerken en voorwaarden van "effectieve professionele ontwikkeling van docenten", en noemt ook het afstemmen tussen de ontwikkeling van docenten en van de school als organisatie, het afstemmen van de specifieke vernieuwing op andere lopende maatregelen en veranderingen en het zorgen voor een balans tussen druk en steun.

Spillane, Halverson en Diamond (2004) en Runhaar et al. (2009) werken het belang van leiding geven aan veranderingsprocessen nader uit, waarbij Spillane et al. (2004) laten zien dat niet alleen de schoolleiding zelf maar ook anderen in de school leiderschapsfuncties kunnen vervullen.

Vernieuwen als een gezamenlijk proces van leren en ontwikkeling

Voor het ontwikkelen van een duidelijke visie op het te realiseren onderwijs benadrukt Adriaenssen (2005) het belang van open relaties tussen schoolleiding, docenten en leerlingen. Als bijvoorbeeld wordt gekozen voor het centraal stellen van meer actief en zelfstandig leren door leerlingen, waarbij de docent geleidelijk minder moet gaan sturen en meer begeleiden, is het heel belangrijk dat dit proces kan verlopen tegen de achtergrond van gedeelde opvattingen over leren en gedeelde normen voor competent gedrag. Voorkomen moet worden dat de betrokkenen op de school blijven steken of terecht komen in wij-zij denken. Echte verandering vraagt volgens Adriaenssen om een school waarin alle geledingen met elkaar in gesprek zijn en meedenken over en betekenis geven aan de koers, waarin veranderingen worden gebaseerd op herkenbare en gedeelde uitgangspunten en een bepaalde kennisbasis en deze in de praktijk worden getoetst, en waarin successen en knelpunten zichtbaar zijn en ook worden besproken. Schoolleiding, docenten en leerlingen ontwikkelen zich in dat geval parallel, in onderlinge interactie, in de context van de ontwikkeling van de school als geheel. Ook Scholtes (2004) wijst op het belang van het voortdurend steunen en volgen van en communiceren over het proces en het vermijden van wrijvingen. Daarnaast noemt hij als voorwaarden dat wordt voorzien in een doordacht inhoudelijk concept dat houvast biedt en dat sprake is voldoende faciliteiten en van concreet ontwikkelwerk. Vanuit een iets algemenere optiek kan worden aangenomen dat veranderingsprocessen bij docenten en in scholen via processen van communicatie en betekenisgeving elkaar beïnvloeden (vergelijk Crossan, Lane & White, 1999; Opfer & Pedder, 2011; Weick, 1995).

Ook specifiek ten aanzien van vernieuwingen gericht op meer zelfstandig leren zijn dergelijke ervaringen en bevindingen gerapporteerd. Butler, Lauscher, Jarvis-Selinger en Beckingham (2004) beschrijven de invoering van zo'n vernieuwing in een school in termen van samenwerking, tussen docenten onderling, tussen docenten en leerlingen, en tussen docenten en onderzoekers die het proces ondersteunen. Leraren en leerlingen blijken daarbij soortgelijke leerprocessen te kunnen doormaken en als het ware parallel te leren. Belangrijke factoren waren dat leraren gelegenheid hebben voor tussentijdse reflectie, een beeld hebben van waar het naar toe moet, daar zelf mede invulling aan kunnen geven, kunnen zien dat de beoogde situatie kan worden gerealiseerd (zelfstandig lerende leerlingen), en leereffecten bij de leerlingen kunnen zien. Swaffield en Macbeath (2006) rapporteren over meerdere scholen die werkten aan meer zelfstandig leren en daarbij het leren van leerlingen en van docenten en de ontwikkeling van de school met elkaar verbonden. Daarbij vonden zij een aantal voorbeelden van koppeling van structurele en culturele factoren: op vergaderingen de vernieuwing tot een vast onderwerp van de agenda maken; stimuleren dat docenten er onderling over praten en ideeën en ervaringen uitwisselen; integreren van de vernieuwing in het beleid van de school; inbouwen van regelmatige evaluatie van het proces; respecteren van de professionaliteit van docenten, aansluiten bij hun praktijk en ontwikkeling, hen stap voor stap voorbij hun comfort zone laten gaan; tolereren van

verscheidenheid in de manier waarop en het tempo waarin de vernieuwing wordt vormgegeven; en zorgen voor een voldoende stabiele context binnen de school in termen van cultuur, waarden en relaties als voedingsbodem voor de veranderingen.
In de benadering van vernieuwen als gezamenlijk proces van leren en ontwikkeling zijn communicatie en afstemming belangrijke voorwaarden om de verandering vorm te kunnen geven. Weick (1995) stelt dat voor betekenisgeving binnen een organisatie intersubjectieve contacten nodig zijn. Als de stroom van communicatie en afstemming verward raakt, begint de organisatie te disfunctioneren, gaan mensen zaken versimpelen en vallen ze terug op hun routines. In een school kan dit bijvoorbeeld gaan spelen op het moment dat bij het invoeren van een vernieuwing niet duidelijk is wat van docenten wordt verwacht (vergelijk Kelchtermans, 2007). Een gemeenschappelijk kader kan alleen ontstaan als opvattingen, ervaringen en interpretaties worden uitgewisseld en daarbij (deels wellicht onbewuste) kennis, verwachtingen, doelen en keuzes in relatie tot het eigen handelen en concrete situaties worden geëxpliciteerd en gearticuleeerd (Homan, 2005; Weick & Sutcliffe, 2005).

Percepties, opvattingen en ervaringen van docenten in de context van vernieuwing
In een context van een traject van vernieuwing op een school en de processen van verandering en communicatie die daardoor worden teweeggebracht, zijn de docenten de centrale actoren. Volgens Adriaenssen (2005) staan docenten vaak welwillend tegenover door de schoolleiding voorgestelde vernieuwingen, maar passen ze hun gedrag niet aan. Dit kan volgens hem worden verklaard doordat veel docenten hun vakkennis graag willen blijven uitdragen, en moeite hebben leerlingen los te laten en het leerproces meer aan hen over te laten. Volgens hem speelt hierbij mee dat docenten verandermoe zijn, en dat zij zich afvragen of leerlingen zelfstandig wel voldoende zullen leren.
Hargreaves (2004) constateert dat leraren vaak negatief staan tegenover veranderingen die zij als opgelegd ervaren. Hij ontdekte echter dat ook veel veranderingen die docenten zien als door henzelf geïnitieerd hun oorsprong hebben in verplichtende beleidskaders. Doorslaggevend voor de beleving bij de docent van de verandering lijkt te zijn of de docent er als professional bij wordt betrokken en de vernieuwing duidelijk is gericht op het belang van de leerlingen. Ook Schmidt en Datnow (2005) wijzen op het belang bij docenten van een "*sense of ownership*". Volgens hen hebben docenten behoefte aan steun en begeleiding bij het geven van betekenis aan een vernieuwing en moeten schoolleiders veel investeren om ervoor te zorgen dat docenten weten wat de vernieuwing inhoudt en de middelen hebben om deze gestalte te geven.
De ervaringen die docenten met een andere aanpak opdoen, vooral waar het gaat om de effecten bij de leerlingen, dragen op hun beurt bij aan veranderingen in opvattingen (Fishman, Marx, Best & Tal, 2003; Guskey, 2002a; Opfer et al., 2011). Leren van docenten en leren van leerlingen kunnen op die manier elkaar beïnvloeden en in beide gevallen kunnen dezelfde factoren en principes van belang zijn: een duidelijk doel voor ogen hebben, voldoende eigen inbreng hebben en zelf keuzes kunnen maken, kunnen

rekenen op voldoende steun, en succeservaringen kunnen opdoen (vergelijk Bergen & Van Veen, 2004; Swaffield & MacBeath, 2006; Teurlings, Van Wolput & Vermeulen, 2006).

Clusters van voorwaarden voor succesvol vernieuwen in een school
De besproken literatuur wijst op het belang van de volgende vier groepen factoren.
De vernieuwing biedt een duidelijke visie op en beeld van het gewenste onderwijs en de weg daarnaar toe. De visie is voldoende concreet maar tegelijk voldoende inspirerend en open voor nadere invulling. De visie is voldoende nieuw om uit te dagen en voldoende herkenbaar om haalbaar te achten. De vernieuwing is afgebakend maar wel verbonden met wat verder in de school speelt. De vernieuwing wekt de verwachting van positieve effecten bij leerlingen en kan in stappen worden uitgeprobeerd.
Het veranderingstraject wordt duidelijk georganiseerd in termen van taken, prioriteiten en fasering en met behulp van faciliteiten en materialen. Er wordt gezorgd voor goede communicatie, steun en coördinatie. Het proces en de effecten worden tussentijds gevolgd, zichtbaar gemaakt, gedeeld en geëvalueerd. Waar nodig vinden bijstellingen plaats.
De schoolleiding en de docenten trekken samen op. De kaders zijn duidelijk maar vragen wel om nadere ontwikkeling en maken deze ook mogelijk. De docenten worden vanaf het begin bij het proces betrokken en zijn daarvan mede-eigenaar. De bij docenten aanwezige expertise wordt erkend en benut en verder ontwikkeld, in relatie tot de ontwikkeling van de school. Er is ruimte voor verschillende uitwerkingen, mits deze bijdragen aan het gestelde doel.
De dialoog tussen alle betrokkenen staat centraal en wordt actief opgezocht en gevoerd. Opvattingen, kennis, ervaringen en normen worden besproken en gedeeld. Op basis van open interpersoonlijke relaties wordt gewerkt aan reflectie, gedeelde betekenisgeving, samenwerking en teamvorming.

Bevindingen en ervaringen in een onderzoek naar een veranderingsproces op een school

De schoolleiding van de afdeling havo/vwo (70 docenten, 600 leerlingen) van een brede scholengemeenschap in een middelgrote gemeente stelde in 2006 aan het docententeam voor om keuzewerktijduren te gaan invoeren. Om inzicht te krijgen in het verloop van het veranderingsproces bij docenten en leerlingen en aanwijzingen voor eventuele bijsturing en de aanpak van toekomstige didactische en organisatorische veranderingen, vroeg de schoolleiding een onderzoek aan. Voorafgaand aan (februari-april 2007) en tijdens het onderzoek (mei 2007-februari 2008) is een aantal gesprekken met de schoolleiding gevoerd over de aanleiding, opzet en bedoeling van de verandering en de tussentijdse uitkomsten van oriënterende gesprekken met individuele docenten en groepjes leerlingen, interviews met docenten en schriftelijke vragenlijsten

voor docenten en voor leerlingen. Het onderzoek leidde tot een eindrapport (Sol & Stokking, 2008a), waaraan we hier het volgende ontlenen.

De schoolleiding

De school kenmerkte zich volgens de schoolleiding altijd door goede verhoudingen tussen docenten en leerlingen, waardoor de noodzaak tot vernieuwen niet sterk werd gevoeld, en stond bekend om haar leerlinggerichte klimaat en haar zorg voor leerlingen met problemen. De school slaagde er echter geleidelijk minder goed in om alles uit de leerlingen te halen. Vanaf 2004 is gediscussieerd over onderwijsvernieuwing, vooral over het bieden van meer keuzemogelijkheden aan leerlingen. Een rapport van de onderwijsinspectie bood een aanleiding en de landelijke invoering van de Vernieuwde Onderbouw (in 2006) een concreet moment om de vernieuwing gestalte te geven.

De schoolleiding stelde voor te beginnen met invoering van keuzewerktijduren en de vorming van vaksectieoverstijgende subteams in de onderbouw en kreeg hier de docentenvergadering in mee. In het lesrooster werden acht wekelijkse lesuren apart bestemd: 1 gewoon mentoruur, 1 à 2 uur bij de mentor werken aan een vakoverstijgend project (vier projecten per jaar), en 5 à 6 uur keuzewerktijd onder begeleiding van een vakdocent, waarbij de leerlingen zelf vooraf kiezen in welk uur ze willen werken aan welk vak. De schoolleiding verwachtte van de docenten dat zij de keuzewerktijduren zelf zouden gaan inrichten en daartoe (meer) met elkaar zouden samenwerken (in de nieuwe subteams).

De schoolleiding zag de invoering van keuzewerktijduren als katalysator voor een cultuuromslag waarbij meer doelen in samenhang worden nagestreefd: beter volgen van het leerproces van de leerlingen, meer inspelen op verschillen tussen leerlingen, aanbieden van mogelijkheden tot herhaling en verdieping en van vakoverstijgende projecten, leerlingen aanzetten tot meer zelfstandig en zelfverantwoordelijk leren, en geven van feedback aan leerlingen op hun leren en hun keuzen. Het hoofddoel was dat docenten meer uit elke individuele leerling zouden halen.

Deze opzet bracht volgens de schoolleiding tijdens het eerste jaar (2006-2007) diverse processen op gang. Het aantal vakspecifieke uren nam af waardoor docenten gingen nadenken over wat hen in de nieuw ingestelde uren te doen stond. Docenten bleken verschillende opvattingen te hebben over en verschillend zicht op het leren van de leerlingen en ook verschillend te staan tegenover de beoogde vernieuwing. De organisatie van de keuzewerktijd (leerlingen worden op basis van door hen gemaakte keuzes ondergebracht bij vakdocenten) bleek niet eenvoudig. Verwachtingen en behoeften van docenten en leerlingen over de invulling bleken uiteen te lopen. Voor een aantal docenten was het wennen meer naar de leerprocessen van leerlingen te kijken en daarop in te spelen, en voor leerlingen was het wennen zelf keuzen te maken: welke leeractiviteiten bij welke vakken op welk moment.

In reactie op de uitkomsten van de oriënterende gesprekken met docenten (zie onder) concludeerde de schoolleiding dat de docenten kennelijk, als gevolg van impliciete en

niet-getoetste opvattingen, worden weerhouden leerlingen te motiveren voor meer zelfstandig leren. Dat docenten opmerken dat ze door de invoering van keuze-werktijduren "in minder tijd dezelfde stof moeten doen" wijst op hun subjectieve ervaring, maar het gaat bij de vernieuwing nu juist om het differentiëren tussen leerlingen die niet allemaal evenveel tijd nodig hebben om de doelen te realiseren. Dat docenten opmerken dat keuzewerktijd "niet geschikt is voor alle vakken en voor alle leerlingen" lijkt een overtuiging, waarbij de vraag is in hoeverre dit echt zo is. De oude situatie op de school was moeizaam en er moest iets veranderen. Wat docenten doen, staat niet los van hun opvattingen en de vraag is hoe mensen kunnen worden geprikkeld hun opvattingen te veranderen.
De schoolleiding constateerde verder dat het vernieuwingsproces nog vooral bij de schoolleiding blijft liggen en dat de docenten zich minder of niet verantwoordelijk voelen. Tijdens vergaderingen komen docenten met weinig ideeën, maar de schoolleiding wil de verantwoordelijkheid wel bij de docenten leggen. De school-leiding ziet het realiseren van de vernieuwing (samenwerken, ontwikkelen en invoeren) als taak van de docenten. In reactie op de suggestie van de onderzoeker dat de docenten misschien te weinig urgentie voelen en de vernieuwing zodanig is ingezet dat docenten hun oude patronen kunnen handhaven, gaf de schoolleiding aan graag zorgvuldig te willen optreden. Het gaat om de meest fundamentele verandering van de organisatie van het onderwijs op de school sinds de Mammoetwet, in een cultuur die niet alleen een zekere vrijblijvendheid kent maar ook veel politiek-ambtelijke trekken heeft.
Naar aanleiding van de uitkomsten van de interviews met docenten (zie onder) gaf de schoolleiding aan de grote variatie in de antwoorden opvallend te vinden. Zij interpreteerde de uitkomsten zo dat de docenten draagvlak geven voor het bieden van keuzemogelijkheden aan leerlingen maar dat nog wel moet worden gewerkt aan randvoorwaarden en het aangeven van wat nu precies in de keuzewerktijduren moet gebeuren. In reactie op de suggestie van de onderzoeker dat het neerzetten van een duidelijke visie belangrijk is zodat docenten zich daarop kunnen richten, gaf de schoolleiding aan te worstelen met de randvoorwaarden, waaronder de beperkingen van het gebouw, en dat er in het onderbouwteam wel over de keuzewerktijd is gesproken maar dat iedereen zijn eigen ding doet. De schoolleiding wilde niet alles van tevoren dichttimmeren.

De docenten
In de oriënterende gesprekken met docenten (n=5) komt naar voren dat zij als hoofddoel van de vernieuwing zien dat wordt bevorderd dat leerlingen meer leren en dat meer uit leerlingen wordt gehaald dan nu gebeurt, en dat recht wordt gedaan aan verschillen tussen leerlingen. Zij plaatsen dit doel tegen de achtergrond van een schoolcultuur met een goede sfeer en een ontspannen omgang tussen leerlingen en docenten. Belangrijke voorwaarden, die volgens hen door docenten moeten worden vervuld, betreffen het ontwikkelen en aanbieden van studieplanners, materialen voor herhaling en verdieping, en vakoverstijgende projecten. Ze zien in dat ze voor dit

ontwikkelwerk, maar ook voor hun nieuwe didactische taken (volgen van en inspelen op individuele leerprocessen) meer moeten gaan samenwerken. Ze geven aan dat de docentenvergadering destijds met het vernieuwingsbesluit akkoord is gegaan onder de voorwaarde dat een aantal condities zou worden vervuld. Een aantal condities is echter nog niet (voldoende) vervuld, zoals geschikte lokalen met voldoende faciliteiten (computers, materialen), meer keuzemogelijkheden voor de leerlingen, betere mogelijkheden om het leerproces van de leerlingen te kunnen volgen, en meer tijd voor de docenten om samen te werken en dingen te ontwikkelen (taakuren, nascholing).
Uit de met docenten gehouden interviews (n=13) blijkt dat alle betreffende docenten meer samenwerking nodig achten maar ervaren daarvoor te weinig tijd te hebben, terwijl de grote verschillen tussen docenten volgens hen juist om meer samenwerking vragen. De samenwerking richt zich nu vooral op het oplossen van (steeds weer nieuwe) praktische problemen. Volgens sommige van de geïnterviewde docenten is het lastig dat veel docenten erg zijn gebrand op hun eigen vak en dat docenten soms nog moeten leren om op dingen te kunnen worden aangesproken zonder zich meteen aangevallen te voelen of het als bemoeizucht te ervaren. Een deel van de docenten vindt dat de docenten nadrukkelijker op hun verantwoordelijkheid moeten worden aangesproken en dat de cultuur van vrijblijvendheid de school nog teveel parten speelt.
De meeste geïnterviewde docenten zijn van mening dat ze in de gewone vaklessen meer zicht hebben op het leren van de leerlingen en daarop beter kunnen inspelen dan in de keuzewerktijduren. De docenten achten de vaklessen meer geschikt om een relatie op te bouwen met de leerlingen, te differentiëren tussen leerlingen en leerlingen te leren meer zelfstandig te werken dan de keuzewerktijduren, en zeer veel meer geschikt om te bevorderen dat leerlingen zoveel mogelijk leren op vakinhoudelijk gebied. De docenten zijn in de vaklesuren erop gericht dat alle leerlingen de stof begrijpen en ze spelen in op verschillen tussen leerlingen door vragen te beantwoorden en waar nodig extra uitleg te geven (klassikaal of tijdens zelfstandig werken). In vergelijking daarmee differentiëren ze veel minder in de opdrachten die ze geven. Ze geven de leerlingen duidelijkheid over wat ze moeten doen en volgen hoe leerlingen zelfstandig werken en geven hen daarbij hulp.
Een aantal van de geïnterviewde docenten is terughoudend in hun verwachtingen over de mate waarin leerlingen meer willen doen dan strikt nodig is en de mate waarin leerlingen kunnen leren meer zelfstandig te leren. Een aantal andere docenten denkt dat leerlingen wel kunnen worden gemotiveerd en kunnen leren hun eigen leerproces meer te sturen, maar ziet (nog) niet goed voor zich hoe dit kan worden gerealiseerd. De docenten denken ook verschillend over wat van hen in de keuzewerktijduren wordt verwacht (extra instructie geven, vragen beantwoorden, iets extra's bieden, begeleiden en feedback geven, zorgen dat de leerlingen werken) en ze verschillen tijdens deze uren in wat ze de leerlingen vragen, toestaan en bieden. De docenten weten veelal van elkaar niet wat ze doen.

De docenten en de leerlingen

Uit de antwoorden van de docenten (n=33) en de leerlingen (n=284) op de eerder genoemde schriftelijke vragenlijsten blijkt dat in hun kijk op de reguliere vaklessen en de daarin gevolgde aanpak, tussen docenten onderling en tussen docenten en leerlingen veel meer overeenstemming bestaat dan in hun kijk op de bedoeling van en de gang van zaken in en rond de keuzewerktijduren. Zowel docenten als leerlingen geven gemiddeld aan dat docenten in de gewone vaklessen ernaar streven dat de leerlingen de stof begrijpen (door vragen stellen, vragen beantwoorden, zo nodig extra uitleg geven) en dat leerlingen veel gelegenheid krijgen om zelfstandig te werken, daartoe duidelijke instructies krijgen en daarbij ook worden begeleid, maar dat leerlingen niet wordt geleerd hoe ze hun werken en leren kunnen plannen en bij het plannen ook weinig worden begeleid. Met de keuzewerktijduren gaan de docenten heel verschillend om, ze werken daarbij weinig samen en ze weten vaak van elkaar niet wat ze doen. De leerlingen maken van de nieuwe keuzevrijheid gebruik en kiezen massaal voor het maken van huiswerk en het leren voor een proefwerk, wat niet direct past bij een aantal van de met de keuzewerktijduren nagestreefde doelstellingen.

Leerlingen en docenten verschillen ook in hun beeld van de zelfstandigheid van de leerlingen. Leerlingen achten zichzelf behoorlijk zelfstandig en bij het inschrijven voor keuzewerktijduren kiezen ze veelal heel bewust, terwijl docenten vinden dat nogal wat leerlingen helemaal niet zo zelfstandig zijn, uit zichzelf niet reflecteren en moeite hebben met plannen. Leerlingen redeneren echter vanuit het perspectief van vrijheid (keuzevrijheid bij inschrijven), zelfstandig werken en huiswerk kunnen maken, docenten redeneren vanuit het perspectief van het door de leerlingen leren van het vak en het vooruit kijken en het ook willen doen van moeilijke taken dat volgens hen daarvoor nodig is.

Terugkijkend

Terugkijkend kan worden vastgesteld dat zowel de schoolleiding als de docenten hebben aangegeven dat sprake was van vrijblijvendheid en dat de leerlingen meer moesten worden geactiveerd en uitgedaagd. Over het probleem en de doelstelling was voldoende consensus.

De schoolleiding zag de invoering van keuzewerktijduren als katalysator van een gewenst proces van verandering, en verwachtte van de docenten dat zij de verantwoordelijkheid daarvoor op zich zouden nemen en dat docenten en leerlingen op elkaar zouden reageren en elkaar zouden meenemen in een gemeenschappelijk leerproces. Docenten en leerlingen zijn inderdaad op de nieuwe situatie gaan reageren. De meeste leerlingen reageerden op dezelfde manier: ze grepen de vrijheid aan om zelf te bepalen waar ze aan werken en daarbij zo nodig hulp van de docent te krijgen, en ze kozen voor het maken van huiswerk en het leren voor een proefwerk. De docenten reageerden meer verschillend, al naar gelang hun opvattingen en mogelijkheden. Docenten en leerlingen reageerden ook op elkaar: leerlingen kozen al dan niet voor een uur bij een docent afhankelijk van hoe die docent met keuzewerktijduren omgaat, en

docenten slaagden er meer of minder in leerlingen te motiveren om zich in te zetten en te leren (net als in de gewone vaklessen). Alle drie partijen, schoolleiding, docenten en leerlingen, merkten op dat sprake is van grote verschillen tussen de docenten. Een aantal individuele docenten is aan de slag gegaan om de aan hen toegewezen keuzewerktijduren nader in te vullen, maar de docenten hebben de hen toegewezen taak niet in gezamenlijkheid op zich genomen.
Een mogelijke interpretatie van de verschillen tussen de boven beschreven overeenstemming over de gewone vaklessen en de verschillen ten aanzien van de keuzewerktijduren is dat de docenten op de school in de loop van de jaren in de vaklessen een bepaalde manier van werken hebben ontwikkeld, en dat zij, nu zij met de keuzewerktijduren plotseling zijn gesteld voor een nieuwe situatie waaraan zij zich moeten aanpassen, daarop onderling heel verschillend reageren, mogelijk door het manifest worden van uiteenlopende persoonlijke opvattingen, aannamen en verwachtingen. Tegenover deze variatie, die ook kan worden geduid als onzekerheid en verwarring, hadden de leerlingen, die in de nieuwe situatie kennelijk heel goed wisten wat ze wilden, min of meer vrij spel.
De concrete invulling en de uitvoering van het veranderingstraject waren door de schoolleiding bij de docenten neergelegd maar beide partijen trokken niet samen op, toegezegde faciliteiten kwamen slechts geleidelijk beschikbaar, en in toewijzing van taken en coördinatie was niet voorzien. De docenten waren wel bij de besluitvorming betrokken geweest maar voelden zich slechts gedeeltelijk mede-eigenaar, hun expertise werd wel erkend en kreeg alle ruimte, maar schoolleiding en docenten bleven elkaar de bal toespelen. De benodigde dialoog kwam niet op gang en dit werd door de docenten gaandeweg ook als groot knelpunt herkend: er is te weinig samenwerking. De docenten stelden dat de daarvoor benodigde tijd ontbrak, en er kwam geen initiatief om deze vicieuze cirkel te doorbreken.

