

B.A.M. van den Berg
W.F. Admiraal
A. Pilot

B.A.M. van den Berg (I.vanden-Berg@ivlos.uu.nl), W.F. Admiraal (W.Admiraal@ivlos.uu.nl) en A. Pilot (A.Pilot@ivlos.uu.nl) zijn werkzaam bij het IVLOS (Interfacultair Instituut voor Lerarenopleiding, Onderwijsontwikkeling en Studievoordigheden) van de Universiteit Utrecht.

Intercollegiale beoordeling van schrijfproducten

Opzet en resultaten van een ontwerponderzoek

In dit artikel wordt verslag gedaan van een praktijkstudie, waarin een ontwerp voor intercollegiale beoordeling van schrijfproducten is ontwikkeld en geïmplementeerd. De evaluatie richt zich op de uitvoerbaarheid van het ontwerp, de aard van de intercollegiale feedback en de effecten ervan op de kwaliteit van de schrijfproducten. Plaats van uitvoering is de opleiding Geschiedenis van de Utrechtse Letterenfaculteit, het ontwerp is beproefd in één cursus.

De resultaten zijn positief: de uitvoering verliep soepel, de meeste studenten zouden vaker met intercollegiale beoordeling willen werken en vrijwel alle werkstukken zijn verbeterd als gevolg van het intercollegiale commentaar. Dit neemt niet weg dat er redenen zijn om het onderwijsontwerp te optimaliseren.

Inleiding

Academische vorming houdt behalve het aanleren van zgn. wetenschapsgebonden bekwaamheden (kennis van het vakgebied, onderzoekservaring en dergelijke) ook het verwerven van algemene beroepsbekwaamheden in (Holleman e.a., 1999). Tot die laatste categorie behoren de communicatieve vaardigheden, zoals schriftelijk rapporteren en mondeling presenteren. Veel universitaire opleidingen zijn de laatste jaren bezig met een inhaalmanoeuvre om leerlijnen op curriculumniveau te ontwikkelen voor dit type vaardigheden, vanuit het inzicht dat het aanleren ervan niet aan het toeval overgelaten mag worden. Daarbij gaat het vooral om het creëren van een doordachte opbouw van de studietaken, waarbij de docentbegeleiding geleidelijk aan van karakter moet veranderen om de zelfsturing van de student te bevorderen.

In deze bijdrage concentreren we ons op het aspect van de begeleiding van schrijftaken. In hoeverre kunnen studenten bepaalde begeleidingstaken van de docent overnemen?

Zijn studenten in staat om adequate feedback te geven op het werk van medestudenten? Wat doen ze met het commentaar dat ze van elkaar krijgen? Wat is een goede manier om een en ander te organiseren?

Om hier meer zicht op te krijgen is een praktijkstudie ondernomen, waarin een ontwerp voor intercollegiale beoordeling van schrijfproducten is ontwikkeld en geïmplementeerd. De evaluatie richt zich op de uitvoerbaarheid van het ontwerp, de aard van de intercollegiale feedback en de effecten ervan op de kwaliteit van de schrijfproducten. De uitvoering vond plaats binnen de opleiding Geschiedenis van de Utrechtse Letterenfaculteit in het kader van een schrijfpracticum voor studenten dat geprogrammeerd is in het derde opleidingsjaar.

Schrijven is voor historici een essentiële beroepsvaardigheid. De opleiding besteedt dan ook veel aandacht aan de ontwikkeling van de schrijfvaardigheid. Studenten Geschiedenis zijn een aanzienlijk deel van hun tijd bezig met het schrijven van werkstukken en de docenten met het corrigeren, becommentariëren en beoordelen ervan. Desondanks heerst bij veel docenten het idee dat het droevig gesteld blijft met de schrijfvaardigheid van hun studenten.

In deze context kwam het concept van intercollegiale beoordeling¹ naar voren als een vorm van ondersteuning die studenten elkaar kunnen bieden bij het verbeteren van hun schrijfvaardigheid. In de beroepswereld is het doorgaans zo dat collega's de fijne kneepjes vooral van elkaar leren, met name waar het gaat om complexe vaardigheden. Soms worden daarvoor speciale coachingstrajecten ingericht, waarin de relatief onervarene begeleid wordt door een collega met meer ervaring. Vaak gaat het meer om intervisie, waarin collega's met een ongeveer gelijk ervaringsniveau elkaar feedback geven en wederzijds problemen bespreken. Deze werkwijze vindt navolging binnen opleidingen, vanuit het idee dat 'ware kennis' niet noodzakelijk van boven afkomstig hoeft te zijn maar dat collega's elkaar veel kunnen leren. In de nascholing van docenten bestaat al een lange traditie van 'peer coaching' (Showers & Joyce, 1996). Ook het taal- en schrijfonderwijs werkt veel met intervisieachtige werkvormen, zoals 'peer revision', 'peer editing', 'peer response' en dergelijke (zie bijvoorbeeld Lockhart & Ng, 1995, McGroarty & Zhu, 1997; Amores, 1997). Meestal gaat het daarbij om de kwalitatieve feedback, maar er zijn ook voorbeelden waarin de studenten hun beoordeling van elkaars schrijfwerk kwantitatief uitdrukken, zoals het geval is bij 'calibrated peer review' (Chapman & Fiore, 1997).

