



**Universiteit Utrecht**

Een Zelfontworpen Academisch Instrument met als Meetpretentie Niveaubepaling van  
Wetenschappelijk Redeneren bij Leerlingen in het VMBO Leerjaar 1 en 2: een Onderzoek  
naar de Psychometrische Kwaliteiten.

Masterthesis Orthopedagogiek (200500130)

Studiejaar 2014-2015

Universiteit Utrecht

Masteropleiding Pedagogische Wetenschappen

Masterprogramma Orthopedagogiek

Thesisbegeleider: Willemijn Schot, Hanna Mulder

Naam: Danny Berendse

Studentnummer: 3670600

### Voorwoord

Vanaf September 2014 zijn mijn thesispartner Lonneke Spierings en ik gestart met het bedenken en ontwikkelen van een eigen meetinstrument, met als meetpretentie wetenschappelijk redeneren bij adolescenten in het vmbo leerjaar 1 en 2. Het thema wetenschappelijk redeneren was voor mij nieuw, net als het ontwerpen en ontwikkelen van een eigen meetinstrument. Nadat er voldoende literatuur was verzameld over het construct wetenschappelijk redeneren en over testontwikkeling, zijn Lonneke en ik gestart met het filmen en monteren van het instrument zelf. Naast het gehele jaar werken aan de masterthesis, heb ik stage gelopen en de colleges met bijbehorende tentamens moeten afleggen. Al met al heeft dit veel van mij gevraagd, waardoor ik niet altijd even veel tijd en aandacht in de masterthesis kon steken. Ik wil daarom allereerst mijn twee thesisbegeleiders, Willemijn Schot en Hanna Mulder, bedanken voor hun behulpzame maar vooral geduldige manier van begeleiding. Willemijn, zonder jouw flexibele houding op het eind had ik deze thesis niet eens op tijd kunnen inleveren. Bedankt daarvoor! Tot slot wil ik mijn thesispartner Lonneke Spierings bedanken voor het vele meedenken en tonen van doorzettingsvermogen op momenten dat ik het juist even niet meer zag zitten.

Danny Berendse

Amersfoort, Juli 2015

### Samenvatting

Wetenschappelijk redeneren omvat een breed scala aan cognitieve processen die nodig zijn voor cognitieve vaardigheden zoals kritisch denken en redeneren. Hoewel eerder onderzoek zich wel heeft gericht op de functie en het belang van wetenschappelijk redeneren, bestaat er voor vmbo leerlingen uit leerjaar 1 en 2 op het voortgezet onderwijs nog geen valide en betrouwbaar meetinstrument om dit cognitieve concept bij hen te meten. Deze studie richt zich op het onderzoeken van de psychometrische kwaliteiten van het zelfontworpen instrument wetenschappelijk redeneren. 84 participanten die gelijk verdeeld zijn over twee middelbare scholen voor vmbo onderwijs, hebben dit instrument ingevuld. In de eerste plaats heeft het onderzoek zich gericht op de betrouwbaarheid, door de factorstructuur te onderzoeken. In de tweede plaats richt het onderzoek zich op de validiteit, door op basis van literatuur verwachte scoreverschillen of- samenhangen te onderzoeken op variabelen die in het onderzoek zijn meegenomen. De betrouwbaarheid van de individuele factoren alsmede het instrument als geheel bleken onvoldoende te zijn. Wat betreft de validiteit zijn de verwachte scoreverschillen op de variabelen opleidingsniveau en leeftijd niet gevonden. Wel is er naar verwachting geen verschil gevonden op de variabele geslacht, alsmede een samenhang met de variabele nieuwsgierigheid naar wetenschap. Echter, omdat het instrument onvoldoende betrouwbaar is bevonden, moeten deze uitslagen in twijfel worden getrokken. Ondanks dat de psychometrische kwaliteiten van het zelfontworpen instrument onvoldoende zijn, biedt dit onderzoek wel ingang voor vervolgstudies. Indien vervolgonderzoek de beperkingen van het huidige onderzoek ondervangt, is de verwachting dat er een meer betrouwbaar en valide meetinstrument ontworpen kan worden.

*Kernwoorden:* wetenschappelijk redeneren, zelfontworpen meetinstrument, psychometrische kwaliteiten, vmbo leerjaar 1 en 2

Abstract

Scientific reasoning contains a wide range of cognitive processes needed for cognitive skills like critical thinking and reasoning. Although previous research has focused on the function and the importance of scientific reasoning, for vmbo students in de first en second grade of secondary education, there still no is a valid and reliable measuring instrument available to measure them in this cognitive concept. This study focuses on the investigation of the psychometric qualities of the self-designed instrument of scientific reasoning. 84 participants equal spread over two high-schools for VMBO education have completed the questionnaire's. First, the research has focused on the reliability, by investigate the factor structure. In the second place, the research focuses on the validity, by studying expected score differences or score cohesions on the basis of literature investigations on variables that are included in the research . The reliability of the individual factors as well as the instrument as a whole turned out to be unreliable. As regards to the validity the expected score differences on the variables seks and educational level are not found. The prediction to found no sex differences is found, and also a interaction effect between curiosity about science and scores on the self-designed instrument. However, because the self-designed instrument has been found unreliable, should these results be doubted. Despite that the psychometric qualities of the self designed instrument are unreliable, this research provides access to follow-up studies. If follow-up research overcomes the limitations of the current research it is expected that there can be disigned a more reliable and valid instrument.