Interpretatie en vervolg
Van de in de literatuur genoemde voorwaarden voor succesvolle verandering was bij het proces op deze school aan enkele voldaan. Er was sprake van een gedeeld belang bij verbetering ten behoeve van de leerlingen. Ook zijn de docenten er duidelijk bij betrokken, is hun expertise erkend en hebben ze veel ruimte gekregen om de vernieuwing zelf in te vullen. Van de andere genoemde voorwaarden was echter een aantal tegelijk niet vervuld, waaronder die van een afgebakende en concreet uitgewerkte vernieuwing, voldoende materialen en faciliteiten, en tijd voor en organisatie van uitwisseling en samenwerking. Daardoor is niet duidelijk welke factor of factoren vooral verantwoordelijk is of zijn geweest voor het niet goed op gang komen van het beoogde proces. Mogelijk schatte de schoolleiding terecht in dat als zij de leiding had genomen, voor de beoogde vernieuwing een reeds concreet uitgewerkte opzet en aanpak had neergelegd en/of eerst een paar jaar had uitgetrokken om de gewenste materialen en voorzieningen te realiseren, dit alleen maar averechts zou hebben uitgepakt.

Het eindrapport over dit onderzoek (Sol & Stokking, 2008a) is door de schoolleiding aan het hele docententeam ter beschikking gesteld en bracht in de school een schokeffect teweeg (april 2008). Dit leidde, mogelijk mede doordat de geesten daarvoor inmiddels rijp waren, tot het op gang komen van overleg en samenwerking tussen de schoolleiding en docenten en in werkgroepen tussen docenten onderling, waardoor de vernieuwing in gezamenlijkheid concreet kon worden uitgewerkt en doorgezet. De in het rapport geformuleerde aanbevelingen staan in figuur 8.1 en een beschrijving van het daarop volgende proces (in 2009 besproken met de schoolleiding) is opgenomen in figuur 8.2.
In oktober 2011 zijn de hiervoor beschreven bevindingen, conclusies en interpretaties, met de aanbevelingen (fig. 8.1) en de beschrijving van het vervolg (fig. 8.2), (nogmaals) voorgelegd aan de tweehoofdige schoolleiding. Deze bevestigde dat zij zich hierin goed kon herkennen en voegde daaraan toe dat de school inmiddels verder op de goede weg was: er is sprake van groeiende consensus onder de collega's over het wat, hoe en waarom van de keuzewerktijduren, en de conclusies en aanbevelingen uit het onderzoek zijn daarbij heel waardevol gebleken (e-mail 28 oktober 2011).
De met het rapport uitgebrachte *feedback* leidde kennelijk tot een proces van voortgezette *dialoog en samenwerking* en vervolgens tot voortgang in het vernieuwingsproces, waarmee het in de literatuur genoemde belang van deze factoren lijkt te worden bevestigd.

8.4 Kunnen omgaan met datafeedback

De eerder genoemde inzichten van Bennebroek Gravenhorst (2002, 2006) ten aanzien van de rol die onderzoekers kunnen spelen bij veranderingsprocessen in organisaties, met name door het geven van feedback, worden bevestigd en aangevuld door literatuur die specifiek gaat over het geven van feedback aan scholen en docenten.
Van Petegem en Vanhoof (2004) en Geijsel en Krüger (2005) gebruiken hiervoor de term datafeedback. Van Petegem et al. houden een pleidooi voor "geïnformeerde schoolontwikkeling" en stellen dat schoolleiders moeten kunnen onderkennen welke gegevens over de eigen school vooral van belang zijn en, om er betekenis aan te kunnen geven, deze ergens mee moeten kunnen vergelijken: met andere scholen, een bepaald criterium, en/of de situatie en ontwikkeling van de eigen school. Volgens deze auteurs is de juiste interpretatie van gegevens niet vanzelfsprekend en moeten schoolleiders leren om daarmee om te gaan en deze doelgericht en geïntegreerd te benutten voor evaluatie en bijsturing.

Naar aanleiding van de uitkomsten van het onderzoek zijn onderstaande aanbevelingen geformuleerd. Bij de keuze voor de inhoud van de aanbevelingen zijn gegevens uit verschillende bronnen betrokken: de situatie en vraagstelling van de school, de praktijkgerichte literatuur over recente vernieuwingen in het voortgezet onderwijs, de wetenschappelijke literatuur over vernieuwingsprocessen in scholen, de resultaten van de rondes van bevraging onder leerlingen en docenten, en de vergelijking van de perspectieven van de betrokkenen. Deze bronnen gaven aanleiding tot de volgende aanbevelingen.

a. Faseer en concretiseer het vernieuwingsproces. Doordenk de onderlinge relaties tussen de doelen en maatregelen, breng op grond daarvan prioriteiten aan, ontwerp een meerjaren-perspectief met concrete acties en voorwaarden, en formuleer criteria voor succes (richting-gevend en motiverend, en geschikt voor tussentijdse evaluaties).
b. Betrek hierbij de docenten en de leerlingen. Geef ze invloed op het beleid (benut hun kennis, waarden en visies) en vraag tegelijkertijd committment (afspraken zijn niet vrijblijvend). Ontwerp samen operationele modellen voor de organisatie en inrichting van de kwt-uren.
c. Organiseer en stimuleer de inhoudelijke dialoog tussen docenten onderling en tussen docenten en schoolleiding over opvattingen en vernieuwingsconcepten (zoals meer zelfstandig (laten) leren, volgen van het leerproces, inspelen op verschillen). Het is belangrijk om met elkaar in gesprek te gaan over betekenissen die men aan begrippen toekent en zo een gemeenschappelijke taal en een gedeeld referentiekader te ontwikkelen.
d. Differentieer de organisatie en inrichting van de kwt-uren indien wenselijk naar groepen leerlingen (met overeenkomstige behoeften en mogelijkheden), soorten uren (bijvoorbeeld herhalen en remediëren, verdiepen, begeleid huiswerk maken, leren voor een proefwerk), en inschrijving (naast losse uren ook voor een aantal uren tegelijk).
e. Houd rekening met de voorwaarden die moeten worden vervuld, rond de inschrijving door leerlingen, de bemensing, beschikbare ruimten, ontwikkeling van materialen, en ondersteuning van docenten bij het gaan realiseren van voor hen nieuwe manieren van werken.
f. Evalueer de noodzaak en de mogelijkheden voor het blijven volgen van de leeractiviteiten en leerresultaten van de leerlingen waar deze meer individueel verschillende keuzes gaan maken.
g. Geef bepaalde docenten zo mogelijk expliciete taken in het proces, zoals het ontwikkelen van materiaal, het verzorgen van bepaalde typen kwt-uren, het volgen van de leerprocessen van de leerlingen. Haal cruciale vernieuwingstaken uit het lopende werk, zodat beide voldoende aandacht kunnen krijgen.
h. Ondersteun en bevorder de gewenste samenwerking tussen docenten. Onderzoek de mogelijkheid hiervoor meer expliciet of op meer effectieve wijze uren in rooster en taakstelling te bestemmen.
i. Betrek de reguliere vaklessen nadrukkelijk in het proces. Bevorder dat de mogelijkheden tot differentiatie die deze lessen bieden (ook al doordat veel docenten tijdens deze lessen de leerlingen al een substantieel deel van de tijd zelfstandig laten werken en leren en hen daarbij willen begeleiden) ten volle worden benut, en dat de vaklessen en de kwt-uren op elkaar worden afgestemd zodat een voor de leerlingen samenhangende leeromgeving ontstaat.
   Koppel de uitkomsten van dit onderzoek terug naar de docenten (en de leerlingen) en breng ze in bespreking, benut ze bij de ontwikkeling van de hier bedoelde koers.
j. Zet een en ander in het teken van een cultuuromslag op de school, waarin heldere afspraken worden gemaakt en taken en verantwoordelijkheden worden toegewezen, die kunnen rekenen op draagvlak doordat ze een duidelijk perspectief schetsen (waar willen we met de school over vijf jaar staan) en tegelijk realistisch en concreet zijn, en waaraan alle betrokkenen elkaar kunnen en zullen gaan houden.

Figuur 8.1: Aanbevelingen aan de school in het eindrapport (Sol & Stokking, 2008a)

Het rapport werd eind maart 2008 verstuurd aan de schoolleiding, die de conclusies en aanbevelingen direct doorstuurde aan alle docenten en het volledige rapport op het intranet van de school beschikbaar stelde. Een tumultueuze bijeenkomst tussen schoolleiding en docententeam over het vernieuwingsproces leidde tot een intensief gezamenlijk traject van aanpassing, aan de hand van de aanbevelingen in het rapport en de inbreng van de docenten. De onderzoekers en de schoolleiding hebben drie maanden later deze periode geëvalueerd en een jaar later nogmaals teruggekeken. We vatten de conclusies daarvan samen.

Volgens de schoolleiding heeft het onderzoek op drie manieren bijgedragen aan de doorbraak in het proces en de daarop volgende aanpak. Ten eerste werd door het onderzoek voor de docenten zichtbaar dat de schoolleiding alle ruimte wilde geven aan docenten om hun opvattingen, ervaringen en ideeën in openheid (tegenover de onderzoekers) naar voren te brengen. Docenten brachten expliciet naar voren dat zij dit zeer op prijs stelden. Ten tweede is een aantal docenten door hun deelname aan het onderzoek en de herhaalde bevraging en terugrapportage aan het denken gezet en is de discussie in het team hierdoor gestimuleerd. Ten derde heeft de inhoud van het onderzoeksrapport bij een aantal docenten een schrikeffect teweeg gebracht en tot de conclusie geleid dat er iets moest gebeuren, en dat het zaak was er, in het belang van de school, samen de schouders onder te zetten.

In overleg tussen schoolleiding en docenten is een aantal werkgroepen geformeerd en, met behulp van draaiboeken, aan de slag gegaan, elk met een eigen onderdeel van het vernieuwingsproces en met de opdracht te komen met concrete voorstellen tot aanpassing en verbetering. De door de onderzoekers geformuleerde aanbevelingen zijn hierbij gebruikt als een checklist: er werd een gezamenlijke koers afgesproken: verhoging van de kwaliteit van de keuze-uren en betere afstemming op de reguliere vaklesuren; het proces werd in onderdelen uiteengelegd; de vernieuwing werd geconcretiseerd, inclusief bijbehorende organisatorische, logistieke en technische aspecten; groepen docenten kregen de ruimte om een en ander verder uit te werken; de schoolleiding voorzag in faciliterende maatregelen; en de opbrengsten van de werkgroepen zijn aan de anderen voorgelegd en door hen geaccepteerd.

Inhoudelijk was sprake van versterking van de functie van de mentoren bij het begeleiden van de leerlingen van het consequent intekenen voor de keuze-uren en het monitoren van hun keuzes en studiegedrag; van concentratie van de keuze-uren per leerjaar en per vak bij één docent, die tijd heeft voor het ontwikkelen van opdrachten en kan optreden als herkenbaar aanspreekpunt; van afstemming tussen de docenten in de vrijheid die leerlingen hebben bij de precieze invulling van hun keuze-uur; van expliciete bestemming van keuze-uren voor verschillende functies (zelfstudie, vakondersteuning, verdieping); en van uitbreiding van het aantal beschikbare projecten, waarmee leerlingen tijdens de keuzeuren diverse vaardigheden kunnen ontwikkelen.

Al met al ging iedereen zich intensief bezighouden met hoe op de school het onderwijs wordt ingericht en legde de gevolgde aanpak, met een aantal werkgroepen, het proces dicht bij de docenten. Er is veel meer interactie ontstaan, tussen schoolleiding en docenten (in de werkgroepen), tussen docenten en leerlingen (leerlingen durven nu meer te vragen om tijdens een keuze-uur ergens aan te mogen werken en dit wordt ook vaker toegestaan), en tussen schoolleiding en leerlingen (via regelmatige gesprekken waarin leerlingen veel vertellen over hoe het gaat).

Een jaar later (voorjaar 2009) kan worden geconstateerd dat het onderzoek de school, de schoolleiding en de docenten een spiegel heeft voorgehouden, en dat flinke vorderingen zijn gemaakt. In het team bestaat veel consensus over de kern van de vernieuwing. Het werken met keuze-uren is uitgebreid van het eerste tot en met het vierde leerjaar, waarbij docenten expliciet kunnen kiezen of zij studiebegeleiding of verdieping aanbieden. Het aanbod van projecten is sterk toegenomen en betreft nu alle vakken.

De communicatie tussen de docenten die keuze-uren begeleiden en de overige vakdocenten is nog wel een aandachtspunt. Het onderwijs in de reguliere lessen is vaak nog behoorlijk traditioneel, een deel van de docenten vindt de vernieuwing niet nodig en teveel gedoe.

Figuur 8.2: Het proces op de school na ontvangst van het eindrapport

Ook Geijsel en Krüger (2005) zien het benutten van datafeedback voor leren en ontwikkeling als niet vanzelfsprekend. Zij voegen toe dat onderzoekers zich moeten richten op gegevens die de school belangrijk vindt, maar dat schoolleiders daar zelf lang niet altijd een duidelijke visie op hebben. Verder benoemen zij dezelfde factoren als cruciaal voor succes als eerder genoemd: een gevoel van eigenaarschap, inbedding in bestaande procedures van de school, voeren van een dialoog die bijdraagt aan betekenisverlening en gedeelde interpretaties en waarderingen, tolereren van en omgaan met verschillen in interpretaties en perspectieven, en gebruiken van datafeedback als leermiddel in plaats van als middel voor sturing en beoordeling (zoals het vaak hardnekkig wordt gezien).

Reed en Stoll (2000) beklemtonen eveneens dat het realiseren van verbeteringen in scholen vraagt om leren, in relatie tot het ontwikkelen van en het voeren van de dialoog over gemeenschappelijke doelen en over adequate aanpakken om die doelen te bereiken. Ze beschrijven vier processen die daarbij een belangrijke rol spelen: actief monitoren van de context waarin de school zich bevindt en op basis daarvan formuleren van doelen; strategisch denken en handelen op basis van kennis (benutten van in de organisatie beschikbare kennis en ervaring en aanbrengen van coherentie); ontwikkelen van een op leren gerichte cultuur (gebaseerd op respect, collegialiteit, openheid, bereidheid tot risico's, onderlinge steun, humor, vieren van successen), in plaats van situaties en processen die leren tegen gaan (zoals hokjesgeest, afweer, neiging tot conformeren, zwarte pieten, ontkenning, vaagheid); en denken in termen van het geheel en blijven zien van de samenhang. Reed en Stoll maken duidelijk dat het benutten van feedback een interactief en iteratief proces is dat op zich moet worden geleerd, en dat scholen zelfs expliciet moeten stilstaan bij dat leerproces op zich. Daarnaast moet natuurlijk de feedback zelf ook aan bepaalde eisen voldoen: niet te omvangrijk en te gedetailleerd, helder (voor de betrokkenen verhelderend), relevant (gericht op de kerntaken en uitdagend), inpasbaar (in bestaand denken en handelen), en goed getimed.

Aanvullend beschrijft Wikeley (2000) aspecten en processen van sociale, culturele en psychologische aard waardoor goed moet worden nagedacht aan wie, wanneer, bij welke gelegenheid, op welke manier, waarover, in welke vorm en in welke volgorde feedback wordt gegeven. Ontvangers van feedback (de betrokkenen in de school) moeten kunnen reageren, er betekenis aan kunnen toekennen, er iets mee kunnen doen. Het kan nodig zijn de feedback te doen vergezellen van gerichte vragen, gekoppeld aan actieperspectieven (doelen, middelen, standaarden). Er kunnen beperkingen en dilemma's naar boven komen, en verschillen in referentiekader, inclusief verschillen in de manier waarop wordt aangekeken tegen leren en veranderen, en verschillen in persoonlijke stijlen van leren. Er kan discussie ontstaan over de hardheid en de relevantie van gepresenteerde gegevens en betrokkenen kunnen verschillen in wat ze belangrijk vinden om na te streven. Conceptualiseringen en perspectieven kunnen gebonden zijn aan contexten (denk bijvoorbeeld aan vaksecties of onderbouw/ bovenbouw). Op de achtergrond zullen assumpties, waarden en (andere) culturele

aspecten meespelen. Als de feedback betrekking heeft op gerealiseerde veranderingen en effecten zal bovendien niet altijd duidelijk zijn waaraan die kunnen worden toegeschreven.

Geven van feedback op basis van onderzoek die moet bijdragen aan leren en veranderen, moet kortom zorgvuldig worden geformuleerd en gegeven, zowel betrekking hebben op het beoogde doel als op de weg daarnaar toe, en worden opgenomen in een dialoog tussen de betrokkenen waarin zij zich de feedback kunnen toe-eigenen en er gezamenlijk betekenis aan kunnen geven, ook in de zin van de consequenties die daaraan kunnen of moeten worden verbonden, waarbij altijd enige ruimte moet blijven bestaan voor persoonlijke interpretaties en invullingen.

De hier bedoelde datafeedback moet ruimer worden opgevat dan alleen in termen van het benutten van gegevens over de leerprestaties van leerlingen, zoals dat momenteel centraal staat in ontwikkelingen rond opbrengstgericht werken en afleggen van externe verantwoording (in de VS bekend als *policy-driven accountability testing* en *curriculum-based performance testing*, zie Meijssels, Atkins-Burnett, Xue, DiPrima Bickel & Son, 2003). Het hier bedoelde ruimere kader (in de VS aangeduid met *data-driven decision-making*, zie Luo, 2008; Park & Datnow, 2009; Wohlstetter, Datnow & Park, 2008) omvat gegevens van uiteenlopende aard:

- gegevens over leerresultaten van leerlingen;
- gegevens over kenmerken van het onderwijs zelf (curriculum, materialen, didactiek, organisatie);
- gegevens die op de eigen school zijn verzameld, zoals toetsresultaten, systematische observaties, tevredenheidsonderzoeken, rapportages van de onderwijsinspectie;
- uitkomsten van onderzoeken naar effectieve scholen en effectieve instructie, landelijke rapporten van de Inspectie, rapporten van het Cito over het onderwijsaanbod (PPON);
- uitkomsten van landelijk onderzoeken naar leerresultaten, normgegevens van toetsen, verwijzings, doorstroom- en slagingspercentages.

Veel van deze gegevens kunnen op een school zowel worden gebruikt voor verbetering (formatief) en als onderdeel van de interne kwaliteitszorg, als voor het afleggen van externe verantwoording, horizontaal en verticaal (summatief).

Eigen ervaringen met het terugkoppelen van verzamelde en verwerkte gegevens naar schoolleiders, teamleiders en docenten op de scholen waar de onderzoeken zijn uitgevoerd waarover in de eerdere hoofdstukken is gerapporteerd, zijn al beschreven in paragraaf 4.4.6. Daar bleek dat een aantal schoolleiders en teamleiders aangaf dat zij bij het interpreteren van de gegevens ondersteuning nodig hebben omdat zij niet gewend zijn zulke gegevens te lezen en te duiden. Dit blijkt ook uit andere onderzoeken naar “datageletterdheid” bij schoolleiders (o.a Geijsel & Krüger, 2005; Luo, 2008; Verhaeghe et al., 2011). Dit pleit voor een aanbod waarbij niet alleen wordt volstaan met een schriftelijke rapportage, maar tevens de dialoog wordt aangegaan door de betreffende onderzoekers met de betrokkenen.

Bij het vormgeven van datafeedback en de daarbij gewenste ondersteuning en dialoog kan mogelijk gebruik worden gemaakt van inzichten uit theorie en onderzoek over *information use environments* (zie Luo, 2008). Luo maakt aannemelijk dat het gebruik van gegevens door schooldirecteuren door verschillende factoren wordt beinvloed, waaronder extern gestelde eisen, de beschikbaarheid van relevante data, de perceptie van de kwaliteit van deze data, en de eigen data-analysevaardigheden, terwijl ook de aard van de problemen waarvoor de data zouden moeten kunnen worden gebruikt hierbij een rol speelt. Bij het benutten van gegevens over het pedagogisch-didactisch handelen van docenten voor hun professionele ontwikkeling en de ontwikkeling van het onderwijs in de school zullen daar in ieder geval ook de factoren bij komen die in dit hoofdstuk aan de orde kwamen, zoals de duidelijkheid, eenduidigheid en gemeenschappelijkheid van de visie op de gewenste ontwikkeling en de mate waarin en manier waarop leiding wordt gegeven aan het proces.

8.5 Conclusie

Voor het bouwen van een brug tussen het huidig handelen van docenten op een school en het handelen dat gewenst wordt geacht, is een aantal condities van belang. In de onderzoeken die in dit proefschrift zijn beschreven, kwamen drie van deze condities duidelijk naar voren en deze zijn in dit hoofdstuk nader verkend.
De uitkomsten van de onderzoeken die zijn beschreven in de hoofdstukken 3 t/m 6 zijn op diverse manieren teruggekoppeld naar de betreffende schoolleiders, teamleiders en docenten, veelal (mede) op hun eigen nadrukkelijke verzoek. Het bleek voor de meeste betrokkenen echter niet gemakkelijk om de verkregen gegevens te duiden en er conclusies, voornemens en beslissingen aan te verbinden. Datafeedback is van belang, juist om de dialoog over de gewenste ontwikkeling handen en voeten te geven, maar gebruik maken van datafeedback moet op zich ook worden geleerd (zie ook 8.4).
In de onderzoeken die zijn beschreven in de hoofdstukken 3, 4 en 6 en in paragraaf 8.3 bleek dat sprake is van aanzienlijke verschillen tussen docenten, ook binnen dezelfde school, in hun opvattingen en/of hun huidig handelen. In hoofdstuk 7 bleek bovendien sprake van verschillen tussen handelen dat gewenst kan worden geacht op basis van wetenschappelijke inzichten, de onderwijsconcepten van de scholen, en de keuzen die betrokkenen op de scholen maken. In reactie op de terugkoppeling van elk van deze onderzoeken bleken de schoolleiders de aangetroffen verschillen opmerkelijk te vinden. In hoofdstuk 7 is vastgesteld dat in de onderwijsconcepten van scholen uiteenlopende inzichten en doelen worden gecombineerd en is geconcludeerd dat aan het verminderen van de genoemde verschillen kan worden bijgedragen door het gezamenlijk formuleren van een duidelijke en (meer) coherente visie (zie ook 8.2).
In het onderzoek dat is beschreven in paragraaf 8.3 is een concreet vernieuwingstraject op een school een poos op de voet gevolgd. Op deze school was ruime consensus over de aanleiding voor en de algemene richting van de ingezette vernieuwing maar bleek het proces niet goed op gang te komen. In de conclusies en aanbevelingen in het

eindrapport werd tot uitdrukking gebracht dat dit waarschijnlijk diende te worden toegeschreven aan het feit dat aan het proces niet voldoende duidelijk leiding werd gegeven. Kennisname van het eindrapport door schoolleiding en docenten bleek beide partijen bij elkaar te brengen, waarna vervolgens in deze lacune werd voorzien.