In een overzichtsstudie van Dochy e.a. (1999) worden werkwijzen waarin studenten commentaar geven op elkaars werk onder de noemer 'peer assessment' bij elkaar gebracht en geëvalueerd. De aangetroffen definities voor deze term lopen uiteen van toetsing waarbij studenten elkaar alleen kwalitatieve feedback geven tot toetsing waarin studenten elkaar ook een cijfer geven. De auteurs concluderen dat peer assessment vooral effectief is als de nadruk ligt op het geven van formatieve feedback, opdat het beoordelen als leerinstrument kan fungeren. De betrokkenheid van de studenten neemt toe, ze staan er over het algemeen positief tegenover en zijn van mening dat het helpt om hun bekwaamheden te ontwikkelen. Rapportages waarin studenten hun beoordeling van elkaars werk moeten uitdrukken in de vorm van een cijfer laten een minder eenduidig positief beeld zien. Studenten neigen soms tot 'friendship marking', 'decibel marking'

(belonen van dominantie) en 'parasite marking' (belonen van meeliftgedrag). De auteurs zien dan ook als het om eindbeoordeling gaat meer in een combinatie met 'co-assessment', waarbij de resultaten van de peer assessment wel meewegen maar de docent het eindoordeel geeft. Zij besluiten hun overzicht met de conclusie dat het evident is dat studenten een substantiële rol kunnen spelen in de beoordeling, maar dat nog weinig bekend is over de voorwaarden waaraan een leeromgeving moet voldoen om peer assessment effectief in te kunnen zetten. Zijzelf achten 'training van studenten in het beoordelen' een belangrijke voorwaarde, evenals 'het samen met de studenten ontwikkelen van de beoordelingscriteria'.

In een review study heeft Topping factoren op een rij gezet die van belang zijn bij het ontwerpen van een geschikte leeromgeving voor peer assessment (Topping, 1998). Deze worden samengevat in tabel 1. Ook in deze studie wordt gerapporteerd over positieve effecten van peer assessment, met name waar het gaat om schrijfproducten. Peer assessment wordt hierbij opgevat als 'an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of learning of peers of similar status' (p. 250). Als belangrijke variabelen die te maken hebben met de functie van beoordelingsinstrument komen naar voren de gebruiksdoelen, focus en object van de beoordeling, de relatie met de docentbeoordeling en het gewicht van het oordeel (tabel 1, 1 t/m 6). Daarnaast zijn er variabelen met betrekking tot het sociale aspect, te weten oordeelsrichting, privacy en contact (7 t/m 9). Variabelen die te maken hebben met de samenstelling van de groep waarin beoordeeld wordt zijn jaar en bekwaamheid (10 t/m 13). Ten slotte worden variabelen gevonden op het vlak van verplichting en beloning (14 t/m 17). Daarnaast destilleert Topping nog acht factoren uit de literatuur die te maken hebben met de implementatie van peer assessment. Hierbij gaat het vooral om de manier waarop de studenten voorbereid worden op hun taken als beoordelaar, bijvoorbeeld door het wederzijds bespreken van doelen en verwachtingen, het verhelderen van de beoordelingscriteria en training in beoordelen (vergelijk ook Dochy e.a., 1999). Tezamen moeten ze een houvast geven voor een systematischer manier van ontwerpen, implementeren en evalueren van peer assessment, zodat ervaringen en uitkomsten beter vergeleken kunnen worden.

Bovengenoemde literatuur versterkte onze verwachting dat intercollegiale beoordeling een geschikte vorm van ondersteuning zou kunnen zijn voor de verwerving van schrijfvaardigheid binnen de opleiding Geschiedenis. In de eerste plaats krijgen de studenten op deze manier meer feedback en in een eerder stadium, die gebruikt kan worden bij het schrijven van de eindversie. Wanneer de feedback die de studenten elkaar geven goed is en wanneer ze de gekregen feedback goed gebruiken, moet de kwaliteit van de schrijfproducten er dus op vooruit gaan. Ook zou de tijd die de docent investeert in het nakijken en bespreken op den duur moeten kunnen afnemen. Daarnaast leek het ons een logische gedachte dat het regelmatig toepassen van de beoordelingscriteria op andermans werk er op langere termijn toe zal leiden dat die criteria worden geïnternaliseerd en als vanzelfsprekende kwaliteitsmaatstaven gaan dienen bij het schrijven en reviseren van het eigen werk.

Bij een schrijfcursus in het derde opleidingsjaar is een omgeving gericht op intercollegiale beoordeling ontworpen, geïmplementeerd en geëvalueerd.