*Key words:* scientific reasoning, self-designed measuring instrument, psychometric qualities, VMBO first and second grade

Een Zelfontworpen Academisch Instrument met als Meetpretentie Niveaubepaling van Wetenschappelijk Redeneren bij Leerlingen in het VMBO Leerjaar 1 en 2: een Onderzoek naar de Psychometrische Kwaliteiten

Wetenschappelijk redeneren omvat een breed scala aan procedurele en conceptuele activiteiten die nodig zijn voor cognitieve vaardigheden zoals kritisch denken en redeneren (Bao et al., 2009; Klahr, Zimmerman, & Jirout, 2011). Een veel gebruikte methode om het niveau van wetenschappelijk redeneren bij kinderen te bepalen, is het met de participant doorlopen van ontdekkingsstaken binnen een wetenschappelijk thema. Niveaubepaling van wetenschappelijk redeneren zou ook bij adolescenten ingezet kunnen worden als een extra indicator voor niveaubepaling in het onderwijs, gezien het niveau van wetenschappelijk redeneren nauw samenhangt met de onderwijsprestaties van leerlingen (Komarraju, Karau, Smeck, & Avdic, 2011). Echter, er bestaat reeds geen valide en betrouwbaar meetinstrument om dit cognitieve concept bij adolescenten te kunnen meten. De doelstelling van het huidige onderzoek is daarom een evaluatie van de psychometrische kwaliteiten van een zelf ontworpen instrument met als meetpretentie het niveau van wetenschappelijk redeneren bij adolescenten.

Relevant voor het niveau van wetenschappelijk redeneren is de mate van nieuwsgierigheid (Klahr et al., 2011). Nieuwsgierigheid speelt namelijk een cruciale rol bij het faciliteren van de cognitieve ontwikkeling (Tamdogon, 2006). Mensen die nieuwsgierig zijn naar aangeboden informatie verwerken de informatie op een dieper niveau, onthouden meer van wat zij hebben gelezen en behalen hogere cijfers op school (Silvia, 2005; Silvia, 2008). Nieuwsgierigheid wordt daarnaast gezien als een faciliterend component voor leren in het algemeen (Dewey, 1910/ 2009 ed.) maar ook als belangrijk cognitief aspect voor het ontwikkelen van een wetenschappelijk manier van denken (Klahr et al., 2011). Wanneer nieuwsgierigheid gericht is op de wereld om ons heen wordt dit wetenschappelijke nieuwsgierigheid genoemd (Krapp & Prenzel, 2011; Spektor-Levy, Kesner Baruch, & Mevarech, 2013).

Wetenschappelijk redeneren in de breedste zin omvat vele procedurele en conceptuele activiteiten zoals vragen stellen, hypothesen opstellen, experimenten ontwerpen, experimenten uitvoeren, observeren, meten, voorspellen, interpreteren van data, bewijs evalueren, statistische berekeningen uitvoeren, conclusies trekken en het opstellen van theorieën en modellen (Keys, 1994; Klahr et al, 2011; Kuhn, Amsel, & O'Loughlin, 1988; Schauble, Glaser, Duschl, Schulze, & John, 1995; Zimmerman, 2007). Vanwege deze

complexiteit is er een tweedeling gevormd (Zimmerman, 2000). Aan de ene kant spreekt men van een domeinspecifieke benadering: de mate van kennis welke een individu heeft van een wetenschappelijk concept zoals biologie. Aan de andere kant spreekt men van een procedurele benadering: de mate van kennis en vaardigheden die een persoon bezit voor het kunnen opzetten en uitvoeren van wetenschappelijk onderzoek (Klahr, Zimmerman & Jirout, 2011). Het gaat hierbij om het in allerlei verschillende situaties kunnen toepassen van onderzoeksvaardigheden zoals: het opstellen van een hypothesen (Klahr & Dunbar, 1988), reeds geformuleerde hypothesen toetsen, het bewijs evalueren (Germann, Aram, & Burke, 1996), het trekken van conclusies, het opstellen van theorieën en modellen (Keys, 1994; Schauble, Glaser, Duschl, Schulze, & John, 1995) en indien nodig het bijstellen van hypothesen (Klahr & Dunbar, 1988; Kuhn et al., 1988). Op basis van bovenstaande literatuur lijkt de procedurele benadering van wetenschappelijk redeneren te kunnen worden verdeeld in drie deelvaardigheden te weten: (1) het opstellen van hypothesen, (2) het ontwerpen van experimenten en (3) het voeren van observaties en evalueren van bewijs (Klahr et al., 2011; Kuhn et al., 1988; Zimmerman, 2007).