In de onderzoeken in dit proefschrift stond het pedagogisch-didactisch handelen van docenten centraal. Het verder ontwikkelen van een breed (gevarieerd) en gedifferentieerd (flexibel) handelingsrepertoire vergt meer dan alleen de drie hier genoemde condities. Scholen en docenten zullen daartoe ook concrete activiteiten moeten ontplooien. Zij kunnen daartoe putten uit de beschikbare literatuur over leren en professionele ontwikkeling van docenten (bijvoorbeeld Opfer et al., 2011; Runhaar et al., 2009; Van Veen, Zwart, Meirink & Verloop, 2010; Vermunt & Endedijk, 2011).

9 Conclusies en discussie

Recapitulatie van context, aanleiding en vraagstelling

Schoolleiders in het voortgezet onderwijs nemen de afgelopen 10-15 jaar steeds meer het initiatief om het onderwijs op hun school te vernieuwen. Zij doen dit naar aanleiding van landelijk beleid, maatschappelijke veranderingen, schoolinterne ontwikkelingen en nieuwe wetenschappelijk inzichten. De gewenste vernieuwingen hebben veelal consequenties voor het pedagogisch-didactisch handelen van docenten. Dit vraagt om een dialoog tussen de betrokkenen die moet kunnen worden gevoed door inzicht in het huidig handelen van docenten en in handelen dat gewenst is in de context van de school.
In de onderzoeken waarover in de voorgaande hoofdstukken verslag wordt gedaan, zijn instrumenten ontwikkeld en gegevens verzameld om het huidige en gewenste handelen in kaart te brengen, zowel wat betreft het pedagogisch-didactisch handelen in de breedte als ten aanzien van enkele specifieke onderdelen daarvan.
De vraagstelling voor het gehele proefschrift is:
In hoeverre en onder welke voorwaarden is het mogelijk het huidige en gewenste pedagogisch-didactisch handelen van docenten in scholen voor voortgezet onderwijs met een al dan niet vernieuwend onderwijsconcept in kaart te brengen, op een zodanige wijze dat dit wetenschappelijk verantwoord gebeurt en wetenschappelijk relevante inzichten oplevert en tegelijk scholen, schoolleiders, teamleiders en docenten kan ondersteunen bij verdere ontwikkeling en vernieuwing?
De overkoepelende onderzoeksvragen zijn:
1 Waaruit bestaat het pedagogisch-didactisch handelen van docenten in het voortgezet onderwijs en hoe kan dit op valide, betrouwbare en bruikbare wijze worden gemeten?
2 In welke mate vertonen docenten dit handelen?
3 Hoe is de variatie in dit handelen en in hoeverre hangt deze samen met kenmerken van docenten (geslacht, leeftijd, ervaring, school) en leerlingen (geslacht, leeftijd, leerjaar, schooltype)?
4 Welk handelen wordt gewenst geacht en door wie en op welke gronden?
5 Wat zijn de verschillen tussen huidig en gewenst handelen en hoe kunnen deze worden overbrugd?

De mogelijke betekenis van de verkregen antwoorden in relatie tot de bestudeerde literatuur

In het algemeen kan onderzoek bestaande theoretische kaders of delen daarvan bevestigen, niet bevestigen (als bepaalde bevindingen niet worden verkregen), tegenspreken, nuanceren, preciseren of aanvullen. De bij de uitgevoerde onderzoeken

betrokken wetenschappelijke literatuur over het algemeen pedagogisch-didactisch handelen van docenten (hoofdstuk 3) en de wenselijkheid daarvan (hoofdstuk 7) en specifiek over geven van feedback en leerzaam beoordelen (hoofdstukken 5 en 6) bevat geen duidelijk concurrerende beweringen waarbij de verkregen bevindingen in het geding zouden kunnen worden gebracht. De onderzoekstradities en publicaties vullen elkaar aan en vertegenwoordigen verschillende perspectieven op onderwijs, leeromgevingen en handelen van docenten. Door het in dit proefschrift gelegde accent op het meetbaar maken van aspecten van het handelen van docenten en de opvattingen en percepties dienaangaande van verschillende betrokkenen, en het min of meer ontbreken in de literatuur van onderzoeken naar en uitspraken over verschillen en (causale) verbanden in relatie tot het algemeen pedagogisch-didactisch handelen, is ook het tegenspreken van beweringen niet aan de orde. Dit zou eventueel wel hebben gekund ten aanzien van de specifieke taakgebieden geven van feedback en leerzaam beoordelen, die in theorie en onderzoek expliciet in relatie zijn gebracht met effecten op motivatie en prestatie bij leerlingen, als we in de betreffende onderzoeken ook zulke effecten hadden kunnen betrekken. Dat was echter niet (hoofdstuk 5) of amper (hoofdstuk 6) het geval. De uitgevoerde onderzoeken kunnen al met al dus alleen bevestigen of niet bevestigen, aanvullen, en nuanceren of (conceptueel) preciseren.
De literatuur bevat wel uitspraken over taken die docenten vervullen of zouden moeten vervullen, maar niet over normen of standaarden, in termen van de frequentie, intensiteit, duur of mate van kwaliteit van het handelen. Daardoor kunnen vergelijkingen met de literatuur voornamelijk worden gemaakt bij de overkoepelende onderzoeksvragen 1 en 4 (de taken die in het handelen van docenten kunnen worden onderscheiden en het gewenste handelen) en hebben de opbrengsten bij de onderzoeksvragen 2 en 3 (aangaande de mate waarin docenten dit handelen vertonen en de variatie daarin) vooral het karakter van empirische bevindingen die aanleiding kunnen zijn voor theorievorming en verder onderzoek. Wel kunnen sommige bevindingen met betrekking tot de mate waarin docenten bepaalde taken vervullen (onderzoeksvraag 2) en de vraag waarmee de variatie in het handelen samenhangt (een deel van onderzoeksvraag 3) met de literatuur in verband worden gebracht.

Antwoorden op de overkoepelende onderzoeksvragen, opbrengsten en interpretaties

Waaruit bestaat het pedagogisch-didactisch handelen van docenten in het voortgezet onderwijs en hoe kan dit op valide, betrouwbare en bruikbare wijze worden gemeten?

Het begrip ‘handelen’ is afgebakend ten opzichte van aanpalende begrippen zoals vaardigheid, gedrag en competentie. Handelen betreft intentioneel gedrag dat door relevante anderen als zodanig kan worden waargenomen en ervaren. Handelen van docenten in interactie met leerlingen betreft wat docenten doen en beogen en door leerlingen kan worden herkend. Vergeleken met handelen bestrijkt vaardigheid ook de

resultaten van handelen, mist gedrag de intenties, en betreft competentie ook de achterliggende kennisbasis (Reynolds, 1992; Roelofs et al., 2008; Shulman, 1987; Vermunt et al., 1999).
Het (algemeen) pedagogisch-didactisch handelen is onderscheiden van andere dimensies in het handelen van docenten zoals het interpersoonlijke, het organisatorische en het vakdidactische (zie onder andere Bransford et al., 2005; Brekelmans et al., 2000).
Van de in de literatuur aangetroffen manieren om het pedagogisch-didactisch handelen te specificeren is die in termen van taken het meest geschikt voor gestructureerde en tussen docenten, vakken en scholen vergelijkbare bevraging van docenten en leerlingen over het handelen van de docenten. Specificatie in termen van rollen blijft te globaal, die in termen van instructiestrategieën en van afzonderlijke verbale uitingen zijn te specifiek, en die in termen van te creëren leeromgevingen is meer voorwaardelijk en betreft ook handelen buiten de interactiesituatie (zie onder meer Brown & Campione, 1996; Collins et al., 1989; De Corte, 2000).
Vergelijking van wetenschappelijke literatuur over didactische modellen (zoals De Corte et al., 1976; Merril, 2002; Valcke, 2007), meta-analyses van factoren voor effectief onderwijs (zoals Scheerens, 2007; Seidel & Shavelson, 2007) en leren en leeromgevingen (zoals Bransford, 2000; Brown & Campione, 1986; Doyle, 1986; Shute, 2008; Webb, 2009), landelijke beleidsdocumenten en onderwijsconcepten van scholen, heeft geleid tot het onderscheiden binnen het pedagogisch-didactisch handelen van 15 rubrieken (in het vervolg aangeduid met 'taakgebieden'), die daarna door een aantal docenten, werkzaam op verschillende scholen, konden worden gespecificeerd en geconcretiseerd in beschrijvingen van handelingen (uitspraken, 'items'). Extra opgenomen (op verzoek van de schoolleiders) is een taakgebied over het werken aan de eigen professionele ontwikkeling.
Met het oog op grootschalige bevraging van docenten en leerlingen, gericht op het in kaart brengen van het huidige en gewenste repertoire van het pedagogisch-didactisch handelen van docenten, is een schriftelijke vragenlijst ontwikkeld waarin dit handelen is uiteengelegd in 16 taakgebieden en rond 150 items. De betrokken schoolleiders en docenten zagen deze taakgebieden en beschrijvingen van activiteiten als herkenbaar en dekkend. De vragenlijst kon worden afgenomen bij 140 docenten en 900 leerlingen op zes scholen, met behulp waarvan de psychometrische kwaliteiten konden worden vastgesteld. Factoranalyses per taakgebied en daarop volgende schaalanalyses brachten aan het licht dat zes taakgebieden elk twee subschalen bevatten. Mede gebaseerd op de onderscheiden taakgebieden en de gevonden subschalen is, op verzoek van de schoolleiders, een verkorte versie van de vragenlijst geconstrueerd waarin het pedagogisch-didactisch handelen is uiteengelegd in 22 taken met rond 100 items (deze verkorte versie wordt in het vervolg aangeduid met 'PDH', en bevat wederom een extra taak met betrekking tot de eigen professionele ontwikkeling).
Door het gevolgde ontwikkelproces zijn waarborgen ingebouwd voor de construct-validiteit (gerichtheid op het construct 'pedagogisch-didactisch handelen') en de

inhoudsvaliditeit (dekking van een breed repertoire van taakgebieden), alsmede voor de ecologische validiteit (herkenbaarheid voor docenten en leerlingen; zie Beijaard, 1990) en de *systemic validity* (congruentie met en bijdragend aan de beoogde professionele ontwikkeling in de context van het onderwijsconcept van de scholen; zie Frederikson & Collins, 1989). Over de constructvaliditeit, inhoudsvaliditeit, ecologische validiteit en *systemic validity* van de PDH kan worden opgemerkt dat de verkorte versie is gebaseerd op het uitgebreide instrument en kan naar dat instrument worden verwezen, onder de kanttekening dat per taak bezien de inhoudsvaliditeit en de *systemic validity* door het schrappen van items mogelijk enigszins zijn verminderd.

Een aanvullende indicatie voor de validiteit van de resulterende schalen kan worden gevonden in de ruim voldoende tot goede (taakgebieden, uitgebreide versie) respectievelijk acceptabele tot goede (taken, PDH) interne consistenties. Zowel de taakgebieden als de taken vormen samen een in hoge mate betrouwbare schaal voor het gehele repertoire van pedagogisch-didactisch handelen. De intraclassecorrelatie varieert per taakgebied van .31 tot .50 en per taak van .27 tot .36, en is voor de overallschaal bij de taakgebieden .50 en bij de taken .35. Met beide versies van het instrument kunnen docenten dus goed worden onderscheiden. Dat de interne consistenties en intraclassecorrelaties bij de taakgebieden hoger zijn dan bij de taken maakt duidelijk dat het loont om het handelen van docenten op uitgebreide wijze te operationaliseren en dat de lagere waarden bij de PDH de prijs zijn die voor een verkorte versie moet worden betaald.

Indicaties voor discriminante validiteit van de PDH zijn verkregen door berekening van correlaties van de taken, de vier onderscheiden takengroepen en de overallscore van de PDH met de sectoren en de dimensies van de VIL. De overall-score van de PDH hangt significant maar gematigd positief samen met beide VIL-dimensies. Daaruit kan worden geconcludeerd dat beide instrumenten twee te onderscheiden constructen meten. De meeste afzonderlijke taken en de vier takengroepen (ten aanzien van ontwikkeling, leerproces, feedback & beoordeling, en samenwerken & ontwikkeling) hangen eveneens significant samen met een of beide dimensies van de VIL.

De gevonden relaties van de VIL-dimensie Invloed met de PDH-takengroep leerproces en de VIL-dimensie Nabijheid met alle vier PDH-takengroepen passen goed bij de aard van de taken die in deze clusters aan de orde zijn. De PDH bestrijkt naast meer traditionele taken die zijn gericht op het cognitieve leerproces een aantal min of meer vernieuwende taken die de afgelopen 15 jaar in het voortgezet onderwijs meer aandacht hebben gekregen en een toenemende gerichtheid op de leerlingen inhouden en die minder met de dimensie Invloed en meer met de dimensie Nabijheid samenhangen.

Gelet op de positieve relaties van de (taken van de) PDH met de VIL-sectoren BS, SB en SO en de negatieve relaties met de sectoren OT en TO, en de in onderzoek vastgestelde relaties van deze sectoren met het leerproces van leerlingen, kan worden verondersteld dat zulke relaties ook bij de (taken van de) PDH kunnen worden gevonden. Een mogelijke interpretatie van de gebleken (gematigd) positieve relatie

tussen de VIL en de PDH is dat adequaat interpersoonlijk handelen een voorwaarde is om het pedagogisch-didactisch handelen tot zijn recht te kunnen laten komen.
Indicaties voor confirmerende validiteit zijn verkregen door variantieanalyses van de verschillen tussen groepen die kunnen worden gevormd met behulp van een aantal kenmerken van de leerlingen (geslacht, leeftijd, leerjaar, schooltype, school) en de docenten (geslacht, leeftijd, school, ervaring op deze school, ervaring totaal), een tweeweg variantie-analyse met als factoren schooltype en school, en multilevelanalyse met drie niveaus (school, docent, leerling) en de kenmerken van docenten en leerlingen als predictoren. Consistent significante verschillen doen zich alleen voor bij de leerlingen tussen jongens en meisjes (jongens scoren gemiddeld iets hoger dan meisjes), terwijl bij de multilevelanalyse sprake is van significante variantie op docentniveau (24%), maar niet op schoolniveau (11%).

Als proeve van meer gedetailleerde uitwerking zijn twee taakgebieden nader onderzocht: geven van feedback en (leerzaam) beoordelen. Uit de literatuur over feedback konden ruim 20 aspecten, inhouden en vormen van feedback worden gedestilleerd die van toepassing kunnen zijn in de situatie dat docenten mondeling feedback geven aan leerlingen die zelfstandig aan het werk zijn. Deze betreffen de context en timing (eenmalig of onderdeel van een interactieproces; in aanwezigheid van anderen of niet; niet te snel en niet te langzaam), de aandacht voor de taak (geven van informatieve en beschrijvende feedback in plaats van louter een cijfer of feedback op de persoon van de leerling), de cognitieve belasting (in relatie tot de hoeveelheid, duidelijkheid en concreetheid), de affectieve lading (in relatie tot de vriendelijkheid (toon), mate van controle en balans positief-negatief), de aandacht voor de standaard (om de actuele prestatie mee te vergelijken), en een aantal mogelijke aanzetten tot voortgezette interactie (zoals geven van uitleg, vereenvoudigen van het probleem, voordoen, geven van suggesties, stellen van vragen). Een aantal kenmerken en vormen van feedback wordt in de literatuur verondersteld positief bij te dragen aan motivatie en leren, zoals: niet eenmalig, goed getimed, cognitief niet te belastend, niet controlerend, gericht op de taak en gerelateerd aan een standaard, en vatbaar voor c.q. bijdragend aan verdere interactie.
Uit de literatuur over beoordelen werd een aantal aanwijzingen verkregen aangaande beoordelingen die kunnen bijdragen aan motivatie en leren. Deze betreffen onder meer de beoordeling als onderdeel van het onderwijsleerproces; de keuze van geschikte taken; het aan de leerlingen verduidelijken van de bedoeling; de voldoende voorbereiding van de leerlingen op het maken van de taak; de afstemming tussen de leerstof, het onderwijs en de beoordeling; het louter beoordelen op de relevante kennis en vaardigheden; het adequaat communiceren van de beoordeling; het verduidelijken hoe de beoordeling tot stand is gekomen en wat deze betekent; het vermijden dat de beoordeling een ongewenste impact heeft; het bevorderen dat de beoordeling een leereffect heeft, dat de leerling er iets mee kan doen, en dat ook doet.

Op beide taakgebieden bleek het mogelijk, en was het onvermijdelijk, om tot een veel meer uitgewerkte en gespecificeerde operationalisering te komen dan in het hiervoor beschreven instrument is gerealiseerd, dat is gericht op het inventariseren van het repertoire van docenten in de breedte van het pedagogisch-didactisch handelen (zie de hoofdstukken 5 en 6).

In welke mate vertonen docenten dit handelen?

De gemiddelde scores over alle docenten (n=141) op de taakgebieden van het pedagogisch-didactisch handelen van het ontwikkelde instrument (hoofdstuk 3) variëren van 3,4 tot 4,3 (op de gebruikte schaal van 1 tot 5) op basis van de antwoorden van de docenten zelf en van 3,3 tot 4,0 bij de leerlingen. De gemiddelde overallscore op basis van de antwoorden van de docenten is 3,8 en op basis van de antwoorden van de leerlingen 3,7.
De gemiddelde scores over alle docenten (n=49) op de taken van de PDH (hoofdstuk 4) variëren van 2,9 tot 4,4 (op de gebruikte schaal van 1 tot 5) op basis van de antwoorden van de docenten zelf en van 3,0 tot 3,9 bij de leerlingen. De gemiddelde overallscore op basis van de antwoorden van de docenten is 3,7 en op basis van de antwoorden van de leerlingen 3,6.
Al met al blijkt dat de docenten gemiddeld in behoorlijke mate de in beide versies van het instrument vertegenwoordigde breedte van het pedagogisch-didactisch handelen bestrijken. Men kan zich afvragen wat dit betekent en waardoor dit komt. Mogelijk zijn de scores gemiddeld relatief hoog doordat de meeste docenten werken op vernieuwende scholen, of doordat veel taakgebieden en taken een leerlinggericht karakter hebben, of doordat docenten gemiddeld geneigd zijn zichzelf relatief hoog in te schatten en leerlingen gemiddeld enigszins geneigd zijn hun docenten positief te beoordelen.
Van de taakgebieden in de uitgebreide versie scoren de taakgebieden 1 (zorgen voor een goed leerklimaat) en 14 (samenwerken met andere leraren) bij de leerlingen en docenten het hoogst en taakgebied 4 (leerlingen het nut en de samenhang laten zien) het laagst. Van de taken in de verkorte versie scoren de hiermee corresponderende taken eveneens relatief hoog respectievelijk laag, en scoren verder vooral taken laag die in de verkorte versie zijn afgesplitst van hun oorspronkelijke taakgebied, op grond van de subschalen die bij de factoranalyses in hoofdstuk 3 naar boven kwamen en waarbij steeds de tweede subschaal lager scoorde dan het eerste. Dit betreft vooral de taken 5 (leerlingen wijzen op verbanden tussen vakken), 8 (leerlingen vaardigheden aanleren), 13 (leerlingen ondersteunen bij zelf maken van keuzes) en 18 (bevorderen dat feedback wordt gebruikt).
Een mogelijke interpretatie van deze verschillen tussen de taakgebieden en taken is dat docenten gemiddeld vrij sterk zijn gericht op hun sociale context (het leerklimaat voor de leerlingen, de samenwerking met collega's) en dat inhoudelijke taken die niet direct

behoren tot het traditionele lesgeven c.q. de overdracht van kennis van hun vak daarmee vergeleken minder aandacht krijgen.

Op de specifieke taakgebieden geven van feedback en beoordelen blijkt het volgende.
Zowel de docenten zelf als hun leerlingen geven aan dat de docenten in ruime mate en op gevarieerde wijze feedback geven, waarbij zij volgens de leerlingen iets meer zijn gericht op het resultaat dan op het proces en minder aandacht hebben voor het benutten van feedback. Ook de observaties van opgenomen lessituaties, aan de hand van categorieën die zijn gebaseerd op in de literatuur geïnventariseerde aspecten, inhouden en vormen van feedback, laten een breed repertoire van feedbackgedrag zien, met iets vaker procesgerichte dan resultaatgerichte feedback. Het (beperkte) verschil in de verhouding tussen de aandacht voor het proces en voor het resultaat tussen de vragenlijsten en de observaties heeft mogelijk (mede) te maken met een verschil in operationalisering, maar kan er ook op wijzen dat leerlingen als zij mondeling feedback krijgen als onderdeel van of leidend tot een gesprekje (van enkele minuten) zij procesgerichte feedback minder vaak als feedback waarnemen of zich als zodanig herinneren. Uit de observaties bleek verder dat de docenten in hun feedback niet zo vaak verwijzen naar een of meer standaarden waaraan het werk moet voldoen.
Wat betreft het handelen van docenten op het taakgebied beoordelen scoren zowel de docenten zelf als hun leerlingen dit handelen gemiddeld minder hoog dan op het taakgebied feedback. Vooral de vergeleken met de traditionele praktijk van veel docenten waarschijnlijk meest vernieuwende activiteiten, zoals het betrekken van leerlingen bij het beoordelingsproces (bijvoorbeeld door hen werk van medeleerlingen en/of van henzelf te laten beoordelen) en het uitgebreider nabespreken van gegeven beoordelingen doen de docenten minder.

Hoe is de variatie in dit handelen en in hoeverre hangt deze samen met kenmerken van docenten (geslacht, leeftijd, ervaring, school) en leerlingen (geslacht, leeftijd, leerjaar, schooltype)?

De individuele overallscore op basis van de antwoorden van de docenten zelf op het in hoofdstuk 3 ontwikkelde instrument varieert van 2,7 tot 4,4 en op basis van de antwoorden van de leerlingen bij deze docenten van 1,3 tot 4,9. De gemiddelden op de zes aan dit onderzoek deelnemende scholen op basis van de antwoorden van de leerlingen verschilden overall significant maar paarsgewijs vergeleken niet. De individuele overallscore op basis van de taken van de PDH (hoofdstuk 4) op basis van de antwoorden van de docenten zelf varieert van 2,4 tot 4,9 en op basis van de antwoorden van de leerlingen van 1,4 tot 4,6. Dit betekent dat sprake is van grote verschillen tussen docenten.
De scores van de docenten in hoofdstuk 4 op de taken van de PDH hangen niet duidelijk samen met hun geslacht (behalve bij taak 9, inspelen op verschillen tussen leerlingen, waarop de vrouwen gemiddeld hoger scoren (4,3) dan de mannen (3,7)),

leeftijd of onderwijservaring (in totaal en op de betreffende school), en verschillen bij slechts vier taken tussen de bij deze analyse betrokken vier scholen. De overallscores van de leerlingen met betrekking tot het handelen van hun docenten hangen samen met hun geslacht (jongens scoren gemiddeld iets hoger dan meisjes), leeftijd (met de laagste scores bij de 14- en 15-jarigen), leerjaar (met de laagste scores in leerjaar 3), schooltype (gemiddeld hogere scores bij de vmbo-leerlingen), en school (waarbij twee van de vier scholen op de meeste taken duidelijk hoger scoren dan de twee andere scholen). Bij de multilevelanalyse was de variantie op schoolniveau niet significant, die op docentniveau wel, en bleek alleen het geslacht van de leerlingen een significante predictor.
Dat geen duidelijke relatie is gevonden tussen de scores op de PDH en de leeftijd en ervaring van de docenten kan komen doordat in dit onderzoek relatief weinig oudere en meer ervaren docenten waren betrokken. Het kan ook komen doordat de docenten in dit onderzoek werkzaam zijn in een context van verandering en vernieuwing. Ook docenten die al de nodige ervaring hebben, kunnen daardoor opnieuw in een leerproces betrokken raken, vanwege nieuwe eisen aan hun pedagogisch-didactisch handelen. Aangezien een deel van de taken van de PDH betrekking heeft op relatief vernieuwend (en daarom wellicht niet gemakkelijk) handelen, is deze mogelijkheid niet denkbeeldig.
Bij afwezigheid in de literatuur van duidelijke beweringen over de (te verwachten) relatie tussen het pedagogisch-didactisch handelen van docenten en de hierboven genoemde kenmerken van docenten en leerlingen, kan het niet vinden van zulke relaties worden opgevat als indicatie voor het ontbreken van bias. De iets lagere scores bij 14/15-jarigen c.q. in leerjaar 3 kunnen worden geïnterpreteerd vanuit de in die leeftijdsperiode bekende meer kritische opstelling van jongeren jegens volwassenen (puberteit). Dominerend in de uitkomsten is het gegeven dat sprake is van grote individuele verschillen tussen docenten.