Tabel 1 Typologie peer assessment (Topping, 1998)

	Variabele	Variatiemogelijkheden
1	vakgebied/onderwerp	op alle terreinen bruikbaar
2	gebruiksdoel	docenttijdsbesparing of leermiddel?
3	focus	kwantitatief/summatief of kwalitatief/formatief of beide?
4	object van beoordeling	schriftelijk werk, vaardigheidsdemonstratie?
5	relatie docentbeoordeling	vervangend of aanvullend?
6	gewicht van het oordeel	bijdrage aan het eindcijfer?
7	oordeelsrichting	eenzijdig of wederzijds?
8	privacy	anoniem, vertrouwelijk of openbaar?
9	contact	in directe aanwezigheid of op afstand?
10	jaar	studenten uit zelfde jaar of ander jaar?
11	bekwaamheid	met vergelijkbare of verschillende bekwaamheden?
12	constellatie beoordelaars	individueel, twee- drietallen, groepsgewijs?
13	constellatie beoordeelde	individueel, twee- drietallen, groepsgewijs?
14	plaats	binnen of buiten de klas?
15	tijd	lestijd/vrije tijd?
16	deelnameverplichting	verplicht of facultatief?
17	beloning	deelname, cijfer(onderdeel), studiepunt, vrijstelling?

Beschrijving van het onderwijsontwerp

Gezocht is naar een kansrijk onderwijsontwerp voor intercollegiale beoordeling dat past binnen het bestaande cursuskader. Daarbij staat de kwalitatieve kant van de feedback en de formatieve functie voorop. De studenten krijgen in elk geval tijd voor revisie van hun conceptwerkstuk na de intercollegiale beoordeling.

Het systeem is simpel:

- Stap 1 de studenten werken aan hun essay, paper of ander schrijfproduct;
- Stap 2 ze beoordelen elkaars tussenproducten aan de hand van in principe dezelfde lijst beoordelingscriteria die de docent hanteert bij de eindbeoordeling;
- Stap 3 de studenten vullen het beoordelingsformulier in en vatten de belangrijkste punten samen in de vorm van minstens drie adviezen voor verbetering; ze geven geen cijfer. Dit verslag geven ze aan de betrokken student, al of niet met een mondelinge toelichting;
- Stap 4 de studenten krijgen de gelegenheid om hun conceptversies te herzien.

De docent heeft hierbij voornamelijk de rol van procesbewaker. Hij ziet erop toe dat deze stappen worden uitgevoerd. Om zicht te kunnen houden op de voortgang van het proces ontvangt de docent de conceptversies en/of andere tussenproducten en de door de studenten uitgebrachte beoordelingen, maar hij onthoudt zich zoveel mogelijk van commentaar. Hij laat eerst de studenten elkaar feedback geven, pas daarna en voor zover hij dat nodig acht vult hij de studentbeoordelingen aan.

Tabel 2 geeft een overzicht van de kenmerken van het ontwerp van de intercollegiale beoordeling in de cursus.

Tabel 2 Typering onderwijsontwerp cf. 'typologie peer assessment' van Topping (1998)

	Variabele	Waarde
1	vakgebied/onderwerp	vakgebied Geschiedenis, specialisatie Internationale Betrekkingen
2	redenen: gebruiksdoelen	vergroting leeropbrengst
3	focus	de beoordeling is kwalitatief en formatief
4	object van beoordeling	schriftelijk werkstuk
5	relatie docentbeoordeling	aanvullend, de docent bepaalt het eindcijfer
6	gewicht van het oordeel	draagt niet bij aan het eindcijfer
7	oordeelsrichting	wederzijds
8	privacy	openbaar
9	contact	beoordeling is deels face to face
10	jaar	studenten uit dezelfde cursusgroep
11	bekwaamheid	studenten hebben in principe dezelfde kennis van het materiaal
12	constellatie beoordelaars	beoordelen in 3-tallen: ieder beoordeelt twee werkstukken; at random bij elkaar gezet
13	constellatie beoordeelde	ieder werkstuk krijgt twee beoordelingen
14	plaats	schriftelijke beoordeling geschiedt buiten contacttijd, mondelinge toelichting in contacttijd
15	tijd	verdisconteerd in de nominale studiebelasting
16	deelnameverplichting	verplicht
17	beloning	1: te beoordelen conceptversie bepaalt 50% van het cijfer voor de eindversie 2: de kwaliteit van het zelf gegeven intercollegiale oordeel telt voor 1/4 punt (10 pts schaal) mee met het eindcijfer

De keuze om studenten elkaar in drietallen te laten beoordelen is gemaakt omdat studenten in duo's maar van één persoon feedback zouden krijgen, wat we te weinig vonden. Beoordelen in groepen van vier of meer studenten zou betekenen dat ze te veel werkstukken moesten beoordelen. Het feit dat in deze cursus alle studenten gewoonlijk dezelfde opdrachten kregen die ze aan de hand van hetzelfde materiaal moesten uitwerken leek

ons bevorderlijk voor de interesse in elkaars werk, dus dat kenmerk hebben we in het ontwerp overgenomen. Omdat het ons belangrijk leek dat de studenten hun geschreven beoordelingen ook mondeling aan elkaar toelichtten kon de beoordeling niet anoniem plaatsvinden. Gezien het beperkte aantal contacturen en de aanzienlijke hoeveelheid leestijd die het beoordelen van twee werkstukken vergt lieten we de studenten de schriftelijke beoordeling buiten collegetijd uitvoeren. De mondelinge toelichtingen daarentegen vonden wel plaats in de werkgroep, in de eerste plaats omdat we vermoedden dat die er anders bij in zouden schieten en ten tweede omdat we wilden registreren wat er gezegd werd.