Eerder onderzoek naar wetenschappelijk redeneren, richt zich met name op de ontwikkeling- en beschrijving van de cognitieve componenten die nodig zijn om tot wetenschappelijk redeneren te komen. Een studie uitgevoerd door Germann, Aram, Odom en Burke (1996) vormt daarop een belangrijke uitzondering voor de huidige studie. Germann, (1996) poogden de ontwikkeling van een onderzoekskader waarmee antwoorden van studenten op taken als het stellen van vragen, identificeren van variabelen en het opstellen van hypothesen konden worden geanalyseerd en geclassificeerd. Op basis van dit kader is het antwoordformat ontwikkeld waarmee studentscores op de zelfontworpen instrument wetenschappelijk redeneren zijn geanalyseerd en geclassificeerd.

In de eerste plaats zal het huidige onderzoek zich richten op het onderzoeken van de factorstructuur van het instrument wetenschappelijk redeneren. Op basis van bovenstaande literatuur over de procedurele benadering van wetenschappelijk redeneren, is de hypothese opgesteld dat ‘er op het instrument wetenschappelijk redeneren een beoogd drie-factorenmodel wordt gevonden te weten: (1) het opstellen van hypothesen, (2) ontwerpen van experimenten en (3) het voeren van observaties en evalueren van bewijs”

In de tweede plaats zal het onderzoek zich richten op de criteriumvaliditeit van het instrument wetenschappelijk redeneren. Dit wordt onderzocht door de op basis van literatuur wel of geen verwachte samenhang te onderzoeken tussen scores op het instrument

wetenschappelijk redeneren en achtergrondvariabelen, namelijk geslacht, leeftijd en opleidingsniveau. In haar review stelt Spelke (2005) dat er wat betreft het niveau van wetenschappelijk redeneren tussen mannen en vrouwen, er geen significante verschillen zijn gevonden in eerder onderzoek. Zimmerman (2000) heeft onderzocht in hoeverre kinderen en adolescenten in staat zijn tot het kunnen leggen van causale verbanden op basis van aangereikte informatie. Hieruit is gebleken dat de variabelen leeftijd en opleidingsniveau positief samenhangen met de mate van causale verbanden kunnen leggen. Op basis van deze literatuur zijn de volgende hypothesen opgesteld, namelijk “de verwachting is dat scores op het instrument wetenschappelijk redeneren significant hoger uitvallen naarmate de leeftijd en het opleidingsniveau van leerlingen hoger liggen” en “de verwachting is dat er geen significant verschil in scores wordt gevonden tussen jongens en meisjes.

In de derde plaats zal het onderzoek zich richten de convergente validiteit van het instrument wetenschappelijk redeneren. Het idee achter de convergente validiteit is dat als de metingen op het construct wetenschappelijk redeneren valide zijn, en als het construct volgens theorie zou moeten samenhangen met een ander construct, eenzelfde samenhang kan worden teruggevonden in het huidige onderzoek. Op basis van de eerder beschreven samenhang tussen wetenschappelijk redeneren en nieuwsgierigheid naar wetenschap, is voor het bepalen van de externe validiteit de hypothese opgesteld “dat scores op het instrument wetenschappelijk redeneren significant correleren met de variabele nieuwsgierigheid naar wetenschap”.

## **Methoden**

### **Populatie van de Steekproef**

De populatie bestond uit 84 participanten, waarvan 41 jongens en 43 meisjes. De gemiddelde leeftijd was 13,7 jaar met een bereik tussen de 12,1 en 15,1. Alle participanten bevonden zich op het Voortgezet Onderwijs (VO) waarvan 42 in leerjaar 1 en eveneens 42 in leerjaar 2. In totaal volgden 23 participanten de basisberoepsgerichte leerweg (BBL), 35 de kaderberoepsgerichte leerweg (KBL) en 26 de theoretische leerweg (TL). De populatie was gelijk verdeeld over twee middelbare scholen voor vmbo onderwijs. In dit onderzoek was er sprake van een 'gemakssteekproef' en een selecte keuze van de populatie.

### **Instrumenten**

**Zelfontworpen instrument voor wetenschappelijk redeneren.** Het eerste instrument dat is afgenomen was het zelfontworpen instrument voor wetenschappelijk redeneren (zie bijlage 1). Bij het ontwerpen van dit instrument was als basis uitgegaan van een assessment

genaamd *Directed Inquiry Approach to Learning Science Process Skills and Scientific Problem Solving* (DIALSPS) uit het artikel van Germann (1989b), welke een format biedt voor het toetsen van het niveau van wetenschappelijk redeneren bij 12 en 13 jarigen (Germann, Aram, Odom & Burke, 1996).

Voor het bereiken van een zo hoog mogelijke indruks-validiteit is het instrument voorgelegd aan docenten werkzaam binnen het vmbo. Tevens is het instrument voorgelegd aan een professor van de Universiteit Utrecht. De verkregen feedback over het taalgebruik van het instrument, formulering van de items en het theoretisch niveau van de items, is verwerkt in de eindversie van het instrument.