Op de twee nader uitgewerkte taakgebieden geven van feedback en leerzaam beoordelen zijn naast grote individuele verschillen tussen de docenten ook verschillen tussen scholen zichtbaar geworden.
De opvattingen over feedback die de docenten in het onderzoek in hoofdstuk 5 inbrachten, verschillen zowel tussen scholen als binnen dezelfde school, terwijl de visies die door de schoolleiders zijn geuit niet duidelijk tussen de scholen verschillen. Hierdoor kunnen deze opvattingen en visies niet per school één op één op elkaar worden gelegd en kan niet worden aangegeven in hoeverre de verschillen tussen de docenten moeten worden gezien als individuele verschillen dan wel als verschillen die schoolgebonden zijn. Het in de vragenlijsten door de leerlingen aangegeven feedbackgedrag van hun docenten verschilt significant tussen de vier deelnemende scholen bij alle vier gebruikte schalen (geven van feedback, zorgen dat feedback wordt gebruikt, resultaatgerichte feedback, procesgerichte feedback). Op twee van de vier scholen (een meer traditionele havo/vwo-school die aan dit onderzoek deelnam met havo-leerlingen, en een sterk vernieuwende vmbo-school) waren de gemiddelde scores

op alle vier schalen hoger dan op de andere twee scholen (beide sterk vernieuwende havo/vwo-scholen). Al met al kan, mede door de kleine onderzoeksgroep, niet worden aangegeven in hoeverre sprake is van verschillen tussen docenten door verschillen tussen de scholen waar ze werken, of van verschillen tussen scholen door verschillen tussen de docenten die daar werken.

Op het taakgebied (leerzaam) beoordelen kon in hoofdstuk 6 worden geconstateerd dat verschillen tussen scholen, tussen vakken, tussen typen taken en tussen docenten met elkaar verweven zijn. Docenten kiezen voor bepaalde typen taken (toetsen, of kleine en/of individuele opdrachten, of grotere, meer complexe en/of in samenwerking uit te voeren opdrachten) in de context van hun school en hun vak. Een vak als Engels vormt behalve een kennisdomein (woordenschat, grammatica) vooral een vak van vaardigheden die worden geoefend en getoetst aan de hand van kleine halfopen opdrachten rond woorden en zinnen. Aardrijkskunde is een breed vak met kennis van feiten en begrippen over allerlei maatschappelijke en natuurwetenschappelijke onderwerpen, en kent vakeigen benaderingen (schaalniveaus, perspectieven) en vakspecifieke vaardigheden (zoals bijvoorbeeld kaartvaardigheden). Dit vak leent zich voor complexe opdrachten waarbij leerlingen kunnen samenwerken en ook verbanden met andere vakgebieden kunnen worden gelegd. Het vak wiskunde heeft een duidelijke eigen structuur en opbouw en wordt meestal geoefend en getoetst aan de hand van vragen naar berekeningen, toepassingen en inzicht, waarbij antwoorden duidelijk goed of fout zijn maar het proces van oplossen vaak belangrijker wordt gevonden dan de uitkomst. In de taken die de docenten hebben ingebracht in het onderzoek zijn deze verschillen tussen de vakken terug te vinden, maar ook de achtergrond van de scholen waar de docenten werken. Dit heeft erin geresulteerd dat bij de docenten Engels en aardrijkskunde op een nieuwe, sterk vernieuwende school alle taken de vorm hadden van opdrachten, die bij de docenten Engels en aardrijkskunde op een iets meer ervaren sterk vernieuwende school deels de vorm van opdrachten en deels die van toetsen, en de taken van de docent wiskunde op laatstgenoemde school en de docenten van alle drie vakken op een meer traditionele school de vorm van toetsen. De genoemde verschillen tussen de scholen en de vakken werken door in specifieke aspecten van het beoordelingsproces: bij een gestructureerd vak met een aantal uren per week zoals wiskunde worden de taken soms vrij uitgebreid vooraf en achteraf besproken, terwijl dit bij een breed vak met weinig uren per week zoals aardrijkskunde veel minder gebeurt.

Beide hoofdstukken 5 en 6 samengenomen valt op dat waar in de literatuur formatieve en summatieve assessment vaak als één domein worden gezien, in het handelen van docenten bij het geven van feedback het meest beoordelende aspect, namelijk vergelijken met de standaard, minder aandacht krijgt, en in het handelen bij beoordelen het misschien wel meest formatieve aspect, namelijk leerlingen betrekken bij de beoordeling van hun eigen werk, minder uit de verf komt. In beide gevallen wordt dan een aspect onderbelicht dat voor leerlingen juist heel leerzaam kan zijn (zie Sadler, 1989). Dat bij feedback een duidelijk beoordelend aspect en bij beoordelen een

potentieel formatief aspect minder aandacht krijgt, lijkt erop te wijzen dat geven van feedback en beoordelen door docenten, anders dan in recente literatuur wordt aanbevolen, als twee duidelijk verschillende taken worden gezien.

Welk handelen wordt gewenst geacht en door wie en op welke gronden?

De gemiddelde score op het gewenst handelen (overallschaal) is op de taakgebieden (zie hoofdstuk 3) bij de docenten 4.3 (range 3,8-4,6) en bij de leerlingen 4,1 (range 3,8-4,5). Bij beide groepen scoren de taakgebieden 1 (leerklimaat) en 14 (samenwerking) het hoogst, en bij de docenten 9 (laten kiezen) en bij de leerlingen 4 (nut en samenhang) het laagst. De intraclassecorrelatie bij de overallschaal is laag (0.10). Uit dit feit dat de overallscore van de leerlingen op het gewenste handelen niet duidelijk onderscheid maakt tussen de docenten kan worden afgeleid dat, bij het beantwoorden per taakgebied per item van de vragen over het huidige en het gewenste handelen, hun score ten aanzien van het gewenste handelen niet sterk is gebaseerd op hun score op het huidige handelen, maar meer op een algemener beeld van wat zij als gewenst handelen van docenten zien (op de betreffende school).
Op de taken (zie hoofdstuk 4) is de gemiddelde score op het gewenst handelen bij de docenten 4,3 (range 3,7-4,7) en bij de leerlingen 4,2 (range 3,6-4,4). Bij de docenten scoren de taken 1 (leerklimaat) en 20 (samenwerking) het hoogst en bij de leerlingen de taken 1 (leerklimaat) en 3 (lesgeven), en bij zowel docenten als leerlingen scoort taak 5 (verbanden tussen vakken) het laagst. Beide groepen hebben dus aan het laatste verhoudingsgewijs niet veel behoefte. Een mogelijke interpretatie hiervan is dat beide groepen het werken en leren binnen vakken wel goed of voldoende vinden. Het belang dat leerlingen hechten aan het leerklimaat, het lesgeven, en de samenwerking tussen docenten, wordt bevestigd in de groepsinterviews met leerlingen die zijn gehouden op de scholen in het onderzoek in hoofdstuk 7, waarin de leerlingen precies dezelfde drie taken naar voren brachten.
Specifiek op de taken geven van feedback en bevorderen dat feedback wordt gebruikt (hoofdstuk 5) scoren de docenten zelf op het gewenste handelen gemiddeld 4,5 (range 4,0-4,8) respectievelijk 3,9 (range 3,5-4,5) en de leerlingen het door hen gewenste handelen van hun docenten gemiddeld 4,3 (range 3,8-4,6) respectievelijk 3,6 (range 3,3-4,1). De leerlingen hebben dus verhoudingsgewijs niet veel behoefte aan handelen van de docent dat erop is gericht dat zij gegeven feedback ook gebruiken. Een mogelijke interpretatie hiervan is dat leerlingen dit graag zelf willen bepalen.
Ten aanzien van het (leerzaam) beoordelen (hoofdstuk 6) willen de leerlingen gemiddeld graag dat hun docenten de criteria en standaarden vooraf bespreken (gemiddelde score 3,8) en de opdracht en beoordeling met hen nabespreken (3,8). Aan het betrekken van de leerlingen bij de beoordeling hebben de leerlingen veel minder behoefte (2,9). Mogelijk zien leerlingen dit toch als taak van de docent. De neiging van docenten om dit minder te doen (zie boven) zou daar dan goed bij passen.

In het onderzoek in hoofdstuk 7 is een vergelijking gemaakt tussen uitspraken over gewenst handelen van docenten vanuit vier bronnen: wetenschappelijke literatuur, beleidsgerichte publicaties, onderwijsconcepten van vernieuwende scholen, en groepen betrokkenen op en bij die scholen.

In het algemeen wordt in de *wetenschappelijke literatuur* van docenten verwacht dat ze goede leeromgevingen creëren, een aantal pedagogisch-didactische taken vervullen en in interactie met leerlingen een gedifferentieerd gespreksrepertoire hanteren. Specifiek wordt vanuit de behavioristische traditie gewezen op het belang van goed klassenmanagement en het gestructureerd aanbieden van informatie en laten oefenen en automatiseren van vaardigheden, vanuit de cognitieve psychologie op het bevorderen van cognitieve en metacognitieve denkprocessen gericht op analyseren, redeneren en problemen oplossen op onderscheiden domeinen van kennis en vaardigheid, en vanuit de sociaal-constructivistische benadering op het belang van interactie en samenwerking en van werken aan betekenisvolle taken. Deze aanbevelingen zijn niet bedoeld als vaste, algemeen geldige procedures maar als principes die docenten, vanuit een goed begrip ervan, flexibel, passend binnen de context, op maat en in onderlinge samenhang kunnen toepassen (Opfer et al., 2011; Shulman, 1987). Daarnaast wordt het belang benoemd van bepaalde persoonlijke eigenschappen van docenten (zoals deskundig, duidelijk, enthousiast, flexibel, taakgericht en met hoge verwachtingen). Recent wordt het belang genoemd van het werken vanuit een professioneel denkkader, waarin inzichten in wat leerlingen moeten leren, hoe zij dat kunnen leren, welke leeromgeving en welk docenthandelen daaraan kunnen bijdragen en hoe kan worden vastgesteld wat er wordt geleerd, in onderlinge samenhang worden ingezet (De Corte, 2000; Bransford et al., 2005). De gronden voor de aanwijzingen in de wetenschappelijke literatuur hebben zowel betrekking op theorie als op empirisch onderzoek.

In *beleidsgerichte publicaties* wordt meer aandacht besteed aan wat leraren zouden moeten weten en kunnen dan aan hoe ze zouden moeten handelen. Traditioneel wordt van docenten een combinatie verwacht van vakkennis en pedagogisch-didactische bekwaamheden. Tegenwoordig worden ook bekwaamheden verwacht op het vlak van samenwerken en professionele ontwikkeling, en recent ook voor het signaleren en begeleiden van individuele leerlingen met problemen en het doen van onderzoek. De gronden voor de hiervoor genoemde aanwijzingen komen voort uit traditioneel denken over docenten en uit ontwikkelingen in de samenleving en in het onderwijs. Beleidsgerichte studies rapporteren dat in antwoord op veranderingen en problemen scholen werken aan de ontwikkeling van meer zelfstandig leren, aanbieden van betekenisvolle contexten en opdrachten (in leergebieden), en samenwerking door leerlingen, en dat deze vernieuwingen kunnen bijdragen aan motivatie en leren maar dat goede begeleiding door een docent daarbij nodig blijft.

In de *onderwijsconcepten* van vernieuwende scholen staat een aantal docenttaken centraal. Belangrijk is dat docenten zorgen voor een goed leerklimaat en bijdragen aan persoonlijke vorming van leerlingen. Zij moeten inspirerende instructie geven, die aansluit bij het niveau van de leerlingen. Ze moeten leerlingen activeren en stimuleren

en daarbij inspelen op verschillen tussen leerlingen en leerlingen keuzes bieden. Zij moeten het leerproces van leerlingen volgen en (liefst op maat) leerlingen begeleiden en coachen bij zelfstandig werken en samenwerken. Zij moeten leerlingen feedback geven en de dialoog aangaan met leerlingen. Docenten moeten ook zelf bereid zijn tot reflecteren en tot het ontvangen van feedback. Ze moeten zich blijven ontwikkelen en samenwerken. In de onderwijsconcepten wordt al met al een breed pallet van taken beschreven. Daarin lijkt een zeker accent te liggen op pedagogische doelen en op het zoveel mogelijk aansluiten bij de leerlingen. In de wetenschappelijke literatuur (zoals Collins et al., 1989) en in de beleidsmatig vastgestelde SBL-competenties (Stichting Beroepskwaliteit Leraren, 2004) is sprake van een organiserend kader (namelijk een model van *cognitive apprenticeship* respectievelijk een zevental te onderscheiden competentiedomeinen). Dit komt in de onderwijsconcepten minder duidelijk naar voren. De gronden voor gewenst docenthandelen in de onderwijsconcepten van scholen zijn te vinden in ontwikkelingen in de samenleving, de leerlingenpopulatie en landelijk beleid en worden per school op specifieke wijze gecombineerd met bepaalde actuele concepten met betrekking tot onderwijs en leren (zoals bijvoorbeeld meervoudige intelligentie, breinleren of natuurlijk leren).

De *schoolleiders, teamleiders en docenten* op de in dit onderzoek betrokken scholen vinden als groep van de aan hen voorgelegde 22 taken van docenten van de PDH de volgende zeven taken de belangrijkste: zorgen voor een goed leerklimaat; geven van les, uitleg en instructie; leerlingen stimuleren en activeren; leerlingen ondersteunen bij het sturen van hun leerproces; geven van feedback aan leerlingen; bijdragen aan persoonlijke vorming van leerlingen; werken aan eigen verdere professionele ontwikkeling. Deze prioriteiten betreffen de meer traditioneel te noemen taken op het vlak van leerklimaat, persoonlijke vorming, lesgeven, stimuleren en ondersteunen bij het leerproces, naast geven van feedback en werken aan verdere eigen professionele ontwikkeling. Taken die te maken hebben met nieuwere opvattingen, zoals het nut van de leerstof laten zien (bijvoorbeeld via betekenisvolle contexten en opdrachten), samenwerking tussen leerlingen en leerlingen zelf keuzes laten maken, komen in de prioriteiten van schoolleiders, teamleiders en docenten minder naar voren, evenals het inspelen op verschillen tussen leerlingen, en het beoordelen. De gronden waarop men zich baseert bij het kiezen van de belangrijkste taken van een docent op de betreffende school, zijn vooral terug te voeren op de eigen ervaring en opvattingen, en minder op gesprekken met anderen, bepaalde theorieën of vakliteratuur (de roep om meer *evidence-based* werken ten spijt), of het onderwijsconcept van de school. Dat wat de leerlingen op de scholen noemen als wenselijk handelen van docenten past goed bij de onderwijsconcepten van hun scholen en heeft vooral betrekking op drie taakgebieden: het leerklimaat, het lesgeven, en de onderlinge samenwerking tussen docenten. Dit past bij de eerdere interpretatie van het gebleken gemiddelde profiel van de docenten op de taakgebieden en taken, waarin het traditionele lesgeven bij veel docenten nog steeds voorop staat.

Vergelijking tussen deze vier bronnen laat overeenkomsten en verschillen zien.
In de voor dit onderzoek geselecteerde wetenschappelijke literatuur ontbreekt de persoonlijke vorming van de leerlingen, maar ook over pedagogische doelen is wetenschappelijke literatuur beschikbaar. De in dit onderzoek bestudeerde literatuur besteedt meer dan de andere bronnen aandacht aan op leren gerichte interacties met leerlingen (tijdens het coachen). De reeks taken van docenten in het *cognitive apprenticeship* model is terug te vinden in de set van 22 taken van de PDH.
In de beleidsgerichte bronnen zijn het vooral de SBL-competenties die een breed scala aan pedagogisch-didactisch handelen beschrijven. De in de wetenschappelijke literatuur genoemde taken komen daarin min of meer allemaal terug. Aanvullend op de bestudeerde wetenschappelijke literatuur wordt in het recente beleid speciale aandacht gevraagd voor het doen van onderzoek, en voor het signaleren en begeleiden van individuele leerlingen die met problemen kampen.
In de onderwijsconcepten van de scholen wordt eveneens een breed pallet van taken beschreven. Daarin lijkt een zeker accent te liggen op pedagogische doelen en op het zoveel mogelijk aansluiten bij de leerlingen. In de wetenschappelijke literatuur (*cognitive apprenticeship*) en de SBL-competenties (organisatorisch, interpersoonlijk, pedagogisch, didactisch en leerlingen, omgeving, de leraar zelf) is sprake van een organiserend kader. Dit komt in de onderwijsconcepten minder duidelijk naar voren.
De betrokkenen op de scholen kiezen, indien hen wordt gevraagd in het geheel van taken prioriteiten te stellen, vooral voor taken die tot het meer traditionele repertoire behoren, zoals zorgen voor een goed leerklimaat, lesgeven, leerlingen stimuleren en activeren, bijdragen aan persoonlijke ontwikkeling. De in de wetenschappelijke literatuur, beleidsdocumenten en onderwijsconcepten genoemde meer vernieuwende taken, gericht op zelfstandig leren en samenwerken van leerlingen, en het meer inspelen op verschillen tussen leerlingen, worden duidelijk minder gekozen.

Wat zijn de verschillen tussen huidig en gewenst handelen en hoe kunnen deze worden overbrugd?

Op de taakgebieden (hoofdstuk 3) is het verschil tussen huidig en gewenst handelen bij de docenten gemiddeld 0,5 en bij de leerlingen 0,4. Bij beide groepen varieert het verschil per taakgebied tussen 0,4 en 0,6. Het grootste verschil doet zich bij de docenten voor op taakgebied 2 (bijdragen aan de persoonlijke vorming van leerlingen) en bij de leerlingen op de taakgebieden 5 (leerlingen leren zelf opdrachten aan te pakken) en 7 (inspelen op verschillen tussen leerlingen). Op de taken (hoofdstuk 4) is het verschil tussen huidig en gewenst handelen bij beide groepen gemiddeld 0,6 maar dit verschilt nogal tussen de taken: bij de docenten tussen 0,2 en 1,0 en bij de leerlingen tussen 0,2 en 0,7. De taken met het grootste verschil tussen huidig en gewenst handelen zijn bij de docenten de taken 18 (beoordeling), 5 (verbanden tussen vakken) en 8 (vaardigheden) en bij de leerlingen de taken 8 (vaardigheden), 9 (inspelen op verschillen) en 10 (inspelen op verschillen bij opdrachten).

Al met al nemen in het handelen van docenten (hoofdstukken 3 en 4) en hun keuzen en prioriteiten (hoofdstuk 7) meerdere vernieuwende taken, zoals die in wetenschappelijke publicaties, beleidsdocumenten en onderwijsconcepten van scholen aandacht krijgen, een minder grote plaats in dan een aantal meer traditionele taken, en sluiten de leerlingen zich daarbij aan. Op scholen met een sterk vernieuwend onderwijsconcept is dit concept wel duidelijk zichtbaar in de organisatie van het onderwijs qua ruimte en rooster, maar (vooralsnog) minder uitgesproken in het pedagogisch-didactisch handelen van de docenten.

Uit literatuur over vernieuwen en veranderen in het onderwijs en uit een eigen onderzoek van een veranderingsproces op een school (hoofdstuk 8) komt naar voren dat ten minste drie condities een belangrijke rol spelen als schoolleiders in de uitvoering van hun taken ten aanzien van onderwijsverbetering en professionele ontwikkeling van docenten een bepaald verschil tussen huidig handelen en gewenst geacht pedagogisch-didactisch handelen van hun docenten willen overbruggen, namelijk het hebben van een duidelijke visie op de gewenste ontwikkeling, het leiding geven aan het veranderingsproces, en het kunnen benutten van datafeedback. De in hoofdstuk 8 beschreven inzichten aangaande het belang van een duidelijke visie en het belang bij veranderingsprocessen van de onderlinge dialoog, zijn na afronding van het onderzoek in hoofdstuk 7 gebruikt om op basis van de uitkomsten van dat onderzoek aanbevelingen voor de betreffende scholen te formuleren (zie ook Sol & Stokking, 2010). We vatten deze hieronder samen.

Uit de analyse van een aantal onderwijsconcepten blijkt dat deze veelal nog de vorm hebben van een enigszins losse verzameling van vrij veel doelen, middelen en activiteiten. Het verdient aanbeveling de onderdelen ervan meer nadrukkelijk met elkaar te verbinden en te doordenken tot een meer samenhangend en geïntegreerd geheel van uitgangspunten en taakstellingen. De onderwijsvisie van de school kan daardoor winnen aan focus en zeggingskracht en het handelen van docenten meer richting geven. Veel betrokkenen in de scholen hebben bepaalde opvattingen over het gewenste handelen van docenten, maar relateren dat niet primair aan het onderwijsconcept van de school; ze maken persoonlijke, verschillende keuzes. Het verdient aanbeveling om in onderlinge dialoog tussen schoolleiding en docenten het onderwijsconcept, als dat voldoende samenhangend en richtinggevend is, en het gewenste handelen van docenten, meer systematisch met elkaar te verbinden. Dit kan bijdragen aan coherentie, realisme en doelgerichtheid. In de dagelijkse onderwijs-praktijk maken docenten allerlei keuzes in de manieren waarop ze hun taken invullen en vormgeven en in de prioriteiten die ze stellen. Het verdient aanbeveling om via uitwisseling, discussie en overleg tussen docenten de dialoog voort te zetten over de betekenissen die docenten aan hun diverse taken geven en de plaats die deze innemen in hun dagelijkse handelen. Dit kan bijdragen aan verdere professionele ontwikkeling door intensivering en verrijking van denken, delen en samenwerken. Veel docenten werken voor een belangrijk deel op basis van ontwikkelde routines, In een steeds meer veranderende omgeving verdient het aanbeveling om door middel van onderzoek en

reflectie te blijven leren van ervaringen, successen en tegenvallers en zo de eigen praktijk te blijven verbeteren. Juist wanneer wordt gewerkt binnen een coherente en richtinggevende visie aan daarop afgestemde en doordachte taken, is zulk leren en verbeteren op doelgerichte wijze mogelijk.
Wat betreft het belang van het kunnen benutten van datafeedback blijkt uit onze eigen ervaringen met het terugkoppelen van de uitgevoerde onderzoeken naar de betreffende schoolleiders, teamleiders en docenten en uit literatuur over dit onderwerp (zie ook hoofdstuk 8) dat het interpreteren en gebruiken van datafeedback op zich een competentie is die (verdere) ontwikkeling behoeft.

Beperkingen van de uitgevoerde onderzoeken

Doordat de onderzoeken zijn uitgevoerd op verzoek van scholen kon geen sprake zijn van invloed op de samenstelling van de onderzoeksgroep (bijvoorbeeld door het trekken van een aselecte steekproef). Dit beperkt de generaliseerbaarheid van de uitkomsten. Aan de onderzoeken die zijn beschreven in de hoofdstukken 3, 5, 6 en 7 namen echter naast sterk vernieuwende scholen ook een of twee meer traditionele scholen deel en bleken de verschillen tussen beide typen scholen niet groot te zijn.
Ook kon weinig invloed worden uitgeoefend op de samenhang tussen de onderzoeken. Mede dankzij de goede samenwerking met de betreffende schoolleiders kon echter in de verschillende vervolgonderzoeken worden voortgebouwd op de reeds uitgevoerde onderzoeken.
Doordat de vragen vanuit de scholen richtinggevend waren en de condities vanuit het betreffende subsidiekader strikt (per onderzoek een bepaald budget en een maximale looptijd), kon relevante wetenschappelijke literatuur minder dan gebruikelijk invloed hebben op de vraagstelling maar wel op de keuzen van variabelen en de ontwikkeling van instrumenten.

Bijdragen aan de ontwikkeling van de onderwijspraktijk

De uitgevoerde onderzoeken hebben om te beginnen instrumenten opgeleverd: een uitgebreide en een verkorte versie van een instrument om het pedagogisch-didactisch handelen in de breedte in kaart te brengen, een vragenlijst en een observatielijst over resultaatgerichte en procesgerichte feedback, interviewleidraden voor leerlingen over leerzaam beoordelen en over gewenst docenthandelen, en vragenlijsten voor leerlingen over verkregen feedback en verkregen beoordelingen.
De verkregen resultaten zijn per onderzoek op maat naar de deelnemende scholen teruggekoppeld en dit heeft bijgedragen aan onderlinge inhoudelijke discussies en inzichten in de samenhang tussen taken, in relevante verschillen tussen teams en tussen docenten, en in overeenkomsten en verschillen tussen de literatuur, de onderwijs-concepten en het handelen van de docenten.