Hoewel ander onderzoek het belang benadrukt van training in beoordelen (Dochy e.a., 1999) hebben wij geen training in het programma opgenomen. In plaats daarvan hebben de studenten schriftelijke informatie gekregen over de procedure en is hun de gelegenheid geboden om vragen te stellen. Omdat de opleiding er naar streeft om met uniforme criteria te werken voor de beoordeling van schriftelijke werkstukken en er inmiddels een beoordelingslijst ligt die breed gedragen wordt door docenten besloten we om de criteria als min of meer gegeven aan de studenten te presenteren, zij het met een uitvoerige toelichting.

Implementatie in de bestaande cursus

Bovenbeschreven onderwijsontwerp is geïmplementeerd in een reguliere schrijfcursus van vier studiepunten voor derdejaarsstudenten. In deze cursus leren studenten korte artikelen (van 4 à 5 bladzijden) te schrijven voor verschillende doelgroepen, gebaseerd op beperkt, maar gedegen onderzoek van uiteenlopende schriftelijke en visuele bronnen. Zoals bij een krant of tijdschrift moet elk stuk worden geschreven binnen een strikt gehanteerde deadline, die afhankelijk van de moeilijkheidsgraad van de opdracht is gesteld op 24 uur of 48 uur (het materiaal moet in de dagen daarvoor zijn bestudeerd). De docent kijkt de artikelen na en geeft ze met bijgeschreven individueel commentaar en een cijfer terug op de eerstvolgende bijeenkomst. De cursus beslaat vijf weken, de studenten krijgen elke week een opdracht, die voor iedereen hetzelfde is. Het eindcijfer wordt gevormd door een gewogen gemiddelde van de opdrachten, waarbij rekening wordt gehouden met de toenemende moeilijkheidsgraad van de opdrachten. Per opdracht wordt een cijfer gegeven voor de kwaliteit van de 'analyse' en een cijfer voor het aspect 'taal'. 'Analyse' bepaalt tweederde van het eindcijfer, 'taal' telt mee voor eenderde.

In de vernieuwde opzet wordt bij één van de opdrachten met intercollegiale beoordeling gewerkt. Om de studenten daar genoeg tijd voor te geven zijn er in totaal vier opdrachten in plaats van de gebruikelijke vijf. De procedure is als volgt: op de eerste cursusbijeenkomst worden de studenten over de procedure geïnformeerd en alvast in drietallen verdeeld (at random). Deze drietallen zullen elkaars uitwerkingen van opdracht 2 in de week daarna beoordelen. In cursusweek 2 wordt van de studenten verwacht dat ze, net als bij de andere opdrachten, hun uitwerking binnen de gestelde tijdslimiet inleveren. Dit keer evenwel leveren ze de opdracht niet alleen bij de docent in, maar ook bij de twee medestudenten met wie ze samen een drietal vormen.

In de dagen daarna gaat ieder individueel aan de slag met het beoordelen van de twee ontvangen uitwerkingen. Dat gebeurt schriftelijk, door middel van een beoordelingsformulier. In dit formulier staan de belangrijkste beoordelingscriteria die de docent gebruikt bij de beoordeling van de einduitwerking. De docent heeft tijdens de eerste bijeenkomst de lijst criteria toegelicht en het gewicht van de verschillende aspecten voor de bepaling van het cijfer besproken. Om de kwaliteit van de *analyse* te beoordelen moet worden gelet op: 'inleiding (waaronder vraagstelling)', 'inhoud (waaronder correcte samenvatting van het materiaal in het licht van de opdracht en argumentatie)', 'betoon (waaronder indeling en samenhang)' en 'conclusie (waaronder aansluiting op de probleemstelling)'. Daarnaast komt het aspect *taal* (waaronder stijl en spelling) aan de orde. De beoordelaar geeft per criterium aan wat de kwaliteit van het stuk is (+ = 'zo houden', +/- = 'kan beter', - = 'moet beter'), gevolgd door een korte toelichting. Hierna wordt de beoordelaar gevraagd om de drie werkstukken, inclusief dat van hemzelf, te vergelijken op de aspecten 'betoon' en 'conclusie' als voorbereiding op de plenaire nabespreking waarin verschillende mogelijkheden voor de structuur van het werkstuk naast elkaar zullen worden gezet. Aan het einde van het formulier volgt de opdracht om het oordeel samen te vatten in de vorm van minstens drie adviezen voor verbetering.