Voor het bereiken van een acceptabele interbeoordelaarsbetrouwbaarheid, hebben de twee beoordelaars 16 procent van de totale hoeveelheid afgenomen instrumenten, individueel van elkaar gescoord en elk afgenomen instrument een totaalscore gegeven. De mate van overeenstemming tussen de twee beoordelaars is gemeten aan de hand van Cohen's Kappa en beoordeeld als goed ( $\kappa = .76, p < .001$ ). Daarna hebben de beoordelaars individueel van elkaar ieder de helft van de resterende instrumenten gescoord.

**SDDRQ voor nieuwsgierigheid naar wetenschap.** Het tweede meetinstrument dat is afgenomen was de '*Scientific Demonstration with Documented Response Questionnaire*' ([SDDRQ], Kesner-Baruch et al, 2014). Dit instrument gebruikt verbale en gedragsmatige reacties als potentiële indicatoren voor de nieuwsgierigheid naar- en de houding tegenover wetenschap. Dit instrument is bij de helft van de totale steekproef ( $n=42$ ) afgenomen. De participanten kregen o.a. de vraag of zij wilden ruiken of voelen aan een vloeistof tijdens het experiment. Een ander soort vraag die zij halverwege het experiment moesten invullen, is of zij konden aangeven wat zij dachten dat er zou gaan gebeuren gaandeweg het experiment. In het oorspronkelijke artikel is dit instrument afgenomen bij peuters, maar uit een testpilot bleek het instrument qua niveau ook geschikt te zijn voor de populatie van het huidige onderzoek. De gedragsmatige reacties zijn gescoord op basis van observaties. De verbale reacties zijn op basis van een Likert-schaal gescoord. Twee beoordelaars hebben individueel van elkaar 100% van de verbale reacties een Lickert-score gegeven, waarbij het getal '1' stond voor een 'negatieve' verbale reactie en het getal '5' voor een 'positieve' verbale reactie. De mate van overeenstemming tussen de twee beoordelaars, is gemeten aan de hand van Cohen's Kappa en beoordeeld als voldoende ( $\kappa = .60, p < .001$ ).

### **Procedure**

De populatie is via twee VO scholen benaderd, vervolgens geïnformeerd over het



onderzoek en uiteindelijk gevraagd tot deelname. Ouders van leerlingen zijn doormiddel van een brief geïnformeerd over het onderzoek. Het instrument wetenschappelijk redeneren is klassikaal afgenomen door de testleider, dit nam ongeveer 30 minuten in beslag. Leerlingen hebben het instrument individueel en in stilte ingevuld. Alle leerlingen hadden dezelfde materialen ter beschikking, namelijk het antwoordformulier, een pen, een potlood en een gum. Er werd gestart met een gestandaardiseerde mondelinge instructie van de testleider, waarna de afname begon en het instrument op video werd aangeboden. Antwoorden op de items werden individueel genoteerd op het daarvoor bestemde antwoordformulier.

Afname van de SDDRQ vond plaats in groepen van vijf participanten die met de testleider aan een tafel gingen zitten. De testleider plaatste twee kopjes op de tafel, welke beiden voor de helft waren gevuld met lauw water. De testleider liet één van de participanten kopje één aanvullen met een theelepel suiker en een eetlepel gist (gist werd benoemd als een ‘geheim ingrediënt’). Kopje twee liet hij aanvullen met alleen een theelepel gist. Er werd geroerd in beide glazen. Aan de participanten werd vervolgens gevraagd om onderdeel één en twee van het antwoordformat in te vullen (zie bijlage 2). Vervolgens zette de testleider de twee kopjes weg. Hij pakte twee andere kopjes met dezelfde inhoud als de zojuist weggezette kopjes, alleen stonden deze kopjes al 45 minuten in de ruimte met een theedoek eroverheen. In die 45 minuten had er een gistingsproces plaatsgevonden in één van de glazen (degene met suiker én gist). Door het gistingsproces is er een soort schuim ontstaan en is de inhoud van het glas zichtbaar in volume toegenomen. De testleider verwijderde de theedoek en de participanten werd gevraagd om onderdeel drie en vier van het antwoordformat in te vullen (zie bijlage 2). Wanneer zij hier klaar mee waren, vroeg de testleider aan de participanten of zij konden uitleggen welke reactie er had plaatsgevonden en waarom dit slechts in één van de twee glazen was gebeurd. Als de participanten dit niet correct konden uitleggen, vulde de testleider aan.

### **Analyse**

Op basis van de literatuur waarmee het instrument wetenschappelijk redeneren is ontworpen was de verwachting een drie-factorenmodel te vinden, namelijk: (1) het opstellen van hypothesen, (2) ontwerpen van experimenten en (3) het voeren van observaties en evalueren van bewijs (Klahr et al., 2011; Kuhn et al., 1988; Zimmerman, 2007). Door middel van een confirmatieve factoranalyse is hierop gecontroleerd. De gevonden factoren zijn aan de hand van Cronbach’s Alfa onderzocht op interne consistentie.

Voorafgaand aan de verdere analyses is er allereerst middels ‘*compute variabele*’ een

variabele ‘wetenschappelijk redeneren’ geconstrueerd, bestaande uit de scores op de items van het zelfontworpen instrument wetenschappelijk redeneren. Voor het bereiken van de hoogst haalbare *Cronbach’s alpha* zijn de items 1D, 2C en 2A uit de variabele ‘wetenschappelijk redeneren’ verwijderd.