Tot slot is zicht verkregen op het huidige en gewenste pedagogisch-didactisch handelen van docenten op meer en minder vernieuwende scholen, de verschillen tussen scholen en tussen docenten, en manieren waarop binnen scholen kan worden gewerkt aan verandering en vernieuwing in de richting van gewenst geacht docenthandelen.

De aard van de onderzoeken: op verzoek van en in samenwerking met scholen

De onderzoeken waarvan in dit proefschrift verslag is gedaan, beoogden een bijdrage te leveren aan het uitbreiden en differentiëren van het conceptuele repertoire van de betrokkenen op de scholen en de ontwikkelde instrumenten en terugkoppelde gegevens hebben daaraan, blijkens de verkregen reacties, inderdaad bijgedragen. Op een deel van de scholen is hierdoor ook sprake geweest van bevordering van de onderlinge dialoog tussen de betrokkenen, zowel door inzicht in relevante theoretische begrippen, hun onderlinge relaties en hun concrete uitwerkingen, en het verkregen gespreksrepertoire, als door de teruggekoppelde data over feitelijk en gewenst handelen van docenten. Deze opbrengst is verkregen door in te gaan op de vragen die vanuit scholen zelf zijn gesteld, daarbij relevante wetenschappelijke theorie en onderzoek te betrekken, en bij de verdere uitwerking en uitvoering nauw met de betrokkenen samen te werken. Dit proefschrift maakt hierdoor aannemelijk dat de vaak genoemde kloof tussen onderwijs-onderzoek en onderwijspraktijk niet onoverkomelijk is en dat het slaan van een brug daartussen in ieder geval op deze manier mogelijk is en waardevol is voor de verdere ontwikkeling van het onderwijs op de scholen.

Vragen voor verder onderzoek

Het is de moeite waard om na te gaan in hoeverre de taakgebieden respectievelijk de taken een Raschschaal vormen. Daarmee zou namelijk ondersteuning kunnen worden verkregen van de constatering dat subtaakgebieden c.q. taken verschillen in moeilijkheid en dit zou een rol kunnen spelen bij het ontwikkelen van beleid met het oog op professionele ontwikkeling van docenten.
Het is interessant met het ontwikkelde instrument aanvullende gegevens te verzamelen bij docenten en leerlingen op een groter aantal traditionele scholen en de overallscore en de scores op de afzonderlijke taken te vergelijken met de scores die in de hier beschreven onderzoeken zijn gevonden.
Tevens zou het interessant zijn de scholen en docenten die hebben geparticipeerd in de onderzoeken te blijven volgen en na een bepaalde periode weer opnieuw gegevens te verzamelen om te zien of het onderwijsconcept en het pedagogisch-didactisch handelen van de docenten aan verandering onderhevig is en, indien dat zo is, in welke richting deze zich dan manifesteert.
Verder ligt het voor de hand om in vervolgonderzoek na te gaan in welke mate het pedagogisch-didactisch handelen van docenten zoals dit kan worden vastgesteld met het ontwikkelde instrument (zie de hoofdstukken 3 en 4) samenhangt met de leer-

prestaties van de leerlingen. Hierbij kan worden aangesloten bij de onderzoekstradities naar docenthandelen die ten aanzien van een aantal onderscheiden taken al bestaan, zoals ten aanzien van feedback en beoordeling (zie de hoofdstukken 5 en 6) maar ook bijvoorbeeld ten aanzien van leerklimaat, differentiatie, samenwerking en zelfsturing.
Aanvullend hierbij zijn ook verdere analyses wenselijk van de relaties tussen motivatie, formatieve (feedback) en summatieve beoordeling, en prestatie c.q. verkregen cijfer, inclusief de mate waarin cijfers van docenten kunnen gelden als valide summatieve beoordelingen van prestaties.
Vervolgonderzoek kan zich ook richten op het blootleggen van vakspecifieke aspecten, bijvoorbeeld in onderzoek naar verschillen in feedback tussen verschillende typen taken en de effecten van feedback op het leren van de leerlingen.
Daarnaast is verder onderzoek interessant naar de onderscheiden groepen taken in relatie tot de verschillen tussen scholen, docenten en leerlingen. Het is intrigerend dat, vergeleken met de verdeling van de variantie in de overallschaal bij de leerlingen over de drie niveaus, er bij de takengroep *leerproces* variantie verschuift van school- naar docentniveau, bij de takengroep *ontwikkeling* van docent- naar leerlingniveau, en bij de takengroep *feedback en beoordeling* van school- en docentniveau naar leerlingniveau.
Aanvullend hierop is het relevant nader te analyseren in hoeverre in het pedagogisch-didactisch handelen van docenten nu sprake is van verschillen tussen docenten door verschillen tussen de scholen waar ze werken, of van verschillen tussen scholen door verschillen tussen de docenten die daar werken.
Ook is verder onderzoek wenselijk naar de consequentiële validiteit van het ontwikkelde instrumentarium, in de zin van het gebruik en de impact met het oog op professionele ontwikkeling van docenten en schoolontwikkeling. Dergelijk onderzoek zou zich mede kunnen richten op het versterken van het onderzoeksmatig denken en handelen van schoolleiders, teamleiders en docenten, het leren stellen van relevante en onderzoekbare vragen en de bekwaamheid in het interpreteren en benutten van datafeedback.
Tot slot is een belangrijke vraag voor nader onderzoek hoe de met de uitgevoerde onderzoeken verkregen kennis en inzichten en de concrete resultaten kunnen bijdragen aan de verdere ontwikkeling van het onderwijs op scholen en kunnen worden verankerd in het dagelijks denken en handelen van docenten.

Literatuur

Adriaenssen, F. (red.). (2005). *Parallelle processen. Over het leren van leerlingen, leraren en schoolleiding.* Antwerpen/Apeldoorn: Garant.

Anderman, E.M., Maehr, M.L., & Midgley, C. (1999). Declining motivation after the transition to middle school: Schools can make a difference. *Journal of Research and Development in Education 32* (3), 131-147.

Argyris, C.& Schön, D. (1974) *Theory in practice: Increasing professional effectiveness*. San Francisco: Jossey-Bass.

Askew, S. & Lodge, C. (2000). Gifts, ping-pong and loops – linking feedback and learning. In: Askew, S. (Ed). *Feedback for learning* (pp. 1-17). London: Routledge Falmer.

Bangert-Drowns, R.L., Kulik, C-L.C., Kulik, J.A., & Morgan, M.T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research 61* (2), 213-238.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, New York: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, New York: Prentice-Hall.

Baxter, G.P., & Glaser, R. (1998). Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice 17* (3), 37-45.

Bennebroek Gravenhorst, K. (2002). *Sterke staaltjes van samenwerking. Survey-feedback voor het aanpakken van belemmeringen bij organisatieverandering*. (Dissertatie). Deventer: Kluwer.

Bennebroek Gravenhorst, K. (2006). Een leerproces vormgeven en veranderingen realiseren. Reflectie, interactie en samenwerking door survey-feedback. In: Boonstra, J., & Caluwé, L. de. *Interveniëren en veranderen. Zoeken naar betekenis in interacties* (pp. 289-306). Deventer: Kluwer.

Bennett, Y. (1993). The validity and reliability of assessments and self-assessments of work-based learning. *Assessment & Evaluation in Higher Education,18,* 83-96.

Bereiter, C.B. (1990). Aspects of an educational learning theory. *Review of Educational Research 60* (4), 603-624.

Bergen, T., & Veen, K. van. (2004). Het leren van leraren in een context van onderwijs-vernieuwingen: waarom is het zo moeilijk? *VELON Tijdschrift voor Lerarenopleiders 25* (4), 29-39.

Beijaard, D.(1990). *Teaching as acting.* Doctoral Dissertation. Wageningen: Wageningen Agricultural University.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In: J. Metcalfe, & A.P. Shimanmura (Eds.). *Metacognition. Knowing about knowing* (pp. 185-205). Cambridge: The MIT Press.

Black, P., & Wiliam, D. (2006). The reliability of assessments. In: J. Gardner (Ed.). *Assessment and Learning* (pp. 119-131). Los Angeles: Sage.

Black, P., Wiliam, W. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice 5*, 1, 7-75.

Blumenfeld, P.C., Soloway, E., Marx, R.W., Krajcik, J.S., Guzdal, M., & Palincsar, A. (1991) Motivating Project-Based Learning: Sustaining the Doing, Supporting the Learning. *Educational Psychologist, 26* (3 & 4), 369-398.

Boeije, H. (2008). *Analyseren in kwalitatief onderzoek.* Den Haag: Boom Uitgevers.
Boekaerts, M. (1991). Subjective competence, appraisals and self-assessment. *Learning and Instruction 1,*1-17.
Boesjes-Hommes, R.W. (1974). *De geldige operationalisering van begrippen.* Meppel: Boom en Zoon.
Bolhuis, S. (2003). Towards process-oriented teaching for self-directed learning: a multidimensional perspective. *Learning and Instruction 13,* 327-347.
Bolhuis, S. & Voeten. M.J.M. (2001). Toward self-directed learning in secondary school: what do teachers do? *Teaching and Teacher Education, 17* (7), 837-855.
Bolhuis, S., & Voeten, M.J.M. (2004). Teachers' conceptions of student learning and own learning. *Teachers and Teaching: theory and practice 10* (1), 77-98.
Borsboom, D., Mellenbergh, G.J., & Heerden, Jaap van. (2004). The concept of validity. *Psychological Review 111*, 4, 1061-1071.
Bosker, R.J., Creemers, B.P.M., & Stringfield, S. (1999). Enduring problems and changing conceptions. In: R. J. Bosker, B. P. M. Creemers, & S. Stringfield. *Enhancing educational excellence, equity and efficiency* (pp. 1-9). Dordrecht/Boston/London: Kluwer Academic Press.
Boud, D. & Falchikov, N. (1989). Quantatives studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education 18*, 529-549.
Boxtel, C.van (2000) Sociale interactie die bijdraagt aan begripsontwikkeling. In: Van der Linden, J., Roelofs, E.(red.) *Leren in dialoog* (pp. 65-88). Groningen: Wolters-Noordhoff.
Bransford, J. (2000). The design of learning environments. In: J. D. Bransford, A.L. Brown & R.R. Cocking (Eds.). *How people learn. Expanded Edition.* Washington: National Academy press.
Bransford, J., Derry, S., Berliner, D., & Hammerness, K. (2005). Theories of learning and their roles in teaching. In: L. Darling-Hammond, J., Bransford, P., LePage, K., Hammerness, & H. Duffy, (Eds.), *Preparing teachers for a changing world: what teachers should learn and be able to do* (pp. 40-87). San Francisco: Jossey-Bass.
Brekelmans, M. (1989). *Interpersoonlijk gedrag van docenten in de klas.* Utrecht: W.C.C.
Brekelmans, M. (2010). *Klimaatverandering in de klas.* Utrecht (Oratie). Universiteit Utrecht.
Brekelmans, M., Sleegers, P., & Fraser, B.J. (2000). Teaching for active learning. In: P.R.J. Simons, J.L. van der Linden & T. Duffy (Eds.), *New Learning.* (pp. 227-242). Dordrecht: Kluwer Academic Publishers.
Broekkamp, H. & Hout-Wolters, B. van. (2006). *De kloof tussen onderwijsonderzoek en onderwijspraktijk. Een overzichtsstudie van problemen, oorzaken en oplossingen.* Amsterdam: Vossiuspers UVA.
Broekkamp, H., & Hout-Wolters, B. van. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review 19*, 401-428.
Brok, P. den. (2001). *Teaching and students outcomes*. (Dissertation). Nieuwegein: Budde-Elinkwijk Grafische producties.
Brok, P. den. (2011). *De docent in perspectieven*. (Oratie) Eindhoven: Technische Universiteit Eindhoven.
Brookhart, S.M., & DeVoge, J.G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education 12* (4), 409-425.

Brown, A.L., & Campione, J.C. (1996). Psychological theory and the design of innovative learning environments: on procedures, principles, and systems. In: L. Schaeble & R. Glaser (Eds.), *Innovations in learning New environments for education* (pp. 289-325). Mahwah: Lawrence Erlbaum.

Brown, J.S., Collins, A., & Duguid, P. (1989). *Situated cognition and the culture of learning. Educational Researcher 18* (1), 32-41.

Brown, S., & McIntyre, D. (1993). *Making sense of teaching.* Buckingham: Open University Press.

Busman, L., Klein, T., & Oomen, C. (2006). *Beweging in beeld. Innovatie-monitor_VO.* Utrecht: Schoolmanagers_VO.

Butler, D.L., Lauscher, H.N., Jarvis-Selinger, S., & Beckingham, B. (2004). Collaboration and self-regulation in teachers' professional development. *Teaching and Teacher Education 20*, 435-455.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology 79* (4), 474-482.

Calderhead, J. (1989). Reflective teaching and teaching education. *Teaching and Teacher Education, 7* (1), 1-8.

Calderhead, J. & Gates, P. (Eds.) (1993). *Conceptualizing reflection in teacher development.* London: The Falmer Press.

Campbell, J., Kyriades, L., Muijs, D., & Robinson, W. (2004). *Assessing teacher effectiveness. Developing a differentiated model.* London/New York: Routledge Falmer.

Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education 31* (2), 219-233.

Carnell, E. (2000). Dialogue, discussion and feedback – views of secondary school students on how others help their learning. In: S. Askew (Ed.), *Feedback for learning* (pp. 46-61). London/New York: Routledge Falmer.

Carter, K. (1990). Teachers' Knowledge and learning to teach. In W.R. Houston (Ed.), *Handbook of research on teacher education* (pp. 291-310). New York: Macmillan.

Chi, M.T.H., Siler, S.A., Jeong, H., Yamauchi, T., & Hausmann, R.G. (2001). Learning from human tutoring. *Cognitive Science 25*, 471-533.

Collins, A, Brown, J.S, & Newman, S.E. (1989). Cognitive apprenticeship: Teaching the Crafts of Reading, Writing, and Mathematics. In Resnick, L.B. (Ed.), *Knowing, learning and instruction* (pp. 453-494). Hillsdale: Erlbaum.

Commissie Leraren. (2007). *Leerkracht! Advies van de Commissie Leraren.* Den Haag: OC&W.

Corte, E. de. (2000). Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology. *Learning and Instruction 10*, 249-266.

Corte, E. de, Geerligs, C., Lagerweij, N., Peters, J. & Vandenberghe, R. (1976). *Beknopte didaxologie*. Groningen: Wolters-Noordhoff.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research 58* (4), 438-481.

Crooks, T.J., Kane, M.T. & Cohen, A.S. (1996). Threats to the valid use of assessments. *Assessment in Education: Principles, Policy & Practice 3 (*3), 265-285.

Crossan, M.M., Lane, H.W., & White, R.E. (1999). An organizational learning framework: From intuition to institution. *Academy of Management Review 24* (3), 522-537.
Davison, M.L., & Woods, P.K. (1983). Fitting unidimensional choice models with non-metric multidimensional scaling. *Applied Psychological Measurement 7* (3), 333-340.
Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry 11* (4), 227-268.
Diephuis, R.A.M., & Kasteren, R.M.M. van., (2003). *Scenario's voor een herontwerp van de Basisvorming.* Utrecht: Vereniging voor management in het Voortgezet Onderwijs.
Dijsselbloem, J.R.V.A. (Red.) (2008). *Parlementair Onderzoek Onderwijsvernieuwingen*. Den Haag: SDU-uitgevers.
Dochy, F. (1992). *Assessment of prior knowledge as a determinant for future learning.* Utrecht: Lemma.
Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education. *Studies in Higher Education 24* (3), 331-350.
Dolk, M. L.A.M. (1997). *Onmiddellijk onderwijsgedrag. Over denken en handelen van leraren in onmiddellijke onderwijssituaties.* Utrecht: Universiteit Utrecht.
Doyle, W. (1983). Academic Work. *Review of Educational Research 53* (2), 159-199.
Doyle, W. (1986). Classroom organization and management. In: M.C. Wittrock (Ed.), *Handbook of research on teaching, third edition* (pp. 392-431). New York: Simon & Schuster Macmillan.
Drenth, P.J.D. (1988). *Inleiding in de testtheorie.* Deventer: Van Loghum Slaterus.
Driel, L, van. (2006). *Professionalisering in school: een studie naar verbetering van het pedagogisch-didactisch handelen.* Culemborg: Twin Design bv.
Driscoll, M.P. (2005). *Psychology of learning for instruction*. Boston etc.: Pearson Education.
Duffy, T.M., & Cunningham, D.J. (1996). Constructivism: implications for the design and delivery of instruction. In: D. H. Jonassen (Ed.). *Handbook of research for educational communications and technology* (pp. 170-197). New York: Simon & Schuster MacMillan.
Duijnhouwer, H. (2010). *Feedback effects on students' writing motivation, process, and performance.* (Dissertation). Utrecht: Universiteit Utrecht.
Dwyer, C.A. (1994). Criteria for Performance-Based Teacher Assessements: Validity, Standards, and Issues. *Journal of Personnel Evaluation in Education 8*, 135-150.
Ellis, B. (1968). *Basic concepts of measurement.* Cambridge: University Press.
Elshout-Mohr, M., & Hout-Wolters, B.H.A.M. van. (1995). Actief leren en studeren: acht soorten instructieleerepisodes. *Pedagogische Studiën 72,* 273-299.
Elshout-Mohr, M., Van Hout-Wolters, B., & Broekkamp, H. (1999). Mapping situations in classroom and research: eight types of instructional-learning episodes. *Learning and Instruction 9,* 57-75.
Eraut, M. (1994). *Developing professional knowledge and competence.* London: The Falmer Press.
Evans, E.D., & Engelberg, R.A. (1988). Student perceptions of school grading. *Journal of Research and Development in Education 21* (2), 45-54.
Fenstermacher, G.D. & Richardson, V. (1993). The elicitation and reconstruction of practical arguments on teaching. *Journal of Curriculum Studies, 25* (2), 101-114.

Fishman, B.J., Marx, R.W., Best, S., & Tal, R.T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education 19*, 643-658.
Fiske, S. T. & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw Hill.
Fox, D. (1983). Personal theories of teaching. *Studies in Higher Education 8* (2), 151-163.
Fraser, B.J., Walberg, H.J., Welch, W.W. & Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research 11*, 145-252.
Frederiksen, J.R. & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher, 18,* (9), 27-32.
Frederiksen, J.R., & White, B.Y. (1997). Cognitive facilitation: a method for promoting reflective collaboration. *Proceedings December 1997*, 53-62. CSCL '97
Garcia, T & Pintrich, P.R. (1996) The Effects of Autonomy on Motivation and Performance in the College Classroom. *Contemporary Educational Psychology 21* (4), 477-486.
Gelder, L. van, Oudkerk Pool, T., Peters, J. & Sixma, J. (1973). *Didactische analyse*. Groningen: Wolters- Noordhoff.
Gennip, H. van, & Vrieze, G. (2008). *Wat is de ideale leraar? Studie naar vakkennis, interventie en persoon.* Nijmegen: ITS.
Geijsel, F.P., & Krüger, M.L. (2005). Leren van onderzoek. De benutting van datafeedback ten behoeve van schoolontwikkeling. *Pedagogische Studiën 82*, 327-342.
Gennip, H. van, Vrieze, G. (2008). *Wat is de ideale leraar? Studie naar vakkennis, interventie en persoon.* Nijmegen: ITS.
Gipps, C. (1994). *Beyond testing. Towards a theory of educational assessment.* London: Falmer Press.
Glaser, R. (1990). The reemergence of learning theory within instructional research. *American Psychologist 4* (1), 29-39.
Graesser, A.G., Person, N.K. & Magliano, J.P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology 9*, 495-522.
Gravemeijer, K., & Kirschner, P. (2007). Naar meer evidence-based onderwijs? *Pedagogische Studiën 84*, 463-472.
Greeno, J.G., Collins, A.M., & Resnick, L.B. (1996). Cognition and learning. In: D.C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York: Simon & Schuster MacMillan.
Grimett, P.P. & MacKennon, A.M. (1992). Craft Knowledge and the education of teachers. *Review of Research in Education, 18,* 385-456.
Geijsel, F.P., & Krüger, M.L. (2005). Leren van onderzoek. De benutting van datafeedback ten behoeve van schoolontwikkeling. *Pedagogische Studiën 82*, 327-342.
Grift, W. van de, Wal, M. van der, & Torenbeek, M. (2011). Ontwikkeling in de pedagogisch didactische vaardigheid van leraren in het basisonderwijs. *Pedagogische Studien 88,* 416-432.
Groot, A.D. de. (1971). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen.* Den Haag: Mouton.
Grossman, P., & Schoenfeld, A. (2005). Teaching subject matter. In: L. Darling-Hammond, J., Bransford, P., LePage, K., Hammerness, & H. Duffy, (Eds.). *Preparing teachers for a changing world: what teachers should learn and be able to do* (pp. 201-231). San Francisco: Jossey-Bass.

Guay, F., Vallerand, R.J. & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion 24* (3), 175 -213.
Guskey, T.R. (1995). Professional development in education. In search of the optimal mix. In: T.R. Guskey & M. Huberman. *Professional Development in Education. New Paradigms & Practices* (pp. 114-131). New York, London: Teachers College Press.
Guskey, T.R. (2002a). Professional development and teacher change. *Teachers and Teaching: Theory and Practice 8* (3/4), 381-391.
Guskey, T.R. (2002b). Does it make a difference? Evaluating professional development. *Educational Leadership* March 2002, 45-51.
Guskey, T.R. (2003). Analyzing lists of the characteristics of effective professional development to promote visionary leadership. *NASSP Bulletin 87* (637), 4-20.
Haertel, E. (1985). Construct validity and criterion-referenced testing. *Review of Educational Research 55* (1), 23-46.
Hargreaves, D.H. (1994). The new professionalism: The synthesis of professional and institutional development. *Teaching & Teacher Education 10* (4), 423-438.
Hargreaves, A. (2001). Emotional geographies of teaching. *Teacher College Record 103* (6), 1056-1080.
Hargreaves, A. (2004). Inclusive and exclusive educational change: emotional responses of teachers and implications for leadership. *School Leadership & Management 24* (2), 287-309.
Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal 16* (2), 207-223.
Harlen, W. (2006). The role of assessment in developing motivation for learning. In: J. Gardner (Ed.). *Assessment and Learning* (pp. 61-80). Los Angeles: Sage.
Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education 10* (2), 169-207.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research 77* (1), 81-112.
Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: the problem of communication assessment feedback. *Teaching in Higher Education 6* (2), 269-274.
Hmelo-Silver. C.E.. Golan Duncan. R. & Chinn. A.A.(2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner. Sweller. and Clark. *Educational Psychologist. 41*(2). 99-107.
Hodgen, J., & Marshall, B. (2005). Assessment for learning in English and mathematics: a comparison. *The Curriculum Journal 16* (2), 153-176.
Hofman, W., Hofman, R., Dijkstra, B., De Boon, J., & Meeuwisse, M. (2007). *Innovaties in het voortgezet onderwijs. Een verkenning van innovaties en effecten in het voortgezet onderwijs*. Groningen/Rotterdam: GION/RISBO.
Homan, T. (2005). O*rgansiatiedynamica. Theorie en praktijk van organisatieverandering.* Den Haag: Sdu Uitgevers.
Hounsell, D., McCune, V., Hounsell, J., & Litjens, J. (2008). The quality of guidance and feedback to students. *Higher Education Research & Development 27* (1), 55-67.
Houtveen, Th., & Reezigt, G. (2000). *Succesvol adaptief onderwijs. Handreikingen voor de praktijk.* Deventer: Kluwer.