De procedure vervolgt met een mondelinge toelichting op dit schriftelijk advies. In de bijeenkomst van de derde cursusweek lichten de studenten hun schriftelijke beoordelingen mondeling in de drietallen aan elkaar toe. Daarnaast maken ze een inventarisatie van de belangrijkste onderlinge verschillen die de vergelijking van hun stukken heeft opgeleverd. De vragen die daaruit voortkomen zijn onderwerp voor de aansluitende plenaire nabespreking. Na afloop van deze bijeenkomst krijgen de studenten naast het schriftelijke intercollegiale commentaar ook het, volgens afspraak globaal gehouden, docentcommentaar (vooralsnog zonder cijfer) terug. Ten slotte stellen ze hun eerste versie bij op basis van alle gekregen feedback, waarna ze het gereviseerde stuk opnieuw binnen de gebruikelijke deadline inleveren. In de week daarna krijgen ze deze versie terug met de eveneens per criterium toegelichte eindbeoordeling en een cijfer

Evaluatie

Onderzoeksvragen

Om te achterhalen wat precies de effecten van intercollegiale beoordeling zijn en in hoeverre het gekozen ontwerp voldoet, zijn vragen gesteld met betrekking tot het *proces*:

- 1 Hoe reageren studenten en docenten op de procedure van intercollegiale beoordeling?
- 2 Hoeveel tijd besteden de studenten aan het beoordelen van elkaars werk, aan het schrijven van hun eerste versie en aan het reviseren ervan?
- 3 Hoeveel tijd besteedt de docent aan het beoordelen en feedback geven?

Daarnaast zijn vragen gesteld die gericht zijn op de *effecten*:

- 4 Wat is de kwaliteit van de intercollegiale beoordelingen?
- 5 Wat doen de studenten met het intercollegiale commentaar dat ze krijgen?
- 6 Zijn de eindversies beter als gevolg van intercollegiale beoordeling?

Met betrekking tot de vragen over de effecten kunnen slechts indicaties worden verkregen die zijn gebaseerd op de evaluatie van één onderwijsontwerp.

Onderzoeksmethode

De groep die deelnam aan het onderwijs bestond uit dertien studenten. De benodigde gegevens zijn hoofdzakelijk verkregen uit twee vragenlijsten afgenomen bij de studenten en een interview na afloop van de cursus met de docent. Daarnaast zijn werkcolleges geobserveerd om informatie te krijgen over de manier waarop de studenten reageerden op de procedure. De eerste vragenlijst is afgenomen nadat de studenten hun schriftelijke beoordeling mondeling aan elkaar hadden toegelicht. Het gaat hier om open vragen, bedoeld om inzicht te krijgen in de tijd die de studenten hebben besteed aan het beoordelen, hun waardering van de beoordelingsinstrumenten en de opbrengst die de studenten ervaren of verwachten. De tweede vragenlijst is afgenomen na de laatste onderwijsbijeenkomst en is bedoeld ter evaluatie van de cursus. De vragen zijn voor het merendeel gesloten en corresponderen voor het merendeel met de gebruikelijke gestandaardiseerde evaluatie van deze cursus. De ruimte voor open vragen is gebruikt om opnieuw naar de waardering van het intercollegiale beoordelen te vragen. De studenten weten dan inmiddels wat hun eindcijfer voor de na intercollegiale beoordeling herziene opdracht is, en hebben daarna nog enkele opdrachten volgens de normale procedure moeten inleveren.

In het (half gestructureerde) interview met de docent werden vragen gesteld over de tijd die de docent had besteed aan het beoordelen na de intercollegiale beoordeling en zijn waardering van het werken met intercollegiale beoordeling. Tevens werd hem de vraag gesteld of hij bepaalde verschillen zag qua betrokkenheid en participatie van de studenten in vergelijking met de keren dat hij de cursus in de normale vorm uitvoerde. Het interview is met een bandrecorder geregistreerd en uitgetypt.

Ten slotte zijn alle geschreven producten verzameld (concepten en definitieve versies), als ook het commentaar van de studenten (zowel de schriftelijke als de mondelinge feedback) op elkaars werk en het commentaar van de docent. De schriftelijke feedback is gegeven in de vorm van een invulformulier. De mondelinge feedback van elk beoordelingsdrieliet is met een bandrecorder geregistreerd en uitgetypt.

Op de mondelinge feedback is inhoudsanalyse toegepast. De schriftelijke feedback, voorgestructureerd aan de hand van de beoordelingscriteria, fungeerde daarbij als uitgangspunt waarna gekeken werd wat de mondelinge toelichting hieraan toevoegde. De verzamelde concept- en eindversies van de werkstukken zijn gebruikt om te onderzoeken wat de studenten hebben gedaan met het intercollegiale commentaar dat ze kregen en wat ze deden met het commentaar van de docent.

De gegevens uit de vragenlijsten, voor zover het de gesloten vragen betreft, worden gepresenteerd door middel van beschrijvende statistiek. De antwoorden op de open vragen en de antwoorden uit het interview zijn gebruikt bij de beschrijving van het proces.

Resultaten

Uitvoering en waardering

De uitvoering van bovenbeschreven onderwijsontwerp verliep zoals bedoeld: de studenten hadden hun conceptversie op tijd af, de uitwisseling van de werkstukken ging zonder problemen. Wel haakten twee studenten op het laatste moment af, om redenen die niets met het experiment te maken hadden, waardoor er twee duo's in plaats van uitsluitend trio's ontstonden.

Hoewel de studenten geen ervaring hadden met de procedure, stelden ze er weinig vragen over, noch vooraf noch tijdens de uitvoering. Er waren geen opmerkingen waaruit bleek dat iemand het idee om studenten elkaar feedback te laten geven ter discussie wilde stellen.