Om te onderzoeken of er een verschil is op scores op het instrument wetenschappelijk redeneren tussen jongens en meisjes is een t-toets uitgevoerd. Om te onderzoeken of er een verschil is op scores op het instrument wetenschappelijk redeneren tussen de verschillende onderwijsniveaus binnen het vmbo is een *one-way* ANOVA analyse uitgevoerd. Tot slot is er een multiële regressie analyse uitgevoerd om de relatie tussen enerzijds de score op het instrument wetenschappelijk redeneren (afhankelijke variabele) en anderzijds de leeftijd in maanden en nieuwsgierigheid naar wetenschap (onafhankelijke variabelen) te onderzoeken. Voor de analyses is aan alle assumpties voldaan (Baarda, de Goede & van Dijkum, 2007).

### Resultaten

Tabel 1

*Beschrijvende statistieken van de totaalscores op de toets wetenschappelijk redeneren op de variabelen geslacht en opleidingsniveau*

Variabele	Groep	N	M	SD
Geslacht	Jongens	41	4,88	1,25
	Meisjes	43	5,17	1,03
Opleidingsniveau	BBL	23	4,91	1,15
	KBL	35	5,15	1,15
	TL	26	5,02	1,16

*Noot: BBL = basisberoepsgerichte leerweg, KBL = kaderberoepsgerichte leerweg, TL = theoretische leerweg*

Als eerste werd middels confirmatieve principale componentenanalyse (PCA) van 3 factoren de factorstructuur op het instrument wetenschappelijk redeneren onderzocht. Factoren met eigenwaarden van  $\geq 1,0$  en factorladingen van  $\geq 0,45$  werden beschouwd als geschikt voor interpretatie. Gekozen is om een item toe te wijzen aan de factor waarop deze het hoogste laad, i.p.v. een item toe te wijzen aan meerdere factoren. Hiervoor is gekozen zodat er verderop in het onderzoek wanneer er wordt gewerkt met de totaalscore op de itemclusters, geen vertekening van de correlaties tussen clusters ontstaat.

Uit de *Bartlett’s sphericity test* is gebleken, dat de correlaties van de geselecteerde

items niet hoog genoeg zijn voor dimensie reductie door PCA ( $p > 0,05$ ). Bij de confirmatieve factoranalyse met 3 factoren, voldeden er drie items niet aan een factorlading van  $\geq .45$  of hoger. Vervolgens is er een exploratieve factoranalyse uitgevoerd en na rotatie bleek er sprake te zijn van een vierdimensionaal factorenmodel: vier factoren hebben een eigenwaarde boven de 1. Tevens vielen alle items onder deze factoren met een factorlading van  $\geq .45$ . De factoren kunnen worden genoemd (1) Ontwerpen van experimenten, (2) het opstellen van hypothesen (3) het voeren van observaties en evalueren van bewijs en (4) een restfactor. De cumulatieve verklaarde variantie betreft 59,65 procent (zie tabel 2). De interne betrouwbaarheid van alle vier de factoren, gemeten in *Cronbach's alpha* is echter onvoldoende. Er zijn namelijk vele redenen om een *Cronbach's alpha* lager dan .7 als onacceptabel te beschouwen (Field, 2013).

Tabel 2

*Resultaten van een factoranalyse op het instrument wetenschappelijk redeneren alsmede de interne betrouwbaarheid van de factoren weergegeven in Cronbach's alpha ( $\alpha$ ).*

Factoren	Bijbehorende items	Eigenwaarde	% Variantie	% Cumulatieve variantie	( $\alpha$ )
1	2D, 3B, 3D	1,45	16,14	16,14	.45
2	1B, 2B	1,35	14,98	31,12	.25
3	1A, 3A	1,30	14,45	45,57	.28
4	1C, 3C	1,27	14,07	59,64	.32

*Noot.* Zie bijlage 1 voor alle items van het instrument wetenschappelijk redeneren.  $\alpha$  = *Cronbach's alpha*

Uit de ANOVA met als afhankelijke variabele wetenschappelijk redeneren en als onafhankelijke variabele opleidingsniveau (BBL, KBL, TL), kwam in tegenstelling tot de verwachting geen significant verschil naar voren. Dit geeft aan dat de gemiddelde scores op het instrument wetenschappelijk redeneren niet significant verschillen op basis van het opleidingsniveau van de participant,  $F(2,81) = 0,35, p = .704, \text{partial } \eta^2 = .01$ .

Uit de t-toets met als afhankelijke variabele wetenschappelijk redeneren en als onafhankelijke variabele geslacht, kwam in overeenkomst met de verwachting geen significant verschil naar voren. Dit geeft aan dat de gemiddelde scores op het instrument wetenschappelijk redeneren niet significant van elkaar verschillen, op basis van het geslacht van de participant  $t = 1,09; df = 82; p > 0,05$  bij eenzijdige toetsing.