Huberman, M. (1990). Professional Careers and Professional Development. In : Guskey, T.R. & Huberman, M. (Eds). (1990). *Professional Development in Education. New Paradigms & Practices* (193-224). New York: Teachers College Press.
Ilgen, D.R., Davis, C.A. (2000). Bearing bad news: reactions to negative performance feedback. Applied Psychology: *An International Review 49* (3), 550-565.
Ilgen, D.R., Fisher, C.D., & Taylor, S. (1979). Consequences of individual feedback on behaviour in organizations. *Journal of Applied Psychology 64* (4), 349-371.
Imants, J. (2003). Two basic mechanisms for organisational learning in schools. *European Journal of Teacher Education* 26 (3), 293-311.
Inspectie van het onderwijs. (2006). *Leerlingen: Boeit 't. Nieuwe vormen van leren.* Utrecht: Inspectie van het onderwijs.
Inspectie van het onderwijs. (2009). *Toezichtkader 2009*. Utrecht: Inspectie van het onderwijs.
Inspectie van het Onderwijs (2011). *Toezichtkader 2011.* Utrecht:. Inspectie van het onderwijs.
Kagan, D.M. (1992). Implications of research on teacher belief. *Educational Psychologist, 27* (1), 65-90.
Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin 112* (3), 527-535.
Kelchtermans, G. (2007). Capturing the multidimensionality of teacher professionalism: broad and deep reflection. In: J. Van Swet, P.Ponte & B. Smit (Eds.), *Postgraduate programmes as platform: A research-led approach.* (pp. 97-109). Rotterdam: Sense Publishers.
King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice 41* (1), 33-39.
Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experimental, and inquiry-based teaching. *Educational Psychologist 4* (12), 75-86.
Klein, T., Oomen, C., Van der Linden, J., & Onderdenwijngaard, J. (2008). *Beweging in beeld III. Feiten en verhalen over innovatie in het voortgezet onderwijs 2008.* Utrecht: VO-raad.
Kluger, A.N., & Denisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin 119*, 254-284.
Kock, A. de, Sleegers, P., & Voeten, M.J.M. (2004). New learning and the classification of learning environments in secondary education. *Review of Educational Research 74* (2), 141-170.
Larson, J.(1984). The performance feedback process: A preliminary model. *Organizational Behavior and Human Performance 33*, 42-76.
Larsson, S. (1983). Paradoxes in teaching. *Instructional Science 12,* 355-365.
Lasky, S. (2005). A sociocultural approach to understanding teacher identity, agency and professional vulnerability in a context of secondary school reform. *Teaching and Teacher Education 21*, 899-916.
Levin, B. (2000). Putting students in the centre of educational reform. *Journal of Educational Change 1,* 155-172.

Levy, J., Den Brok, P., Wubbels, Th. & Brekelmans, M. (2002). Students' perceptions of interpersonal aspects of the learning environment. *Learning Environments Research 6,* 5-36.

Lűdtke, O., Robitzsch, A., Trautwein, U. & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modelling. *Contemporary Educational Psychology, 34* (2), 120-131.

Luo, M. (2008). Structural Equation Modeling for High School Principals' Data-Driven Decision Making: An Analysis of Information Use Environments. *Educational Administration Quarterly 44* (5), 603-643.

Luttenberg, J. (2000). *Wat te doen? Een onderzoek naar de normatieve professionaliteit van docenten.* Leuven- Apeldoorn: Garant.

Luttenberg, J. Imants, K., Veen, K. van, & Carpay, T. (2009). Zoeken naar samenhang. De rol van "betekenisafstemming" in de interactie tussen docent en vernieuwing. *Pedagogische Studiën, 84* (6), 443-458.

Lyons, N. (1990). Dilemmas of knowing: ethical and epistemological dimensions of teachers' work and development. *Harvard Educational Review 60* (2), 159-180.

Maanen, M. van. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry 6* (3), 205-228).

Mainhard, M.T. (2009). *Time Consistency in Teacher-Class Relationships.* Doctoral Dissertation. Utrecht: Ipskamp Drukkers B.V.

Mayer, R.E. (2004a). Should there be a three-strikes rule against pure discovery learning? *American Psychologist 59* (1), 14-19.

Mayer, R.E. (2004b). Teaching of subject matter. *Annual Review of Psychology 55,* 715-744.

Meisels, S.J., Atkins-Burnett, S., Xue, Y., DiPrima Bickel, D., & Son, S.H. (2003). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores. *Education Policy Analysis Archives 11* (9), 1-18.

Merril, M.D. (1971). Fundamentals of Instructional Design. In M.D. Merill (Ed.). *Instructional Design: readings.* New Yersey: Prentice-Hall.

Merill, M.D. (2002). First Principles of instruction. *Educational Research and Development 50* (3), (43-59).

Messick, S. (1984). The psychology of educational measurement. *Journal of Educational Measurement 21* (3), 215-237.

Messick, S. (1995a). Standards of validity and the validity of standard in performance assessment. *Educational Measurement. Issues and Practice, 14* (4), 5-8.

Messick, S. (1995b). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist 50 (*9), 741-749.

Ministerie van OC&W. (1993). *Vitaal Leraarschap. Beleidsreactie naar aanleiding van het rapport "Een beroep met perspectief" van de Commissie Toekomst Leraarschap.* Den Haag: OC&W.

Ministerie van OC&W. (2004). *Vooruit! Innoveren in het onderwijs.* Beleidsnotitie. Den Haag: OC&W.

Ministerie van OC&W. (2008a). *Monitor 2005-2008. Blijvend in beweging. Vier jaar onderbouwontwikkeling.* Den Haag: OC&W.

Ministerie van OC&W. (2008b). *Krachtig Meesterschap. Kwaliteitsagenda voor het opleiden van leraren, 2008-2011*. Den Haag: OC&W.
Mory, E.H. (2004). Feedback research revisited. In: Jonassen, D. (Ed.), *Handbook of Research on Educational Communication and Technology* (pp. 745-783). Mahwah: Lawrence Erlbaum.
Moss, P.A. (1992). Shifting Conceptions of Validty in Educational Measurement: Implications for performance Assessment. *Review of Educational Research, 62* (3), 229-258.
Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist 22* (3), 155-175.
Onderwijsraad. (1998). *Advies "Toekomst leraren" aan de minister van OC&W*. Den Haag: Onderwijsraad.
Onderwijsraad. (2005a). *Advies AMvB Bekwaamheidseisen Onderwijspersoneel*. Den Haag: Onderwijsraad.
Onderwijsraad. (2005b). *Advies "Vernieuwing Basisvorming"*. Den Haag: Onderwijsraad.
Onderwijsraad. (2007). *Leraarschap is eigenaarschap. Advies*. Den Haag: Onderwijsraad.
Oostdam, R., Peetsma, T., & Blok, H. (2007). *Het nieuwe leren in basisonderwijs en voortgezet onderwijs nader beschouwd: een verkenningsnotitie voor het Ministerie van Onderwijs, Cultuur en Wetenschap*. Amsterdam: SCO-Kohnstamm Instituut.
Opfer, V.D., & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research 81* (3), 376-407.
Oshige, M. (2003). *Exploring task understanding in self-regulated learning: Task understanding as a predictor of academic succes in undergraduate students*. (Masterthesis). University of Victoria.
Paris, S.G., Lawton, T.A., Turner, J.C., & Roth, J.L. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher 20* (5), 12-20.
Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: district and school connections in data-driven decision-making. *School Leadership & Management 29* (5), 477-494.
Pellegrino, J.W., Baxter, G.P., Glaser, R. (1999). Addressing the "two disciplines" problem: Linking theories of cognition and learning with assessment and instructional practice. *Review of Research in Education 24* (1), 307-353.
Petegem, P. van, & Vanhoof, J. (2004). Een pleidooi voor geïnformeerde schoolontwikkeling. *Impuls 34* (4), 192-197.
Pol, J. van de, Volman, M., & Beishuizen. (2010). Scaffolding in teacher-student interaction : A decade of research. *Educational Psychology Review 22*, 271-296.
Pratte, R., & Rury, J.L. (1991). *Teachers, professionalism, and craft. Teacher College Record 93* (1), 59-72.
Prawat, R.S. (1992). Teachers' beliefs about teaching and learning: a constructivist perspective. *American Journal of Education 100* (3), 354-395.
Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science 28*, 4-13.
Reed, J., & Stoll, L. (2000). Promoting organisational learning in schools – the role of feedback. In: Askew, S. (Ed.). *Feedback for learning* (pp. 127-143). London/New York: Routledge Falmer.
Reynolds, M. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research, 62,* 1-35.

Reynolds, S., Martin, K., & Groulx, J. (1996). Patterns of understanding. *Educational Assessment 3* (4), 363-371.
Ridder, J.M.M. van de , McGaghie W.C., Stokking K.M., & Ten Cate Th.J. (2008). What is feedback in clinical education? *Medical Education 42.* 189-197.
Roelofs, E. C., Nijveldt, M. & Beijaard, D.(2008). Ontwikkeling van een zelfbeoordelings-instrument voor docentcompetenties. *Pedagogische Studiën 85,* 319-341.
Runhaar, P., Sanders, K., & Sleegers, P. (2009). *De school als ontwikkelplek voor leraren. Een literatuuronderzoek naar organisatiefoactoren die implementatie van nieuwe onderwijsconcepten bevorderen.* Enschede: Universiteit Twente.
Russell, T. & Mundy, H. (1991). Reframing: The role of experience in developing teachers' professional knowledge. In D.A. Schön (Ed), *The reflective turn. Case studies in and on educational practice* (pp. 164-187). New York: Teachers College Press.
Ryan, R.M. & Deci, E.L. (2000). Instrinsic and Extrinsic Motivation: Classic Definitions and New Directions. *Contempory Educational Psychology 25* (1), 54-67.
Sadler, (1989). Formative assessment and the design of instructional systems. *Instructional Science* 18, 119-144.
Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science 18,* 119-144.
Sadler, D.R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice 5* (1), 77-85.
Sadler, D.R. (2009a). Transforming holistic assessment and grading into a vehicle for complex learning. In: J. Joughin (Ed.). *Assessment, learning and judgement in higher education* (pp. 45-63). Springer Science & Business Media.
Sadler, D.R. (2009b). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education 34* (2), 159-179.
Sadler, D.R. (2010a). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education 35* (5), 535-550.
Sadler, D.R. (2010b). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education 35* (6), 727-743.
Schaaf, M.F. van der, Stokking, K.M., & Verloop, N. (2003). Developing performance standards for teacher assessment by policy capturing. *Assessment & Evaluation in Higher Education, 28* (4), 395-410.
Scheerens, J. (2007). *Een overzichtsstudie naar school- en instructie-effectiviteit*. Enschede: Universiteit Twente. Enschede: University of Twente.
Schmidt, H.G., Loyens, S.M.M., Gog, T, van & Paas, F. (2007). Problem-Based Learning is Compatible with Human Cognitive Architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42* (2), 91-97.
Schmidt, M., & Datnow, A. (2005). Teachers' sense-making about comprehensive school reform: The influence of emotions. *Teaching and Teacher Education 21*, 949-965.
Schön, D.A. (1983). *The reflective practioner. How professionals think in action.* New York: Basis Books.
Schön, D.A. (1987). *Educating the reflective practitioner.* San Franscisco: Jossey-Bass.
Scholtes, H.H.M. (2004). *Schoolontwikkeling: noodzaak, voorbeelden en handreikingen*. Zwolle: Taakgroep Vernieuwing Basisvorming.
Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research 77* (4), 454-499.

Severiens, S. (2005). Nederlandse vertaling van items van de MSLQ, opgevraagd naar aanleiding van: Blom, S., Severiens, S., Broekkamp. H. & Hoek, D. (2005). *Zelfstandig leren van allochtone en autochtone leerlingen in het studiehuis.* Amsterdam: ILO.

Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research 78* (1), 153-189.

Shuell, T.J. (1993). Toward An Integrated Theory of Teaching and Learning. *Educational Psychologist, 28* (4), 291-311.

Shuell, T. J. (1996). Teaching and learning in a classroom context. In: D. C. Berliner & R.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 726-764). New York: Simon & Schuster MacMillan.

Shulman, L.S. (1986). Paradigms and research programs in the study of teaching: a contemporary perspective. In: M.C. Wittrock (Ed.), *Handbook of research on teaching* (third edition) (pp. 3-36). New York: Simon & Schuster Macmillan.

Shulman, L.S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review 57* (1), 1-22.

Smith, P.L., & Ragan, T.J.. (1993). Designing instructional feedback for different learning outcomes. In: J.V. Dempsey, G.C. & Sales (Eds.). *Interactive instruction and feedback* (pp. 75-104). Englewood Cliffs: Educational Technology Publications.

Sol, Y.B., Stokking, K.M. (2007). *Tutoren in het voortgezet onderwijs. Een pilot met vijfdejaars vwo-leerlingen die tweedejaars havo/vwo-leerlingen begeleiden bij wiskunde en Nederlands.* Utrecht: Universiteit Utrecht.

Sol, Y.B., Stokking, K.M. (2008a). *Keuzewerktijd en verandering van leercultuur.* (nr. 70 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Sol, Y.B., & Stokking, K.M. (2008b). *Het handelen van docenten in scholen met een vernieuwend onderwijsconcept.* (nr. 71 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Sol, Y.B., Stokking, K.M. (2008c). *Leerlingenparticipatie in het VO. Vormen, processen, effecten.* (nr. 72 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Sol, Y.B. & Stokking, K.M. (2009). *Mondelinge feedback bij zelfstandig werken. Interactie tussen docenten en leerlingen in het VO.* (nr. 74 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Sol, Y.B. & Stokking, K.M. (2010). *Gewenst docenthandelen in het voortgezet onderwijs. Vernieuwingen, perspectieven, keuzes en gronden.* (nr. 78 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Sol, Y.B., Stokking, K.M. (2011). *Leren van beoordelingen. Leerlingen en docenten in het voortgezet onderwijs over maken en beoordelen van toetsen en opdrachten.* (nr. 83 reeks Kortlopend Onderwijsonderzoek). Utrecht: Universiteit Utrecht.

Solzbacher, C. (2006). Improving learning competence in schools – what relevance does empirical research in this area have for teacher training? *European Journal of Teacher Education 29* (4), 533-544.

Spillane, J.P., Halverson, R., & Diamond, J.B. (2004). Towards a theory of leadership practice: a distributed perspective. *Journal of Curriculum Studies 36* (1), 3-34.

Spillane, J.P., Reiser, B.J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research 72* (3), 387-431.

Stevens, L., Beekers, P, Evers, M., Wentzel, M., & Werkhoven, W. van. (2004). *Zin in school* Meppel: Giethoorn ten Brink.

Stichting Beroepskwaliteit Leraren (SBL). (2004). *In bekwame handen. Bekwaamheidseisen voor leraren*. (CD-rom).
Stokking, K.M. (1984). *Interpreteren en evalueren. Methodologie rondom de uitkomsten van onderzoek*. Deventer: Van Loghum Slaterus.
Stokking, K.M. (2001). Leren van vaardigheden. In: A. Wald & J. van der Linden (Red.). *Leren in perspectief* (pp. 15-29). Leuven/Apeldoorn: Garant.
Stokking, K., Schaaf, M. van der, Jaspers, J., Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal 30* (1), 93-116.
Stokking, K. & Sol, Y. (2010). *Van participatie naar dialoog. Voorwaarden voor het realiseren van meer leerlingbetrokkenheid.* Utrecht: Universiteit Utrecht.
Stoof, A., Martens, R.L., Van Merriënboer, J.J.G., & Bastiaens, T.J. (2002). The boundary approach of competence: a constructivist aid for understanding and using the concept of competence. *Human Resource Development Review 1*, 345-365.
Straub, R. (1996). The concept of control in teacher response: Defining the varieties of "directive" and "facilitative" commentary. *College Composition and Communication 47* (2), 223-251.
Straub, R. (1997). Students' reactions to teacher comments: An exploratory study. *Research in the Teaching of English 31* (1), 91-119.
Stuurgroep Beroepskwaliteit Leraarschap. (1997). *Tekenen voor kwaliteit. Advies inzake een beroepsstandaard, register en beroepsgroep van leraren*. Groningen: Wolters-Noordhoff
Sung, Y.T., Chang, K.E., Chang, T.H., & Yu, W.C. (2010). How many heads are better than one? The reliability and validity of teenagers' self- and peer assessments. *Journal of Adolescence 33*, 135-145.
Swaffield, S., & MacBeath, J. (2006). Embedding learning how to learn in school policy: the challenge for leadership. *Research Papers in Education 21* (2), 201-215.
Sweller, J, Kirschner, P.A. & Clark, R.E.(2007). Why Minimal Guided Teaching Techniques Do Not Work: A Reply to Commentaries. *Educational Psychologist, 42* (2), 115-121.
Tartwijk, J. van. (2011). *Van onderzoek naar onderwijs, of de kunst van de toepassing.* (Oratie). Utrecht: Universiteit van Utrecht.
Teurlings, C., Wolput, B. van, & Vermeulen, M. (2006). *Nieuw leren waarderen. Een literatuuronderzoek naar effecten van nieuwe vormen van leren in het voortgezet onderwijs*. Utrecht: Schoolmanagers_VO.
Torrance, H. & Pryor, J. (1998). *Investigating formative assessment. Teaching, learning and assessment in the classroom*. Maidenhead/ Philadelphia: Open University Press.
Tweede Fase Adviespunt. (2001). *De implementatie van de vernieuwingen in de tweede fase van havo en vwo. Eindverslag van de schoolbezoeken van het Tweede Fase Adviespunt.* Den Haag: Tweede Fase Adviespunt.
Valcke, M. (2007). *Onderwijskunde als ontwerpwetenschap. Een inleiding voor ontwikkelaars van instructie en voor toekomstige leerkrachten.* Gent: Academia Press.
Veen, K. van. (2011). Het niveau en de kwaliteit van leraren in het basisonderwijs en voortgezet onderwijs: wat is het probleem? *Pedagogische Studien 88*, 433-442.
Veen, K. van, Zwart, R., Meirink, J., & Verloop, N. (2010). *Professionele ontwikkeling van leraren. Een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren.* Leiden: ICLON.

Veenhoven, J. (2004). *Begeleiden en beoordelen van leerlingonderzoek. Een interventiestudie naar het leren ontwerpen van onderzoek in de tweede fase bij Aardrijkskunde (dissertatie)*. Utrecht: Universiteit Utrecht.
Veenman, S. (1992). Effectieve instructie volgens het directe instructiemodel. *Pedagogische Studiën 69,* 242-269.
Verhaeghe, G., Vanhoof, J., Vakcke, M. & Van Petegem, P. (2011). Effecten van ondersteuning bij schoolfeedbackgebruik. *Pedagogische Studiën, 88* (2), 90-106
Verloop, N., & Lowyck, J. (Red.). (2009). *Onderwijskunde. Een kennisbasis voor professionals*. Groningen/Houten: Wolters-Noordhoff.
Verloop, N. & Wubbels, Th. (1994). Recente ontwikkelingen in het onderzoek naar leraren en lerarenopleiding. *Pedagogische Studiën 71* (3), 168-186.
Vermunt, J.D., & Endedijk, M.D. (2011). Patterns in teacher learning in different phases of the professional career. *Learning and Individual Differences 21,* 294-302.
Vermunt, J.D. & Verloop, N. (1999). Congruence and friction between learning and teaching.*Learning and Instruction 9,* 257-280.
Veugelers, W., & Volman, M. (Red.). (2007). De pedagogische functie van het onderwijs. Themanummer Pedagogische Studiën. *Pedagogische Studiën 84* (2), 75-152.
Volman, M.L.L. (2006). *Jongleren tussen traditie en toekomst. De rol van docenten in leergemeenschappen*. (Oratie). Amsterdam: VU.
Volman, M. (2011). *Kennis van betekenis. Betrokkenheid als kwaliteit van leerprocessen en leerresultaten.* Amsterdam (Oratie). Universiteit van Amsterdam.
VO-raad. (2008). *Beweging in beeld III. Feiten en verhalen over innovatie in het voortgezet onderwijs 2008.* Utrecht: VO-raad.
Wang, M.C., Haertel, G.D., & Walberg, H.J. (1993). Toward a knowledge base for school learning. *Review of Educational Research 63* (3), 249-294.
Waslander, S. (2007). *Leren over innoveren.* Utrecht: VO-raad.
Webb, N. (1997). Assessing Students in Small Groups. *Theory into Practice 36* (4), 205-213.
Webb, N.M. (2009). The teachers' role to promote collaborative dialogue in the classroom. *British Journal of Educational Psychology 79* (1), 1-28.
Weick, K.E. (1995). *Sensemaking in organizations*. Thousand Oaks: Sage Publications.
Weick, K. E., & Sutcliffe, K.M. (2005). Organizing and the Process of Sensemaking. *Organization Science 16* (4), 409-421.
Wikeley, F. (2000). Learning from research. In: Askew, S. (Ed.). *Feedback for learning* (pp. 97-109). London/New York: Routledge Falmer.
Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal 22* (5), 537-548.
Wilson, B.G., & Cole, P. (1996). Cognitive teaching models. In: D.H. Jonassen (Ed.). *Handbook of research for educational communications and technology* (pp. 601-621). New York: Simon & Schuster MacMillan.
Windschitl, M. (2002). Framing Constructivism in Practice as the Negotiation of Dilemma´s: An Analysis of the Conceptual, Pedagogical, Cultural and Political Challenges Facing Teachers. *Review of Educational Research, 72* (2), 131-175.
Wittwer, J., Nickles, M., & Renkl, A. (2006). Improving human tutoring by improving tutor-generated explanations. In: G. Clarebout & J. Elen (Eds.). *Avoiding simplicity confronting complexity (pp. 349-359).* Rotterdam: Sense Publications.

Wohlstetter, P., Datnow, A., & Park, V. (2008). Creating a system for data-driven decision-making: applying the principal-agent framework. *School Effectiveness and School Improvement 19* (3), 239-259.
Wood, R. (1987). Aspects of the competence-performance distinction: Educational, psychological and measurement issues. *Journal of Curriculum Studies, 19* (5), 409-424.
Woolfolk, A., Hughes, M., & Walkup, V. (2008). *Psychology in Education*. Harlow: Pearson.
Wubbels, Th., Brekelmans, M. & Hooymayers, H.P. (1991). Interpersonal Teacher Behavior in the Classroom. In B.J. Fraser & H. Walberg (Eds.), *Educational Environments* (pp. 141-160). Londen: Pergamon Press.
Zimmerman, B. J. (2000). Attaining self-regulation: a social cognitive perspective. In: M. Boekaerts, P. P. Pintrich & M. Zeidner. *Handbook of self-regulation* (pp.14-39). San Diego: Academic Press.

Samenvatting

Schoolleiders in het voortgezet onderwijs zien zich, in antwoord op maatschappelijke ontwikkelingen, overheidsbeleid, wetenschappelijke inzichten en veranderingen bij hun leerlingen, steeds meer genoodzaakt in hun scholen vernieuwingen te initiëren. Vaak gaat het daarbij om gewenst geachte veranderingen in het pedagogisch-didactisch handelen van de docenten. In een periode van vier jaar hebben de schoolleiders van acht scholen, in wisselende combinaties, vijf aanvragen ingediend voor kortlopende onderwijsonderzoeken (met een maximale looptijd van één jaar). In aanvulling op het eerste onderzoek, waarin de ontwikkeling van een instrument voor het in kaart brengen van het pedagogisch-didactisch handelen van docenten centraal stond, is door de auteur, in samenwerking met vier van de scholen, in een zesde onderzoek van dat instrument een verkorte versie gemaakt en in de praktijk nader gevalideerd. Dit proefschrift beschrijft (in de hoofdstukken 3 t/m 8) de opzet en de voornaamste resultaten van deze zes studies.

De ontwikkeling van een instrument voor het in kaart brengen van het pedagogisch-didactisch handelen van docenten (hoofdstuk 3) begon met de afbakening van de begrippen "handelen" en "pedagogisch-didactisch" en de uitwerking van de interne structuur daarvan. Op basis van wetenschappelijke literatuur, landelijke beleids-documenten, onderwijsconcepten van een zestal scholen en bevraging van een aantal docenten, kon een indeling worden gemaakt van het pedagogisch-didactisch handelen in 15 taakgebieden, op verzoek van de schoolleiders aangevuld met een extra taak-gebied over activiteiten voor professionele ontwikkeling.
Met het oog op het in de scholen (met opleidingen variërend van vmbo tot gymnasium) beoogde gebruik en de gewenste kwaliteit van het instrument en de daarmee te verkrijgen resultaten, is gekozen voor de vorm van een schriftelijke vragenlijst voor docenten en leerlingen waarin elk taakgebied is vertegenwoordigd door een aantal uitspraken over het handelen van een docent op dat taakgebied en wordt gevraagd naar het huidige handelen van de docent en naar het (volgens de invuller) gewenste handelen. De na een try-out bijgestelde versie is ingevuld door 141 docenten en 900 leerlingen op zes scholen. De uitspraken per taakgebied bleken steeds een voldoende betrouwbare schaal te vormen. De betreffende schalen vormen samen een zeer betrouwbare schaal die kan worden opgevat als maat voor het repertoire van het pedagogisch-didactisch handelen van een docent. Bij zowel de schalen van de taakgebieden als de overallschaal bleek dat op basis van de gegevens van de leerlingen in relatief hoge mate betrouwbaar onderscheid kan worden gemaakt tussen docenten. Factoranalyses per taakgebied resulteerden bij zes taakgebieden in het onderscheiden van twee subtaakgebieden die elk eveneens een voldoende betrouwbare schaal vormen en die kunnen worden geïnterpreteerd als een minder respectievelijk een meer vernieuwende invulling van het betreffende taakgebied.