Men vulde de beoordelingsformulieren over het algemeen vrij volledig in, daarnaast werd ook veel op de werkstukken zelf aangetekend om het voor de beoordeelde duidelijk te maken op welke passages de opmerkingen van de beoordelaar precies betrekking had. De mondelinge besprekingen in de drietallen waren geanimeerd en taakgericht en werden door de meeste studenten als een waardevolle uitbreiding op het schriftelijk commentaar ervaren. De toelichting maakte het commentaar concreter en doordat ook de andere beoordelaar met feedback kwam ontstond er meer duidelijkheid over prioriteiten bij de revisie. Zoals verwacht was de beoordeelde het niet altijd eens met het commentaar en werd de gelegenheid benut om eigen standpunten te verduidelijken. Dit leidde echter niet waarneembaar tot gespannen situaties. Waarschijnlijk omdat iedereen beurtelings in de rol van beoordelaar en beoordeelde gezet wordt en niemand langer dan vijftien minuten 'de klos' is, bleef de stemming over het algemeen vriendelijk en vrij zakelijk.

Ook de plenaire nabespreking die erop volgde was levendig: de onderwerpen werden door de studenten ingebracht als punten waar ze onderling niet uitkwamen en de docent nu wel graag over wilden horen.

Niemand had problemen met het feit dat de intercollegiale beoordeling niet anoniem plaatsvond. Als je elkaar toch nauwelijks kent, maakt het niet veel uit wiens stuk je beoordeelt, vond men. Het werken met een beoordelingsformulier werd door sommigen iets te beperkend ervaren, er was qua lay-out te weinig ruimte voor alle kritiek en suggesties.

Uiteindelijk bleken elf van de dertien de studenten het werken met deze vorm van intercollegiale beoordeling positief te waarderen. Ze vonden de tijd die het kostte om de stukken van hun collega-studenten te beoordelen niet te veel en een nuttige investering. Sommigen voegden eraan toe dat het vergelijken van hun eigen stuk met dat van anderen hen meer zicht gaf op de kwaliteit van hun eigen werk. Het constateren van overeenkomsten gaf hun naar eigen zeggen meer zelfvertrouwen, terwijl het vaststellen van verschillen juist kritischer naar de eigen tekst deed kijken. Anderen zagen het nut vooral in het feit dat je ideeën voor je eigen werkstuk op doet, met name ten aanzien van indeling, stijl en formuleringskwesties.

De mogelijkheid tot reviseren met behulp van het gekregen commentaar werd zeer gewaardeerd, bijna iedereen vond dat er veel nuttig commentaar was gegeven en wilde daar iets mee doen. Op twee studenten na zou iedereen vaker met intercollegiale beoordeling willen werken. Voor de meesten van hen is er geen reden om aan de opzet ervan iets te veranderen.

De docent hield zich volledig aan de afgesproken procedure, al vond hij het naar eigen zeggen moeilijk om zich terughoudend op te stellen bij zijn commentaar op de eerste versie van de opdracht. Zijn vrees was dat sommigen uit zijn terughoudendheid ten onrechte zouden afleiden dat ze niet veel meer aan hun werkstuk hoefden te doen. Niettemin is zijn oordeel aan het eind van de cursus dat intercollegiale beoordeling een werkwijze is die structureel in het curriculum zou moeten worden opgenomen. Hij is van plan om intercollegiale beoordeling als vast onderdeel op te nemen in deze cursus, in min of meer dezelfde vorm.

Tijdsinvestering studenten

De tijd die de studenten hadden besteed aan het lezen en beoordelen van twee werkstukken bedroeg anderhalf uur per werkstuk, hetgeen overeenkomt met de tijd die ons minimaal nodig leek. Wat betreft hun schrijftijdsinvestering viel het ons op dat studenten met een laag cijfer relatief veel meer tijd hadden besteed aan het herschrijven dan de betere studenten. Daarentegen hadden ze minder tijd besteed aan het schrijven van hun eerste versie.

De middelmatige studenten zitten wat betreft die verhouding tussen de goede en de zwakke studenten in. Tabel 3 geeft een overzicht van de schrijftijdsinvestering per cijfergroep.

Tabel 3 Cijfers in relatie tot schrijftijd van beide versies

(v1 = eerste versie; v2 = tweede versie; s1 = student 1; s2 = student 2; s3= student 3; tijd v1 en v2 is in minuten)

Studenten	Cijfer v1	Cijfer v2	Tijd v1	Tijd v2	Totaal	Tijd v1/v2
<i>Cijfers 7,5-8</i>						
s1	7,5	8	420	180	600	2,3
s2	7,5	8	390	75	465	5,2
s3	8	8,5	600	75	675	8
Gemiddeld			470	110	580	4,3
<i>Cijfers 6-7</i>						
s4	6	7	480	360	840	1,3
s5	6,5	7,5	720	120	840	6
s6	7	7,5	240	120	360	2
Gemiddeld			480	200	680	2,4
<i>Cijfers 3,5-5,5</i>						
s7	3,5	4,5	300	300	600	1
s8	4	4,5	570	465	1035	1,2
s9	4,5	4,5	600	360	960	1,6
s10	4,5	5	255	255	510	1
s11	5	5,5	270	90	360	3
s12	4,5	7	480	300	780	1,6
s13	5	7	360	90	450	4
Gemiddeld			405	266	670	1,5

Misschien kan het verschil in tijd die aan revisie besteed is deels verklaard worden uit onze bevinding dat de zwakkere studenten minder gedetailleerd commentaar kregen dan de goede studenten. In dit geval evenwel is de groep te klein om er verder reikende uitspraken over te doen, vervolgonderzoek zal meer licht moeten werpen op de vraag of het hier gaat om een toevallig verschijnsel of dat het structureel is.