Wat betreft de samenhang tussen enerzijds scores op het instrument wetenschappelijk redeneren en anderzijds de nieuwsgierigheid naar wetenschap, is allereerst de Pearson's productmomentcorrelatiecoëfficiënt berekend. Omdat er een samenhang tussen de twee variabelen wordt verwacht is er 'one tailed' getoetst. Er blijkt sprake te zijn van een positief verband  $r = 0.43$ ;  $p = 0.01$ .

Vervolgens is er een multipele regressie analyse uitgevoerd. Allereerst is de score op het instrument wetenschappelijk redeneren voorspeld op basis van de leeftijd in maanden (X1). Dat leverde een correlatie op van  $r = -0.38$ ,  $p = < 0,05$ . Van de variantie van de score op het instrument wetenschappelijk redeneren kan 12 procent verklaard worden door de leeftijd in maanden (zie tabel 3). De conclusie is dat er een negatief zwak verband bestaat tussen scores op het instrument wetenschappelijk redeneren en de leeftijd in maanden van de participanten.

Daaropvolgend is de variabele nieuwsgierigheid naar wetenschap toegevoegd, aan de stapsgewijze multipele regressie analyse met de variabele leeftijd in maanden (X1+X2). Dat leverde een multipele correlatie op van  $r = 0.50$ ,  $p < 0,05$  (zie tabel 3). Van de scores op het instrument wetenschappelijk redeneren kan 25 % verklaard worden uit een combinatie van de variabelen leeftijd in maanden en nieuwsgierigheid naar wetenschap. De conclusie is dat er sprake is van een matige voorspellende waarde.

Tabel 3

*Resultaten meervoudige regressie analyses: leeftijd in maanden en nieuwsgierigheid naar wetenschap als voorspellers op wetenschappelijk redeneren*

Model	R	R <sup>2</sup>	Aangepaste R <sup>2</sup>	Std. meetfout	p
1	0,38	0,15	0,12	1,16	0,01**
2	0,50	0,25	0,21	1,10	0.03*

Noot: 1 = Voorspeller (constant), leeftijd in maanden; 2= Voorspeller (constant) Leeftijd in maanden, Nieuwsgierigheid naar wetenschap. \* $p < .05$ . \*\* $p < .01$

### Discussie en Conclusie

Dit onderzoek beschrijft de ontwikkeling van- en het onderzoek naar de betrouwbaarheid en validiteit van het zelfontworpen instrument wetenschappelijk redeneren. Dit is een instrument met als meetpretentie niveaubepaling van wetenschappelijk redeneren bij leerlingen in het vmbo leerjaar 1 en 2.

Op basis van de literatuur over het achterliggende construct wetenschappelijk

redeneren, is bij het ontwikkelen van het instrument als uitgangspunt gehanteerd om alle items onder te verdelen in drie factoren te weten: (1) het opstellen van hypothesen, (2) ontwerpen van experimenten en (3) het voeren van observaties en evalueren van bewijs. De eerste resultaten van het onderzoek naar de psychometrische kwaliteiten laten zien dat het instrument wetenschappelijk redeneren de drie beoogde factoren laat zien, alsmede een vierde factor. Alle vier de factoren zijn echter onvoldoende betrouwbaar, de items binnen de factoren correleren onderling te weinig. Wanneer alle items tezamen worden genomen om een totaalscore op het instrument wetenschappelijk redeneren te berekenen, vertoont deze schaal eveneens een onvoldoende betrouwbaarheid.

Een mogelijke verklaring voor de lage betrouwbaarheid, is het lage aantal items (2 tot 3) per factor. Daarentegen kan dit niet de hoofdoorzaak zijn, omdat er ook sprake is van een lage betrouwbaarheid op de schaal wetenschappelijk redeneren welke bestaat uit 9 items. Een daarom meer voor de hand liggende verklaring is dat de kwaliteit van de items van invloed is op de lage betrouwbaarheid van de individuele factoren, alsmede op de schaal wetenschappelijk redeneren. Geen van de items kon worden overgenomen uit een bestaand en betrouwbaar bewezen vragenlijst met eenzelfde meetpretentie. Mogelijk zit er daardoor te veel verschil in de moeilijkheid van de verschillende items binnen een factor en het instrument als geheel. Als de moeilijkheidsgraad te sterk verschilt kunnen de items nooit sterk correleren en kunnen ze (ten onrechte) laden op andere factoren. Dit kan de correlatie sterk beïnvloeden. Daarom wordt aangeraden om in vervolgonderzoek meer *trials* uit te voeren, om vervolgens op basis van betrouwbaarheidsanalyses de items aan te passen totdat de betrouwbaarheid tussen factoren en items voldoende is (Barry, Chaney, Piazza-Gardner & Chavarria, 2014).