Op verzoek van de schoolleiders is van het ontwikkelde instrument een verkorte versie gemaakt. Daarin zijn de gevonden subtaakgebieden apart opgenomen en is het aantal items van de resulterende 22 taken teruggebracht tot gemiddeld vijf uitspraken per taak. Deze verkorte versie (PDH genoemd: Pedagogisch-Didactisch Handelen) is web-based toegankelijk gemaakt. Op data verkregen van 49 docenten en 502 leerlingen is ter nadere validering van de PDH een aantal analyses uitgevoerd (hoofdstuk 4). De betrouwbaarheid van de schalen voor de afzonderlijke taken (na verwijdering van enkele items met een te lage item-rest correlatie) en de overallschaal bleek acceptabel tot goed en hun onderscheidingsvermogen substantieel, maar de verkregen waarden waren iets lager dan bij de uitgebreide versie met taakgebieden. Voor het inkorten van de schalen (van gemiddeld 10 naar gemiddeld 5 items) en de daardoor minder uitgebreide operationalisering is dus een zekere prijs betaald.
In een op andere scholen uitgevoerd onderzoek waarbij de auteur was betrokken, zijn 136 docenten bevraagd naar hun oordeel over het belang van de 22 taken en op basis van deze gegevens konden met behulp van factoranalyse vier groepen taken worden onderscheiden. Deze bleken in de onderhavige PDH-data elk een betrouwbare schaal te vormen, gericht op respectievelijk de ontwikkeling van leerlingen, het (vak-inhoudelijke) leerproces, feedback en beoordeling, en samenwerking en professionele ontwikkeling.
Door 17 docenten en 188 leerlingen is naast de PDH ook de VIL ingevuld (Vragenlijst Interpersoonlijk Leraarsgedrag). De schalen van beide instrumenten (taken, groepen taken en overallscore van de PDH en sectoren en dimensies van de VIL) hangen, volgens een bepaald patroon, gematigd met elkaar samen, waaruit blijkt dat beide instrumenten elk een eigen dimensie in het handelen van docenten vertegenwoordigen (pedagogisch-didactisch respectievelijk interpersoonlijk). Een mogelijke interpretatie van de gevonden samenhangen is dat (de kwaliteit van) het interpersoonlijk handelen een voorwaarde vormt om het pedagogisch-didactisch handelen tot zijn recht te kunnen laten komen.
Analyses van verschillen tussen scholen en verbanden tussen de PDH-scores en kenmerken van docenten en leerlingen leverden op dat het pedagogisch-didactisch handelen van docenten, ook op de betreffende vernieuwende scholen, in hoge mate tussen docenten varieert. Blijkens multilevelanalyse ligt rond 25% van de variantie in de PDH-totaalscore in de data van de leerlingen op docentniveau en is de variantie op schoolniveau (rond 10%) niet significant. De scores hangen niet sterk samen met docentkenmerken zoals geslacht, leeftijd en ervaring. Dat geen duidelijke relatie is gevonden tussen de scores op de PDH en de leeftijd en ervaring van de docenten kan komen doordat in dit onderzoek relatief weinig oudere en ervaren docenten waren betrokken. Het kan ook komen doordat de docenten in dit onderzoek werkzaam zijn in een context van verandering en vernieuwing. Ook docenten die al de nodige ervaring hebben, kunnen daardoor opnieuw in een leerproces betrokken raken, met gevolgen voor hun pedagogisch-didactisch handelen. Aangezien een deel van de taken van de

PDH betrekking heeft op relatief vernieuwend handelen dat door veel docenten wordt ervaren als niet gemakkelijk, is deze mogelijkheid niet denkbeeldig.

Op verzoek van een aantal schoolleiders is een vervolgonderzoek uitgevoerd naar een van de taakgebieden, namelijk het geven van mondelinge feedback als leerlingen zelfstandig aan het werk zijn (hoofdstuk 5). Uit de literatuur zijn ruim 20 potentieel relevante aspecten, inhouden en vormen van feedback gedestilleerd (procesgericht en resultaatgericht). Op vier scholen (met opleidingen variërend van vmbo tot gymnasium) zijn de schoolleiders en in totaal 12 docenten (drie per school, van diverse vakken) mondeling en schriftelijk bevraagd over hun opvattingen, intenties en praktijken op dit taakgebied. Per docent is uit leerjaar 3 of 4 een groep leerlingen schriftelijk bevraagd over hun motivatie voor het maken van taken voor het betreffende vak bij de betreffende docent en over de praktijk van het geven van feedback door deze docent. Bij elke docent is een les aan de betreffende groep leerlingen opgenomen en geobserveerd waarin de leerlingen zelfstandig werkten aan taken. Aan het eind van de les zijn de leerlingen schriftelijk bevraagd over de van de docent verkregen feedback. Met elke docent is de gemaakte opname bekeken en besproken (stimulated recall interview). Tot slot ontving elke docent een persoonlijke terugrapportage met een overzicht van alle verkregen resultaten en is met hem of haar daarover een eindgesprek gevoerd.
De opvattingen en intenties van de docenten bij het geven van feedback bleken goeddeels overeen te komen met de aanwijzingen in de literatuur, en hun handelen bleek, zowel volgens de vragenlijsten als volgens de observaties, daarmee grotendeels in overeenstemming: zij beschikten over een breed en gedifferentieerd repertoire en pasten dit ook toe, en hun betreffende handelen werd door hun leerlingen in vrij hoge mate gewaardeerd. In tegenstelling tot de verwachting van de schoolleiders en ook van een aantal docenten, bleek dit niet samen te hangen met de leeftijd of ervaring van de docent.
De docenten lijken primair gericht op de leerlingen. Geven van feedback staat voor hen in het teken van persoonlijk contact, vertrouwen en aandacht. Feedback moet voor alles positief en constructief zijn. Het benoemen van wat (nog) niet goed is en het beoordelen van een prestatie aan de hand van een duidelijke standaard lijkt verhoudingsgewijs iets minder te gebeuren. Feedback blijkt meer een reactie op een probleem van de leerling, die iets niet begrijpt en daardoor niet verder kan. De feedback is er dan op gericht dat de leerling het wel begrijpt en weer verder kan. Gezien vanuit de literatuur draagt dit positief bij aan de acceptatie van de feedback maar minder aan het bevorderen van zelfevaluatie door de leerlingen.
De breedte van het repertoire van het volgens de leerlingen gebruikelijke feedback-gedrag van de docent en de motivatie van de leerlingen voor het maken van taken voor het vak bij de docent hangen, blijkens de analyse van een daartoe opgesteld *structural equations model*, onderling significant en substantieel positief samen. Dit kan worden opgevat als een bevestiging van de in de literatuur beschreven invloed van feedback op

motivatie, maar het kan ook zijn dat gemotiveerde leerlingen het pedagogisch-didactisch handelen van hun docenten beter waarnemen en onthouden. Een nog andere interpretatie is dat docenten aan meer gemotiveerde leerlingen meer (diverse) feedback geven, mogelijk omdat hun werk daartoe aanleiding geeft en/of zulke leerlingen hen daartoe meer stimuleren.

Een tweede vervolgonderzoek op verzoek van een aantal schoolleiders betrof het taakgebied beoordelen, gericht op de vraag hoe beoordelingen voor leerlingen leerzaam kunnen zijn (hoofdstuk 6). Uit de literatuur is een aantal kenmerken van beoordelingen gedestilleerd die naar verwachting kunnen bijdragen aan motivatie en leren van leerlingen. Op drie scholen zijn in totaal negen docenten (drie per school: Engels, aardrijkskunde, wiskunde) mondeling en schriftelijk bevraagd over hun beoordelingspraktijk en zijn daarnaast leerlingen (per docent een klas uit leerjaar 3 havo/vwo) schriftelijk bevraagd over hun motivatie voor het maken van taken voor het vak van de docent en over de beoordelingspraktijk van de docent. Per docent zijn twee of drie taken in het onderzoek betrokken die de leerlingen in de loop van enkele maanden dienden te maken en door de docent werden beoordeeld. Per taak is, nadat de leerlingen hun beoordelingen hadden verkregen, met een groepje leerlingen een interview gehouden over de taak, het maken daarvan, de verkregen beoordeling, en de eventueel verkregen feedback. Van de docenten zijn de aan deze leerlingen gegeven beoordelingen verkregen. Tot slot hebben de leerlingen een vragenlijst ingevuld over kenmerken van taken waarvan zij verwachten te kunnen leren.
De opvattingen en intenties van de docenten bij beoordelen bleken goeddeels overeen te komen met de aanwijzingen in de literatuur, maar bij de beoordeling betrekken van leerlingen en beoordelingen achteraf met de leerlingen bespreken, doen ze in relatief beperkte mate. De taken die de leerlingen moesten maken waren deels toetsen, met gesloten en open vragen, en deels opdrachten, variërend in omvang en complexiteit, deels individueel te maken en deels in tweetallen of groepjes leerlingen. Uit de mondelinge en schriftelijke bevraging achteraf bleek dat tweederde van de leerlingen vond dat de taak aansloot bij het voorafgaande onderwijs en niet bijzonder moeilijk was en ook tweederde de verkregen beoordeling wel had verwacht. De leerlingen waren gemiddeld behoorlijk tevreden over de beoordeling maar wilden wel vaak de volgende keer een beter resultaat. De leerlingen begrepen niet altijd hoe de beoordeling tot stand was gekomen, schreven het behaalde resultaat aan verschillende factoren toe, en namen zich voor het vervolg verschillende dingen voor. Uit de afsluitende vragenlijst bleek dat leerlingen vinden dat ze zich het beste kunnen voorbereiden op een taak als deze duidelijk is (bedoeling, aanpak, verwacht resultaat, beoordelingscriteria) en er dan ook het meeste van kunnen leren.
De beoordelingspraktijk varieerde tussen de docenten in relatie tot hun vak, het type taak en de (meer of minder vernieuwende) school waar ze werkten. Bij toetsen vonden de leerlingen de inhoud en de beoordeling duidelijker dan bij opdrachten, dachten ze

meer na over de verkregen beoordeling en konden ze ook beter aangeven wat ze ervan hadden geleerd.
Aanvullend op het vorige onderzoek werden ook in dit onderzoek, nu met padanalyses, relaties gevonden tussen de motivatie van de leerlingen en de waargenomen feedback, nu nader gespecificeerd naar de aard van de motivatie en de waardering van de feedback, en werden ook relaties zichtbaar met de verkregen beoordeling.

Een derde vervolgonderzoek op verzoek van een aantal schoolleiders was gericht op het verkrijgen van meer inzicht in welk pedagogisch-didactisch handelen van docenten op hun scholen gewenst is en waarop dat kan worden gebaseerd (hoofdstuk 7). Dit verzoek kwam voort uit de constatering in het eerste onderzoek (hoofdstuk 3) dat sprake is van verschillen tussen het handelen van de docenten, de onderwijsconcepten van de scholen, en aanwijzingen in de wetenschappelijke literatuur. Het onderzoek was gericht op het in kaart brengen van de overeenkomsten en verschillen tussen aanwijzingen voor gewenst docenthandelen vanuit wetenschappelijke literatuur, landelijke beleidsdocumenten, onderwijsconcepten van vier scholen, en opvattingen en prioriteiten van de betrokkenen op de scholen (schoolleiders, teamleiders, docenten, leerlingen, stakeholders).
In de voor dit onderzoek geselecteerde wetenschappelijke literatuur ontbreekt de persoonlijke vorming van de leerlingen, maar ook over pedagogische doelen is wetenschappelijke literatuur beschikbaar. De in dit onderzoek bestudeerde literatuur besteedt meer dan de andere bronnen aandacht aan op leren gerichte interacties met leerlingen. De reeks taken van docenten in het *cognitive apprenticeship* model is terug te vinden in de PDH.
In de beleidsgerichte bronnen zijn het vooral de SBL-competenties die een breed scala aan pedagogisch-didactisch handelen beschrijven (dat eveneens is terug te vinden in de PDH). De in de wetenschappelijke literatuur genoemde taken komen daarin min of meer allemaal terug. Aanvullend op de bestudeerde wetenschappelijke literatuur wordt in het recente beleid aandacht gevraagd voor het doen van onderzoek door docenten, en voor het signaleren en begeleiden van individuele leerlingen die met problemen kampen.
In de onderwijsconcepten van scholen wordt eveneens een breed palet van taken beschreven. Daarin lijkt een accent te liggen op pedagogische doelen en op het zoveel mogelijk aansluiten bij de leerlingen. In de wetenschappelijke literatuur (*cognitive apprenticeship*) en de SBL-competenties (organisatorisch, interpersoonlijk, pedagogisch, didactisch) is sprake van een organiserend kader. Dit komt in de onderwijsconcepten minder duidelijk naar voren.
De betrokkenen op de scholen kiezen, indien gevraagd om prioriteiten te stellen, vooral voor taken die tot het meer traditionele repertoire behoren, zoals zorgen voor een goed leerklimaat, lesgeven, leerlingen stimuleren en activeren, bijdragen aan persoonlijke ontwikkeling. De in wetenschappelijke literatuur, beleidsdocumenten en onderwijsconcepten genoemde meer vernieuwende taken, gericht op zelfstandig leren en

samenwerken van leerlingen, en het meer inspelen op verschillen tussen leerlingen, worden duidelijk minder vaak gekozen.
Multidimensionele schaalanalyses op basis van data verkregen door middel van paarsgewijze vergelijkingen tussen taken, brachten aanvullend aan het licht dat de verschillen in de mate waarin beide taken van elk aangeboden paar door de respondenten werden gekozen als belangrijkste taak, kunnen worden geordend op twee dimensies. Over alle scholen samen liggen daarbij de taken 'geven van les, uitleg en instructie' en 'leerlingen vaardigheden aanleren' in het midden. Deze zou men kunnen zien als traditionele basistaken van een docent. Op de ene dimensie die vervolgens kan worden onderscheiden, ligt de taak 'bijdragen aan de persoonlijke vorming van leerlingen' aan de ene kant en de taak 'laten en leren samenwerken' aan de andere. Mogelijk beschouwt men deze vormingsaspecten enigszins als tegenpolen (het individuele en het sociale). Op de andere dimensie staat aan de ene kant de taak 'ondersteunen van het leerproces' (taak 15) en aan de andere kant de taak 'inspelen op verschillen tussen de leerlingen' (taak 9). Mogelijk wordt het ondersteunen van het leerproces primair gerelateerd aan de leerinhouden en niet direct in verband gebracht met inspelen op verschillen tussen leerlingen. Kennelijk kunnen taken die vanuit onderwijskundig oogpunt duidelijk met elkaar samenhangen op de onderscheiden dimensies op grote afstand van elkaar worden geplaatst en worden deze mogelijk niet direct op elkaar betrokken. Een aanvullende aanwijzing voor dit laatste is te vinden in de hoofdstukken 5 en 6. Daarin valt op dat waar in de literatuur formatieve en summatieve beoordeling vaak als één domein worden gezien, in het handelen van docenten bij het geven van feedback het meest beoordelende aspect, namelijk vergelijken met de standaard, minder aandacht krijgt, en in het handelen bij beoordelen een sterk formatief aspect, namelijk leerlingen betrekken bij de beoordeling van hun eigen werk, minder uit de verf komt. In beide gevallen wordt dan een aspect onderbelicht dat voor leerlingen juist heel leerzaam kan zijn. Dat bij feedback een duidelijk beoordelend aspect en bij beoordelen een potentieel formatief aspect minder aandacht krijgt, lijkt erop te wijzen dat geven van feedback en beoordelen door docenten, anders dan in recente literatuur wordt aanbevolen, als twee duidelijk verschillende taken worden gezien.
Al met al worden de mogelijkheden die er zijn om de onderwijsconcepten te verbinden met wetenschappelijke inzichten en met landelijk beleid op dit moment niet volledig benut. De onderzochte onderwijsconcepten vertegenwoordigen meer een verzameling van verschillende ambities dan een samenhangende en richtinggevende visie, de prioriteiten die de betrokkenen stellen hangen niet duidelijk samen met de meer of minder vernieuwende onderwijsconcepten van de scholen, en de gronden die zij daarvoor aanvoeren hebben weinig te maken met beschikbare (gepubliceerde) kennis over onderwijs en leren.

Op basis van de ervaringen, opgedaan bij het terugkoppelen en bespreken op de scholen van de resultaten van de voorgaande vijf onderzoeken, literatuur over

processen van verandering en vernieuwing op scholen, en een zesde onderzoek naar een veranderingsproces op één school, zijn drie belangrijke condities voor verandering en vernieuwing naar voren gekomen en nader uitgewerkt (hoofdstuk 8). Deze betreffen het belang van het hebben van een visie op de gewenste ontwikkeling, het leiding geven aan het veranderingsproces, en het kunnen omgaan met datafeedback.

In de meeste van de in dit proefschrift beschreven onderzoeken bleek sprake van aanzienlijke verschillen tussen docenten, ook binnen dezelfde school, in hun opvattingen en/of hun huidige handelen. Ook zijn er verschillen tussen handelen dat gewenst kan worden geacht op basis van wetenschappelijke inzichten, de onderwijs-concepten van de scholen, en de keuzen die betrokkenen op de scholen maken. In reactie op de terugkoppeling van de onderzoeken bleken de schoolleiders de aangetroffen verschillen opmerkelijk te vinden. In de onderwijsconcepten van scholen worden vaak uiteenlopende inzichten en doelen gecombineerd en geconcludeerd is dat aan het verminderen van de genoemde verschillen kan worden bijgedragen door het gezamenlijk formuleren van een duidelijke en (meer) coherente visie.

Het belang van een duidelijke visie wordt ondersteund door de bestudeerde literatuur, die in totaal vier groepen voorwaarden bestrijkt voor het succesvol doorvoeren van veranderingen. Op basis daarvan kan een optimale context als volgt worden beschreven. *De vernieuwing* biedt een duidelijke visie op en beeld van het gewenste onderwijs en de weg daarnaar toe. De visie is voldoende concreet maar tegelijk voldoende inspirerend en open voor nadere invulling. De visie is voldoende nieuw om uit te dagen en voldoende herkenbaar om haalbaar te achten. De vernieuwing is afgebakend maar wel verbonden met wat verder in de school speelt. De vernieuwing wekt de verwachting van positieve effecten bij leerlingen en kan in stappen worden uitgeprobeerd. *Het veranderingstraject* wordt duidelijk georganiseerd in termen van taken, prioriteiten en fasering en met behulp van faciliteiten en materialen. Er wordt gezorgd voor goede communicatie, steun en coördinatie. Het proces en de effecten worden tussentijds gevolgd, zichtbaar gemaakt, gedeeld en geëvalueerd. Waar nodig vinden bijstellingen plaats. *De schoolleiding en de docenten* trekken samen op. De kaders zijn duidelijk maar vragen wel om nadere ontwikkeling en maken deze ook mogelijk. De docenten worden vanaf het begin bij het proces betrokken en zijn daarvan mede-eigenaar. De bij docenten aanwezige expertise wordt erkend en benut en verder ontwikkeld, in relatie tot de ontwikkeling van de school. Er is ruimte voor verschillende uitwerkingen, mits deze bijdragen aan het gestelde doel. *De dialoog tussen alle betrokkenen* staat centraal en wordt actief opgezocht en gevoerd. Opvattingen, kennis, ervaringen en normen worden besproken en gedeeld. Op basis van open interpersoonlijke relaties wordt gewerkt aan reflectie, gedeelde betekenis-geving, samenwerking en teamvorming.

Uit de in dit proefschrift gerapporteerde ervaringen met het terugkoppelen van verzamelde en verwerkte gegevens naar schoolleiders, teamleiders en docenten, bleek dat een aantal schoolleiders en teamleiders aangaf dat zij bij het interpreteren van de gegevens ondersteuning nodig hebben omdat zij niet gewend zijn zulke gegevens te

lezen en te duiden. Dit blijkt ook uit diverse onderzoeken naar "datageletterdheid" bij schoolleiders. Dit pleit voor een aanbod waarbij niet wordt volstaan met een rapportage, maar door de onderzoeker tevens actief de dialoog wordt aangegaan met de betrokkenen op de scholen.

Voor het proefschrift als geheel zijn enkele overkoepelende onderzoeksvragen geformuleerd en beantwoord (hoofdstuk 1 respectievelijk hoofdstuk 9). De eerste vraag betrof de aard en inhoud van het pedagogisch-didactisch handelen van docenten in het voortgezet onderwijs. Het betreffende handelen kan worden beschreven aan de hand van een aantal taakgebieden of taken die volgens de betrokkenen op de scholen goed dekkend zijn en waarin ook de beleidsmatig vastgestelde docentcompetenties (SBL) en de verkregen aanwijzingen uit de bestudeerde wetenschappelijke literatuur terug-komen. Deze brede dekking van de onderscheiden taken neemt niet weg dat als wordt ingezoomd op afzonderlijke taakgebieden, zoals geven van feedback of beoordelen, het handelen van docenten op basis van de daarop betrekking hebbende wetenschappelijke literatuur nog (veel) verder kan worden gespecificeerd en dergelijk meer gespecificeerd handelen ook is terug te vinden in de opvattingen, intenties en waarneembare gedragingen van docenten.
De tweede vraag betrof de mate waarin docenten dit handelen vertonen. Het antwoord luidt dat zij dit blijkens de uitgevoerde onderzoeken in gemiddeld behoorlijk hoge mate doen, zowel volgens de docenten zelf als volgens hun leerlingen. Mogelijke verklaringen voor deze uitkomst zijn: dat de meeste scholen die aan de onderzoeken deelnamen vernieuwende scholen zijn die van hun docenten een breed handelings-repertoire verwachten; dat veel onderscheiden taken een leerlinggericht karakter hebben; dat docenten gemiddeld geneigd zijn zichzelf relatief hoog in te schatten; en dat leerlingen gemiddeld geneigd zijn hun docenten positief te beoordelen. Er is echter sprake van vrij grote verschillen tussen de taken, waarbij blijkt dat taken die niet direct behoren tot of passen bij het traditionele lesgeven relatief minder aandacht krijgen (ook op scholen met een vernieuwend onderwijsconcept).
De derde vraag betrof de variatie in het handelen en waarmee deze variatie samenhangt. Dominerend in de uitkomsten van de uitgevoerde onderzoeken is dat sprake is van grote individuele verschillen tussen de docenten. Deze verschillen hangen soms niet en soms wel samen met de scholen waar de docenten werken. De overallscore (voor het repertoire als geheel) verschilt significant tussen docenten maar niet tussen de scholen. Op het taakgebied geven van feedback is geen duidelijk verschil geconstateerd tussen docenten die werken op een meer traditionele school en docenten die werken op scholen met een sterk vernieuwend onderwijsconcept, maar op het taakgebied beoordelen was dit verschil er wel.
De laatste twee vragen waren die naar welk handelen gewenst wordt geacht en door wie en op welke gronden, en wat de verschillen zijn tussen huidig en gewenst handelen en hoe die kunnen worden overbrugd. De antwoorden op deze vragen zijn voor een groot deel te vinden in de hoofdstukken 7 en 8 (zie hierboven). Over de mate waarin

docenten en leerlingen zelf bepaald docenthandelen gewenst vinden en over het verschil tussen het door hen aangegeven huidige en gewenste handelen bieden de hoofdstukken 3 t/m 6 informatie. Daaruit blijkt dat docenten en leerlingen de meeste nadruk leggen op het belang van het leerklimaat, het lesgeven, en de samenwerking tussen docenten, terwijl beide groepen aan het aandacht geven aan verbanden tussen vakken de minste waarde hechten. Het belang dat leerlingen hechten aan het leerklimaat, het lesgeven, en de samenwerking tussen docenten, wordt bevestigd in de groepsinterviews met leerlingen die zijn gehouden op de scholen in het onderzoek in hoofdstuk 7, waarin de leerlingen precies dezelfde drie taken naar voren brachten.
Kijkend naar de grootte van het verschil tussen huidig en gewenst handelen blijkt dat docenten gemiddeld het duidelijkste meer aandacht zouden willen geven aan het bijdragen aan de persoonlijke vorming van de leerlingen, het wijzen op verbanden tussen vakken, en de beoordeling, terwijl de verschillen bij de leerlingen het grootst zijn voor het inspelen op verschillen tussen leerlingen en het leerlingen leren zelf opdrachten aan te pakken. Op deze manier bekeken, komen dus ook minder traditionele taken naar voren.