Tijdsinvestering docent

Volgens de docent was de tijd die hij besteedde aan de cursus in deze opzet niet veel groter of kleiner geweest dan gewoonlijk. Eén opdracht viel weg en hij moest een tweede versie nakijken. Hij had eigenlijk verwacht dat het nakijken van twee versies minder tijd zou kosten dan het nakijken van een weggevalle vijfde opdracht, maar het vergelijken en het aangeven van de verbeteringen, ook het invullen van de beoordelingsformulieren kostte meer tijd dan gedacht. Hij vond dat evenwel geen bezwaar, ook omdat hij aan het commentaar van de studenten zag dat ze serieus met elkaars stukken bezig waren geweest en van daaruit elkaar vaak zinvolle aanwijzingen voor verbetering hadden gegeven.

Kwaliteit intercollegiale beoordeling

Om een indicatie te krijgen van de kwaliteit van de intercollegiale beoordelingen is gelet op de gedetailleerdheid waarmee het beoordelingsformulier werd ingevuld. Daarnaast is het mondelinge commentaar geanalyseerd om te zien of dit een verdieping en/of uitbreiding gaf op het schriftelijk commentaar. Het intercollegiale commentaar is vergeleken met de feedback van de docent, om te zien wat de verschillen en overeenkomsten zijn. Ten slotte is het oordeel van de docent over de kwaliteit van het studentcommentaar gebruikt als indicatie.

Uit het schriftelijke commentaar dat de studenten aan elkaar geven blijkt dat ze er redelijk in slagen om alle beoordelingscriteria toe te passen. Alle studenten vullen alle vragen op de formulieren in, sommigen gebruiken de geboden ruimte daarbij helemaal, anderen kunnen met minder toe. Ze beoordelen elkaars stukken op alle criteria. Daarnaast maken ze vaak aantekeningen op de te beoordelen werkstukken.

Vergelijking van de schriftelijke beoordelingen met het docentcommentaar laten in de meeste gevallen een behoorlijke overlap zien, zowel wat betreft de beoordelingsaspecten vallend onder 'analyse' als die onder 'taal'. Het studentcommentaar is meestal uitvoeriger dan het docentcommentaar dat, zoals afgesproken, globaal gehouden is.

De docent vond de kwaliteit van het schriftelijk commentaar wisselend, maar over het algemeen behoorlijk. Van de dertien studenten gaf hij er acht een '+' of '++' voor hun uitgebrachte beoordeling. De docent verklaarde in het interview na afloop van het onderwijs dat hij vond dat de studenten heel zinvolle punten inbrachten voor de plenaire nabespreking. De mondelinge toelichtingen in de subgroepen heeft hij niet beluisterd.

Dit duidt er ons inziens op dat het merendeel van de studenten in deze cursus er behoorlijk in is geslaagd om de werkstukken van hun medestudenten kritisch tegen het licht te houden. Wel is het zo dat de zwakke studenten minder goede feedback (in termen van gelijkend op die van de docent) geven, ook al besteden ze meer tijd dan de betere studenten aan het lezen en beoordelen van de werkstukken.

Wat opvalt is dat de goede studenten zowel van hun medestudenten als van de docent vaak gedetailleerder commentaar krijgen en concretere aanwijzingen dan de zwakke studenten.

Verwerking van het intercollegiale commentaar

Vergelijking van de eerste en tweede versies laat zien dat bijna alle studenten wel iets gedaan hebben met de intercollegiale feedback, zowel met het schriftelijke als met het mondelinge commentaar. In veel gevallen wordt datgene gebruikt uit het intercollegiale commentaar, dat ook in het commentaar van de docent (dat conform de procedure globaler gehouden is) wordt aangestipt. Niet alle commentaar wordt overgenomen, noch dat van de docent, noch van de medestudenten. Het docentcommentaar bevat naast globale aantekeningen bij structuur en inhoud ook tamelijk gedetailleerde correcties en suggesties met betrekking tot taal. Opmerkelijk is dat juist deze eenvoudig te verwerken zaken nogal eens blijven liggen bij het reviseren.

Kwaliteit van de eindversies

De cijfers voor de tweede versies zijn in vergelijking met die voor de eerste versie bij vrijwel alle studenten omhoog gegaan. De cijfervoortgang wordt gepresenteerd in tabel 4. Bij acht studenten is het cijfer met 0,5 punt verhoogd, bij twee studenten is het 1,0 punt verbeterd, bij twee studenten zelfs 2,0 tot 2,5 punt. Niettemin lukt het vijf studenten niet om aan een voldoende te komen. Eén student (s9) boekte geen voortgang.