In tegenstelling tot de hypothesen, zijn er geen verschillen gevonden wat betreft de score wetenschappelijk redeneren op de variabelen onderwijsniveau en leeftijd in maanden. De zojuist besproken beperking van de betrouwbaarheid is hiervoor een eerste verklaring. Een tweede mogelijke verklaring is dat het construct wetenschappelijk redeneren ‘te breed’ is meegenomen in het huidige onderzoek. In de literatuur (Zimmerman, 2000) is namelijk alleen gevonden dat opleidingsniveau en leeftijd van invloed zijn op de mate waarin een persoon in staat is om causale verbanden te kunnen leggen. Hoewel de testitems zo veel mogelijk zijn ontwikkeld en geformuleerd op basis van dergelijke literatuur over wetenschappelijk redeneren, was wellicht de hoeveelheid literatuur te beperkt om het beoogde theoretisch construct te meten. Daarom wordt aangeraden om in vervolgonderzoek

meer literatuur te gebruiken over het construct wetenschappelijk redeneren. Zodoende kan er sterker worden onderbouwd of het construct wetenschappelijk redeneren inderdaad uit een drie-factorenstructuur bestaat alsmede of er verschillen op de score wetenschappelijk redeneren worden veronderstelt op basis van de variabelen geslacht, opleidingsniveau en leeftijd.

Hoewel er voldoende aandachtspunten zijn voor vervolgonderzoek, zijn er ook een aantal sterke punten te benoemen van dit onderzoek. De manier waarop het instrument wordt aangeboden aan de doelgroep is vernieuwend, namelijk via Youtube. Dit medium sluit aan bij de beleving van de doelgroep, wekt hen interesse op en brengt hen op een aantrekkelijke manier in aanraking met het thema ‘wetenschap’. Het is van belang om de interesse van de doelgroep aan te wakkeren, aangezien de factor interesse van invloed is op de prestaties (Veerman, 2014). Onderzoekstechnisch gezien zijn sterke punten de grootte van de steekproef waarmee de analyses zijn uitgevoerd en de voldoende tot goede mate van interbeoordelaarsbetrouwbaarheid op de instrumenten. Daarnaast is er in lijn met de verwachting wel een significante samenhang gevonden tussen scores op het instrument wetenschappelijk redeneren en de nieuwsgierigheid naar wetenschap.

### **Conclusie**

Met het huidige onderzoek is getracht een eerste stap te zetten in de ontwikkeling van een betrouwbaar en valide instrument, met als meetpretentie het niveau van wetenschappelijk redeneren bij leerlingen in het vmbo leerjaar 1 en 2. De lage betrouwbaarheid van de factoren op zichzelf alsmede het instrument als geheel, suggereert in sterke mate dat het zelfontworpen instrument nog niet geschikt is om in te zetten voor niveaubepaling van wetenschappelijk redeneren. Toch is het huidige onderzoek een eerste stap geweest in de ontwikkeling van een dergelijk instrument, gezien het positieve resultaat van de correlatie analyse. Wanneer in vervolgonderzoek rekening wordt gehouden met de beperkingen van het huidige onderzoek, is de verwachting dat het ontwikkelen van een betrouwbaar en valide meetinstrument mogelijk is.

Literatuur

- Baarda, D.B., de Goede, M.P.M., & van Dijkum, C.J. (2007). Basisboek statistiek met SPSS. Noordhoff uitgevers Groningen: Houten.
- Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., ... & Wu, N. (2009). Learning and scientific reasoning. *Science*, 323, 586-587. doi:10.1126/science.1167740
- Baruch, Y.K., Spektor-Levy, O., & Mashal, N. (2014). Pre-schoolers verbal and behavioral responders as indicators of attitudes and scientific curiosity. *International Journal of Science and Mathematics Education*,
- Barry, A.E., Chaney, B.H., Piazza-Gardner, A.K., & Chavarria, E.A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education and Behavior*, 41, 12-18.  
doi:10.1177/1090198113483139
- Dewey, J. (1910). How we think (2009<sup>th</sup> ed.). Mineola: Dover Publications Inc.
- German, P.J. (1989). Directed-inquiry approach to learning science process skills: Treatment effects and aptitude-treatment interactions. *Journal of Research in Science Teaching*, 26 (3), 237-250.
- German, P.J., Aram, R., & Burke (1996). Identifying patterns and relationships among the responses of seventh-grade students to the science process skill of designing experiments. *Journal of Research in Scientific Thinking*, 33, 79-99.
- German, P.J., Aram, R., Odom, A.L., & Burke, G. (1996). Student performance on asking questions, identifying variables, and formulating hypotheses. *Journal of Science and Mathematics*, 96 (4), 192-201.
- Field, A. (2013). *Discovering Statistics Using SPSS* (4de ed.). London: SAGE.
- Keys, C.W. (1994). The development of scientific reasoning skills in conjunction with collaborative writing assignments: An interpretive study of six ninth-grade students. *Journal of Research in Science Teaching*, 31, 1003-1022. doi: 10.1002/tea.3660310912
- Klahr, D., & Dunbar, K. (1988). Dual search space during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Klahr, D., Zimmerman, C., & Jirout, J. (2011). Educational interventions to advance children's scientific thinking. *Science*, 333, 971-975. doi:10.1126/science.1204528
- Komarraju, M., Karau, S.J., Schmeck, R.R., & Avdic, A. (2011). The big five personality