Al met al nemen in het handelen van docenten en in hun keuzen en prioriteiten meerdere vernieuwende taken, zoals die in wetenschappelijke publicaties, beleidsdocumenten en onderwijsconcepten van scholen aandacht krijgen, een minder grote plaats in dan een aantal meer traditionele taken, en sluiten de leerlingen zich daarbij aan. Op scholen met een sterk vernieuwend onderwijsconcept is dit concept wel duidelijk zichtbaar in de organisatie van het onderwijs qua ruimte en rooster en ook wel in opvattingen en intenties van docenten, maar (vooralsnog) minder in het feitelijk pedagogisch-didactisch handelen van de docenten.
Doordat de onderzoeken zijn uitgevoerd op verzoek van scholen kon geen sprake zijn van invloed op de samenstelling van de onderzoeksgroep (bijvoorbeeld door het trekken van een aselecte steekproef). Dit beperkt de generaliseerbaarheid van de uitkomsten. Aan de onderzoeken die zijn beschreven in de hoofdstukken 3, 5, 6 en 7 namen echter naast sterk vernieuwende scholen ook een of twee meer traditionele scholen deel en bleken de verschillen tussen beide typen scholen over het algemeen niet groot te zijn.
De opvattingen en intenties van de docenten op alle in de onderzoeken betrokken scholen waren vaak zeer gedifferentieerd en passen bij een aantal van de vernieuwende ambities zoals beschreven in wetenschappelijke publicaties, beleidsdocumenten en onderwijsconcepten. Hun feitelijk pedagogisch-didactisch handelen varieert echter sterk.

Summary

School leaders in secondary education feel themselves increasingly forced by social developments, national policy, new scientific insights, and changes within their student population to initiate innovations in their schools. The changes that seem desirable often concern the pedagogical behaviour of the teachers.[3] During a period of four years the school leaders of eight schools, in variable combinations, have submitted five applications for short-term educational research projects (with a maximum term of one year).[4] In addition to the first project, in which an instrument is developed for measuring the pedagogical behaviour of teachers, in collaboration with four of the schools the author has made a shorted version of this instrument and has validated this in the educational practice of these schools. This dissertation describes (in chapters 3-8) the design and the main results of these six studies.

The development of an instrument for measuring the pedagogical behaviour of teachers (chapter 3) started with the analysis and delineation of the concepts "behaviour" (Dutch: handelen) and "pedagogical" and the specification of their internal conceptual structure. Based on scientific literature on education and learning, national educational policy documents, the educational visions of six schools and input from a number of teachers, teachers' pedagogical behaviour could be classified into 15 task domains, at the request of the school leaders supplemented by an extra task domain concerning teachers' professional development activities.
In view of the intended use of the instrument in the schools (which offer courses from 'vmbo' to 'gymnasium')[5] and the quality desired from the instrument and the results it can produce, we decided to give it the form of a written questionnaire for teachers and students in which each task domain is represented by a number of statements about the possible behaviour of a teacher on that task domain. The teachers and students are asked to indicate the teacher's present behaviour as well as the behaviour the respondent would prefer. After having been piloted the adjusted version was filled out by 141 teachers and 900 students from the six schools. For each task domain the specified statements formed a reliable scale (in terms of item-rest correlations and Cronbach's alpha). Together these scales combine into a very reliable (second-order) scale which can be seen as a measure for a teacher's total repertoire of pedagogical behaviour. Based on the data from the students, it appeared that by using the scales per task domain as well as the overall scale it is possible to discriminate reliably between

[3] In Dutch we use the word "handelen", including both observable behaviour and underlying intentions.
[4] We refer here to the Dutch national subsidy arrangement known as "Kortlopend Onderwijsonderzoek".
[5] The schools cover the complete range of courses in the Dutch selective secondary education system.

teachers. Factor analyses per task domain for six of the domains resulted in the discovery of two subtask domains, each of them forming a reliable scale as well, which can been interpreted as a respectively less or more innovative version of the task domain concerned.
At the request of the school leaders a shorted version of the instrument was made, In this version the discovered subtask domains were represented as separate tasks and the number of statements per task was reduced, resulting in 22 tasks with an average of five statements ('items') per task. This short version (which was given the name PDH) has been made available as a Web-based questionnaire. Using this shorted instrument data from 49 teachers and 502 students have been gathered and analysed for validation purposes. The reliabilities of the scales per task and the overall scale were acceptable to good and their discriminative power was substantial, but the Cronbach's alphas and intraclasscorrelations were a little lower than with the more elaborated version. Apparently, for the reduction of the scales (from 10 to 5 items on average) and the accompanying less extensive operationalization we paid a price.

In another research study in which the author participated, 136 teachers have been asked to judge the importance of the 22 tasks and factor analysis of this data showed that four groups of tasks could be distinguished. In the present PDH data each of these groups did form a reliable scale, concerning respectively students' development, the (cognitive) learning process, feedback and assessment, and collaboration (with colleagues) and professional development.
In addition to the PDH, 17 teachers and 188 students also filled out the QTI (Questionnaire on Teacher Interaction; in Dutch: Vragenlijst Interpersoonlijk Leraarsgedrag, VIL). The scales of both instruments (tasks, groups of tasks and the overall scale of the PDH and sectors and dimensions of the QTI) correlate moderately positive according to a certain pattern, so these two instruments each represent a different dimension of the behaviour of teachers (pedagogical versus interpersonal). One plausible interpretation of the moderate relationships found is that (the quality of) the interpersonal behaviour is a condition for the pedagogical behaviour to show its potential.
Analyses of the differences between schools and the relationships between the PDH scores and characteristics of teachers and students showed that the pedagogical behaviour of teachers to a high extent varies individually, also within the participating innovative schools. Multilevel analysis resulted in the finding that in the data of the students a quarter of the variance (25%) in the PDH total score lies on the teacher level and that the variance on the school level (about 10%) is not significant. The scores do not substantially relate with teacher characteristics such as gender, age and experience. That no clear relations are found between the PDH-scores and the teachers' age and experience could be the result of the fact that in this study relatively few older and more experienced teachers participated. It could also be a result of the fact the teachers in this study are working in a context of change and renewal. In such a situation,

teachers who already have quite a lot of experience can once more get involved in a learning process, with consequences for their pedagogical behaviour. Because some of the tasks of the PDH are concerned with relatively innovative behaviour that is experienced by quite a lot of teachers as not easy to perform, this possibility is not imaginary.

At the request of a number of school leaders a follow-up study was conducted into one of the task domains, namely giving feedback to students working independently on certain tasks (chapter 5). From the literature a good 20 potential relevant aspects, content areas and forms of process-oriented and result-oriented feedback were obtained. On four schools the school leaders and 12 teachers (three per school, from different school subjects) were interviewed and filled out written questionnaires about their beliefs, opinions, intentions and practices on this task domain. For each teacher a group of students from grade 9 or 10 filled out a questionnaire about their motivation for making tasks concerning that specific subject of that specific teacher and about the way in which this teacher normally gives feedback. With each teacher a lesson for the group students concerned was recorded on video and observed, in which the students were working independently on tasks. At the end of the lesson the students filled out a short questionnaire about the feedback the teacher had given them. With each teacher the video was observed and discussed in the form of a stimulated recall interview. Finally each teacher got a personal report with a survey of all results and a in a final conversation this report was systematically discussed.
The beliefs, opinions and intentions of the teachers pertaining to giving students feedback appeared largely to correspond with the directions in the literature and their behaviour was consistent with their intentions, according to both the questionnaire data and the observations. They disposed of a broad and differentiated repertoire which they also applied in practice and their behaviour was appreciated by their students to a fairly high extent. Contrary to the expectations of the school leaders and a number of teachers as well, their feedback behaviour appeared not to be related to their age or their experience as a teacher.
The teachers seemed to be largely focussed on the students. Giving feedback for them was a matter of personal contact, confidence and attention. Before everything else feedback for them must be positive and constructive, Indicating to the student that something was not (yet) good and assessing a performance on a clear standard seemed to occur relatively less often. Feedback appeared to be more a reaction to a problem of the student who does not comprehend something and therefore can not proceed with the task he is working on. The feedback was aimed so that the student does comprehend and can proceed. Seen from the literature this teacher practice supports the acceptance of the feedback but less so the student's self-evaluation competence.
According to the analysis of a structural equation model specified for this, the scope of the teachers' normal feedback behaviour as indicated by their students in their questionnaires is significantly, positively and substantially related to their motivation

for working on tasks for the subjects the teachers were teaching. This result can be taken as a confirmation of the impact of feedback on students' motivation as described in the literature, but it could also be that motivated students better observe and remember the pedagogical behaviour of their teachers. Still another interpretation is that teachers give motivated students more or more diverse feedback, possibly because the performance of such students gives rise to this and/or such students stimulate them to do this.

A second follow-up study, requested by a number of school leaders, was focussed on the teacher task domain of assessing and marking students' work and products; the question was in which way such assessments could be instructive for the students (chapter 6). From the literature a number of characteristics of assessment were derived that are expected to contribute to the motivation and learning of students. In three schools nine teachers (three teachers per school, teaching English, geography, and mathematics) were interviewed and filled out written questionnaires about their assessment practice and also students (per teacher a group of students in grade 9 havo/vwo[6]) filled out a written questionnaire about their motivation for making tasks for the teacher's subject and this teacher's assessment practice. Two or three tasks per teacher were chosen for the study, which the students had to make in a period of several months and which would be assessed and marked by the teacher. Per teacher per task, after the students had received their teacher's judgments, a group of four students were interviewed about the task, their work on that task, the judgments they received and the accompanying feedback, if any. The judgments were received from the teachers. Finally, all students who had participated in the group interviews filled out a short questionnaire about the characteristics of tasks from which they expected to learn (that is: which would make the task instructive).
The beliefs, opinions and intentions of the teachers appeared to largely correspond with the directions in the literature, but involving the students in the assessment and discussing the assessments given by the teachers with the students they do relatively less frequent. The tasks the students had to make partly were (knowledge) tests, including pre-structured as well as open questions, and partly were individual or group assignments, varying in magnitude and complexity. From the interview and questionnaire data gathered after each task it showed that on average two-thirds of the students thought that the task aligned up with the lessons given by the teacher and had not been too difficult and that also two-thirds had expected the judgments they had received. On average, the students were rather satisfied with the judgments but often they wanted to get a better result next time. The students did not always comprehend the basis for the judgement; they attributed the result to different factors and quite different resolutions. The results of the final questionnaire showed that the students think they can best prepare themselves for a task which is clear (concerning its goal,

[6] The two highest streams of the four streams in the selective Dutch secondary education system.

approach, expected result, assessment criteria) and also that they can learn most from such a task.
The assessment practice differed between teachers in relation to their subject, the type of task and the (more or less innovative) school at which they worked. The students on average thought tests to be clearer than assignments, they reflected more upon them, and also could better indicate what they had learned from them.
In addition to the foregoing study, again relationships were found, this time by using path analyses, between the students' motivation for the task and the feedback they reported their teacher had been giving, now specified into several kinds of motivation and several types of appreciation of the feedback, and also certain relations were found with the mark they had received.

A third follow-up study, again at the request of a number of school leaders, was aimed at getting more insight into the pedagogical behaviour which is desirable for the teachers on their schools and on what grounds this can be based (chapter 7). This request followed from the finding in the first study (chapter 3) that there are differences between the behaviour of the teachers, the educational philosophies of the schools, and the directions from scientific literature on education and learning. The goal of the study was to map the similarities and differences between directions for desirable teacher behaviour from four different sources: scientific literature, national policy documents, educational philosophies of four schools, and the beliefs and priorities of the people concerned at these schools (school leaders, team leaders, teachers, students, and external stakeholders such as parents).
In the scientific literature which was selected for this study, the teacher task of contributing to the students' personal development was missing, but about pedagogical goals such literature is available as well. The literature studied, more than the other sources, pays attention to learning-focussed interactions with students. The series of tasks for teachers in the 'cognitive apprenticeship' model is well represented in the PDH.
In policy documents especially the so-called SBL teacher competences (laid down by Dutch national law) include a broad range of pedagogical behaviour (which also is well-represented in the PDH). More or less all of the teacher tasks mentioned in the scientific literature can be found in these competence descriptions. In addition to the scientific literature in recent national educational policy documents attention is asked for conducting research by teachers and for observing and coaching individual students having problems.
In the educational philosophies of the schools a broad palette of tasks is being described as well. In these documents the emphasis seems to lie on pedagogic goals and on linking up as much as possible with the students. The scientific literature (cognitive apprenticeship model) and the SBL competences (distinguishing between organizational, interpersonal, pedagogical, and didactical competences) are based on an organizing framework. In the educational philosophies this is less the case.

The people concerned at the schools, when asked to set priorities, choose particularly tasks belonging to the more traditional repertoire, such as caring for a good learning climate, giving lessons, stimulating students, and contributing to their personal development. The more innovative tasks mentioned in the scientific literature, national policy documents, and educational philosophies of the schools, aiming at students' independent learning and collaboration and at adapting to individual differences between students, are chosen clearly less often.

Multidimensional scale analyses based on data gathered from the school leaders, team leaders, teachers and external stakeholders using pair-wise comparisons between tasks showed that the differences in the extent to which both tasks in each pair are chosen by the respondents as most important task can be represented on two dimensions. Across the schools the tasks 'giving lessons, explanation and instruction' and 'teaching students skills' lie in the middle. These tasks might be seen as traditional basic tasks of a teacher. On the vertical dimension the tasks 'contributing to students' personal development' lies on one end and the task 'letting and teaching students work together' on the other. Possibly these tasks are seen somewhat as opposites (the individual and the social aspect). On the horizontal dimension at one end lies the task 'supporting the learning process' and on the other end 'adapting to individual differences between students'. Possibly for the respondents supporting the learning process is related to the learning content and is not directly connected to adapting to differences between students. Obviously tasks which from an educational point of view are clearly connected can be positioned on the underlying dimensions on a considerable distance from each other and possibly in the minds of the respondents these tasks are not directly related.

An additional indication for the just-mentioned assumption can be found in chapters 5 and 6. In these chapters it strikes home that while in the literature formative and summative assessment often are seen as one domain, in the teachers' practice of giving feedback the most summative aspect, namely comparing the student's performance with a standard, gets less attention and in the practice of their summative assessment (including marking) the most formative aspect, namely involving the students in the assessment, does not appear clearly. Therefore, in both cases an aspect is underexposed which for the students can be very instructive. The fact that when giving feedback a clearly summative aspect and when judging a potentially formative aspect is getting less attention seems to indicate that, contrary to the recommendations in the literature, teachers see these two tasks as rather different.

In sum, the possibilities to connect the educational philosophies of the schools with scientific insights and with national policy at this moment are not completely used. The educational philosophies studied represent to a higher extent a collection of different ambitions than a coherent and guiding vision, the priorities of the people concerned do not correspond clearly to the more or less innovative educational mission statements of their schools, and the grounds they mention for their choices have little to do with the available (published) knowledge about education, teaching and learning.

From the experiences gained while reporting and discussing on the schools the results of the foregoing five studies, scientific literature about processes of change and innovation in schools, and a sixth study into a change process within one school, three important conditions for change and innovation came to the forefront and have been explicitly elaborated (chapter 8). These conditions are the importance of: having a clear vision on the desired development, leading the change process, and being able to handle (interpret and use) data feedback.

In most studies in this thesis teachers differed considerably, also within the same school, in their views and behavior. Also there are differences between teacher behavior which can be seen as desirable from the literature, the educational philosophies of the schools and the choices people in the schools make. In reaction to the datafeedback given, the schoolleaders said they found the differences remarkable. In the educational philosophies of the schools often rather diverse insights and goals are combined and it was concluded that reducing the differences can be furthered by collaboratively formulating a clear and (more) coherent vision.

The importance of a clear vision is supported by the literature, which covers four groups in total of conditions for succesfully carrying through desired changes. Based on this an optimal context can be described in the following way. The *innovation* offers a clear vision and picture of the desired educational approach and how to get there. The vision is sufficiently concrete but also inspiring and open for personal interpretation. It is sufficiently new to challenge and recognizable enough to judge it to be feasible. It is marked out but also connected with ongoing developments in the school. It arouses expectations of positive effects for students and can be tried stepwise. The *route of change* is being transparently organized in terms of tasks. priorities and steps and the facilities and materials needed. Good communication. support and coordination are being cared for. as well as the monitoring. making visible. sharing and evaluating of the process and the effects. Adjustments needed are being carried through. The *schoolleader and teachers* are advancing together. The frameworks are clear but demand and make possible further development. The teachers are involved from the beginning and are also owners of the process. Their expertise is recognized and used and further developed in connection to the school development. There is space for different elaborations provided that they contribute to the goals set. The *dialogue between all involved* is central and is being actively looked out for and carried on. Beliefs. knowledge. experiences and norms are being discussed and shared. Based upon their interpersonal relationships the participants reflect. share meanings. collaborate and work on teambuilding. it is being worked on.

The experiences described in this thesis with feeding back data to schoolleaders. teamleaders and teachers showed that a number of them are not used to examining and interpreting such data. The same has been found in studies on “data literacy” of schoolleaders. This speaks for researchers not only supplying a report but actively involving themselves in the dialogue with the people working in the schools.

For this dissertation as a whole several overarching research questions have been formulated and answered (respectively chapters 1 and 9). The first question concerned the nature and content of the pedagogical behaviour of teachers in secondary education. This behaviour can be described in terms of a number of task domains or tasks which, according to the people concerned at the schools, are representing the teachers' practice to a high extent and in which the teacher competences laid down by Dutch national law as well as the recommendations from the scientific literature are well represented. This broad covering quality of the tasks does not alter the fact that when focussing on separate task domains, such as feedback or assessment, using the relevant more specific scientific literature the teacher's professional behaviour can be specified still to a fairly higher extent and that such more specified professional behaviour can also be found in the beliefs, opinions, intentions and observable behaviour of teachers.

The second question was to what extent teachers show this behaviour. The answer is that, according to the studies conducted, they do this on average to a rather high extent, according to the teachers themselves as well as according to their students. Possible explanations for this result are that most of the schools which participated in the studies are innovative schools which expect from their teachers a broad repertoire of teaching behaviour; that most of the tasks we distinguished have a student-oriented nature; that teachers on average are inclined to judge themselves relatively positively; and that students on average are inclined to judge their teachers positively. However, the tasks differ quite considerably and tasks which do not belong to or fit with traditional teaching get relatively less attention (also in schools with an innovative educational philosophy).

The third question pertained to the variation in the teachers' behaviour and to what this variation is connected. Dominating in the results of our studies are the big individual differences between teachers. These differences are sometimes related to the schools the teachers are working in and sometimes not. The overall score (on the repertoire of pedagogical behaviour as a whole) is significantly different between teachers but not between schools. On the task domain feedback no clear difference was found between teachers working in a more traditional school and teachers working in schools with a more innovative educational philosophy but on the task domain assessment a difference between these two categories of school clearly existed.

The last two questions were which teacher behaviour is judged as desirable and by whom and on which grounds, and how the current behaviour and the desired behaviour differ and how this difference can be bridged. The answers to these question can largely be found in chapters 7 and 8 (see above). On the extent to which teachers and students themselves see certain teacher behaviour as desirable and on the difference between their answers about the current and the desirable behaviour chapters 3-6 give information. From these results it appears that both teachers and students give the most weight to the importance of a good learning climate, the quality of the lessons, and the collaboration among teachers, while both groups give the least weight to giving attention to connections between school subjects. The importance the students attach to

the learning climate, the lessons, and the collaboration among teachers is independently confirmed in the group interviews held with students on the schools in the study in chapter 7, in which the students put forward exactly the same three tasks.
Looking at the size of the differences between current and desired behaviour in the data of the teachers and the students, it appears that teachers on average most clearly want to give more attention (than they do now) to contributions to students' personal development, attention to connections between subjects, and their assessment practice, while in the data of the students the biggest differences are concerned with adapting to the differences among students and teaching students how to approach and make assignments. In this way, therefore, less traditional tasks also emerge.

All in all, in the behaviour of teachers and in their choices and priorities several innovative tasks, as described in scientific publications, national policy documents and educational philosophies of schools have a less prominent place than a number of more traditional tasks, and the students link up with this. In schools with a strong innovative educational philosophy this philosophy is clearly visible in the organization of the education as for rooms and timetable and in the beliefs, opinions and intentions of teachers but as yet less in the actual pedagogical behaviour of the teachers.
Because the studies have been conducted at the request of schools it was not possible to have influence on the composition of the research group, for instance by drawing a random sample. This restricts the generalizability of the results. However, in the studies described in chapters 3, 5, 6 and 7, as well as strongly innovative schools one or two traditional schools also participated and the differences between both types of schools generally appeared not to be great.
The beliefs, opinions, and intentions of the teachers in all participating schools often were very differentiated and were matching a number of innovative ambitions as described in the scientific literature, national policy documents, and educational philosophies. Their current pedagogical behaviour, however, does strongly vary.

Curriculum Vitae

Yvette Sol werd geboren in Amersfoort op 30 november 1961. Zij voltooide haar middelbare schoolopleiding op het Stedelijk Gymnasium Johan van Oldenbarnevelt te Amersfoort waar zij in 1980 haar diploma gymnasium-B behaalde.
Na een studie fysiotherapie, die zij in 1984 afrondde, was zij een aantal jaren werkzaam in de eerstelijns gezondheidszorg in verschillende gezondheidscentra en in haar eigen praktijk. Vanuit haar specialisatie als kinderfysiotherapeute is zij daarna een aantal jaren werkzaam geweest binnen diverse scholen voor speciaal basisonderwijs en een kinderdagverblijf voor verstandelijke gehandicapte kinderen. Naast het geven van individuele begeleiding aan patiënten vervulde zij vanuit haar inhoudelijke expertise als professional tevens een rol bij het bevorderen van de deskundigheid van artsen, pedagogen, logopedisten, leerkrachten en groepsleidsters.
Vanaf 2001 heeft zij haar werk en de zorg voor een gezin gecombineerd met de deeltijdstudie Onderwijskunde aan de Universiteit Utrecht. Deze studie heeft zij in 2005 afgerond met het predicaat ‘cum laude”

Yvette heeft sinds 2005 als onderwijskundige eerst gewerkt op de Politie Academie te Apeldoorn. Zij hield zich daar als toetsdeskundige vooral bezig met het verbeteren van competentiegerichte toetsing.
In 2006 maakte zij de overstap naar de Universiteit Utrecht en heeft zij, aanvankelijk bij de vakgroep Onderwijskunde en later bij het voormalige IVLOS, als onderzoeker en projectleider een aantal extern gefinancierde onderzoeken uitgevoerd. In deze periode heeft zij zowel haar Basiskwalificatie onderwijskundig advies en ontwikkelwerk als haar Basiskwalificatie Onderwijs behaald.
Inmiddels werkt Yvette als adviseur/onderzoeker bij de afdeling Onderwijsadvies en Training van het Centrum voor Onderwijs en Leren van de Universiteit Utrecht. Hierbij biedt zij ondersteuning aan scholen (schoolleiders en docenten) bij het werken aan verbeteringen op het vlak van het pedagogisch-didactisch handelen van docenten, in het bijzonder het geven van feedback, toetsen en beoordelen, en op het terrein van kwaliteitszorg, benutten van datafeedback en vormgeving van vernieuwingsprocessen.
Daarnaast houdt zij zich bezig met onderzoek naar Wetenschap en Techniek in het basisonderwijs.
Als docent en cursuscoördinator verzorgt Yvette onderwijs in de Academische Lerarenopleiding Primair Onderwijs (ALPO) en geeft zij cursussen over toetsen en beoordelen, onderwijsverbetering en evaluatie, en academisch schrijven. Tevens begeleidt zij bachelorstudenten van de ALPO en masterstudenten van de opleiding Onderwijskunde bij hun afstudeeronderzoek. Tot slot participeert zij als docent in een professionaliseringstraject voor leerkrachten in het basisonderwijs, waarin de nadruk ligt op het leren begeleiden van studenten bij het uitvoeren van onderzoek.