Tabel 4 Cijfervoortgang na intercollegiale beoordeling (versie 1 > versie 2)

Beginscore	Gr	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8	8,5
<i>Cijfers 7,5-8</i>												
s1	A									→	→	
s2	D									→	→	
s3	E										→	→
<i>Cijfers 6-7</i>												
s4	C						→	→				
s5	D							→	→	→		
s6	A								→	→		
<i>Cijfers 3,5-5,5</i>												
s7	E	→	→	→								
s8	A		→	→								
s9	B			→								
s10	C			→	→							
s11	B				→	→						
s12	E			→	→	→	→	→	→			
s13	D				→	→	→	→	→			

Zoals tabel 5 laat zien verschilt de cijferwinst per groepje en daarbinnen de cijferwinst per individueel groepslid. Zo is bij groepje D (een goede, een middelmatige en een zwakke student) en E (een goede en twee zeer zwakke studenten) de gemiddelde cijferwinst per groepje en per individu aanzienlijk groter dan bij groepje B (een duo van twee zwakke studenten) en C (een duo van een zwakke en een middelmatige student). Ook in groepje A (twee goede studenten en een zeer zwakke student) is de cijferwinst per groepje en per individu niet zo groot.

Tabel 5 Winst per groepje (A t/m E) en gemiddelde cijferwinst per individueel groepslid

Groep	Tot.scores v1	Tot.scores v2	Winst per gr	Winst pp
A (3)	18,5	20	1,5	0,5
B (2)	9,5	10	0,5	0,25
C (2)	10,5	12	1,5	0,75
D (3)	19	22,5	3,5	1,17
E (3)	16	20	4,0	1,33

De verbeteringen bij de studenten die al een goed cijfer hadden voor hun eerste versie zijn relatief klein en betreffen meestal zowel de 'analyse' als de taalaspecten. De aangebrachte veranderingen in de tekst zijn meestal gering. De studenten die hun cijfer behoorlijk wisten op te trekken hadden dit vooral te danken aan de verbeteringen die ze aanbrachten op de aspecten met betrekking tot 'analyse' (die, zoals ze wisten, het zwaarst meewogen in het eindcijfer). Maar, zoals gezegd, vijf studenten slaagden er ondanks hun inspanningen niet in om hun stuk met succes te reviseren.

Conclusie

We weten nu dat dit ontwerp heeft gewerkt. De procedure was voor iedereen uitvoerbaar, de gevraagde tijdsinvestering werd door de studenten als redelijk en lonend ervaren en hun werkstukken zijn beter geworden als gevolg van de intercollegiale beoordeling. Zowel de studenten als de docent zouden vaker met intercollegiale beoordeling willen werken, wat hen betreft hoeft er aan het ontwerp niets te veranderen. Toch hebben we ook geconstateerd dat intercollegiale opzet in deze vorm niet voor elke student optimaal werkt. In deze bescheiden opzet was het niet mogelijk om er achter te komen waarom. Evenmin weten we wat er precies werkt in het ontwerp. Van de gemaakte keuzes hebben we er geen enkele hoeven te betreuren, maar we weten niet hoe dit ontwerp uitpakt bij andere studenten en met een andere docent als uitvoerder. Wel levert de analyse van de gegevens enkele suggesties voor vervolgonderzoek op. Zo denken we dat de samenstelling van de beoordelingsgroepjes in termen van het creëren van een optimale mix tussen goede, middelmatige en zwakke studenten meer aandacht vraagt. Ook is het nodig om goede feedbackprocedures te vinden, waar de zwakkere studenten hun voordeel mee kunnen doen.

Noot

¹ Wij gebruiken de term 'intercollegiale beoordeling' in plaats van termen als 'peer assessment' en 'peer review' omdat de studenten en docenten van de opleiding Geschiedenis de voorkeur gaven aan een Nederlands equivalent.

Literatuur

- Amores, M.J. (1997) A new perspective on peer-editing. *Foreign Language Annals*, 513-522.
- Chapman, O. L. & Fiore, M.A. (1997) *Calibrated Peer Review*. Regents University of California.
- Dochy, F., Segers, M. & Sluijsmans, D. (1999) The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, 331-349.
- McGroarty, M.E. & Zhu, W. (1997) Triangulation in classroom research: a study of peer revision. *Language Learning*, 1-43.
- Holleman, J.W., Oost, H. A., Milius, J.J., Berg, B.A.M. van den & Admiraal, W.F. (1999) *Gevraagd: academicus (m/v)*. IVLOS-Mededeling nr.61. Utrecht: Universiteit Utrecht.
- Lockhart, C. & Ng, P. (1995) Analyzing talk in ESL peer response groups: stances, functions, and content. *Language Learning*, 605-655.
- Showers, B. & Joyce, B. (1996) The evolution of peer coaching. *Educational Leadership*, 53,(6), 12-17.
- Topping, K. (1998) Peer assessment between students in colleges and universities. *Review of Educational Research*, 249-276.