- traits, learning styles, and academic achievement. *Personality and Individual Differences*, 51 (4). doi: 10.1016/j.paid.2011.04.019
- Krapp, A., & Prenzel, M. (2011). Research on interest in science: Theories, methods, and findings. *International Journal of Science Education*, 33, 27–50. doi: 10.1080/09500693.2010.518645
- Kuhn, D., Amsel, E., & O’Loughlin, M. (1988). The development of scientific thinking. New York: Harcourt, Brace, Jovanovich.
- Schauble, L., Glaser, R., Duschl, R. A., & Schulze, S. (1995). Students’ understanding of the objectives and procedures of experimentation in the science classroom. *Journal of the Learning Sciences*, 4, 131–166. doi:10.1207/s15327809jls0402\_1
- Silvia, P. J. (2005). What is interesting? Exploring the appraisal structure of interest. *Emotion*, 5, 89–102. doi:10.1037/1528-3542.5.1.89
- Silvia, P. J. (2008). Interest—the curious emotion. *Current Directions in Psychological Science*, 17, 57–60. doi:10.1111/j.1467-8721.2008.00548.x
- Spektor-Levy, O., Kesner Baruch, Y., & Mevarech, Z. (2013). Science and scientific curiosity in pre-school—the teacher’s point of view. *International Journal of Science Education*, 35, 2226–2253. doi:10.1080/09500693.2011.631608
- Spelke, A.S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, 60 (9), 950-958.
- Tamdogen, O. G. (2006). Creativity in education: Clearness in perception, vigorousness in curiosity. *Education for information*, 24, 139-151.
- Veenman, M.V.J. (2005) Faalangst, een dobbelsteen met zes zijden. *Remediaal*, 4 (5), 3-9.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20, 99-149. doi:10.1006/drev.1999.0497
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172-223. doi:10.1016/j.dr.2006.12.001



**Bijlage 1: Instrument wetenschappelijk redeneren**

**Instrument Wetenschappelijk redeneren**

Zo meteen krijg je een filmpje te zien. Bij dit filmpje horen een aantal vragen. Na elke vraag krijg je even tijd om hier over na te denken en het antwoord op te schrijven. Je zal merken dat sommige vragen makkelijker zijn en sommige vragen wat moeilijker. Schrijf op wat jij het beste antwoord vindt, fouten maken is hierbij niet erg. De uitkomst van de test telt niet mee voor je rapport. Als je jouw antwoord hebt opgeschreven wacht je rustig en in stilte op de volgende vraag.

Vul zo meteen eerst je naam, leeftijd en klas, en de taal die je thuis spreekt in.

,

Succes!

Naam: ..... Klas: .....

Leeftijd: ..... Taal die we thuis spreken: .....

**Vraag 1a:**

Beschrijf in stappen hoe de professor met deze materialen een raket kan bouwen die de lucht in gaat.



pomp



ventiel



kurk



vinnen



fles



plakband

- Stap 1: .....
- Stap 2: .....
- Stap 3: .....
- Stap 4: .....
- Stap 5: .....
- Stap 6: .....
- Stap 7: .....
- Stap 8: .....
- Stap 9: .....
- Stap 10: .....

- Ja, ik weet al hoe ik een raket moet bouwen
- Nee, ik heb nooit eerder gehoord hoe ik een raket kan bouwen

**Vraag 1b**

Wat denk jij dat er gaat gebeuren?

.....

.....

.....

**Vraag 1C**

Verklaar hoe het komt dat de raket de lucht in schiet.

Zet de 5 stappen in de juiste volgorde. *Zet het juiste nummer in elk vakje.*

- De kurk schiet uit de fles
- Lucht komt in de fles
- De fles gaat de lucht in
- Door het pompen wordt lucht verplaatst
- De luchtdruk in de fles bouwt zich op

**Vraag 1D**

Ging het zoals jij verwacht had? *Kruis aan. ('nee' vul dan nu de juiste volgorde in)*

Ja, het ging precies zoals ik bij de vorige vraag had opgeschreven  
Nee, ik zag dat het in een andere volgorde gebeurde. Vul hier de goede volgorde in:

- De kurk schiet uit de fles
- Lucht komt in de fles
- De fles gaat de lucht in
- Door het pompen wordt lucht verplaatst
- De luchtdruk in de fles bouwt zich op

**Vraag 2A**

Kruis de **twee** juiste kenmerken aan, die de professor in dit experiment kan onderzoeken:

- Snelheid van de raket
- Grootte van de raket
- Tijd dat de raket in de lucht blijft
- Hoogte van de raket
- Het aantal vinnen op de raket

**Vraag 2B**

Welk kenmerk kan de professor veranderen? Vul in.

Als de professor ..... verandert, verandert .....(het andere kenmerk) ook.

**Vraag 2C**

Welke raket blijft langer in de lucht?  
*Zet een rondje op het goede antwoord:*

de kleine raket      *of*      de grote raket

**Vraag 2D**

Leg uit waarom de professor denkt dat de grote raket langer in de lucht blijft.

.....

.....

.....

.....

.....

**Vraag 3A**

Om te onderzoeken of de grote raket langer in de lucht blijft heeft de professor de volgende onderdelen nodig, kruis aan:

- |                          |             |                          |              |                          |           |
|--------------------------|-------------|--------------------------|--------------|--------------------------|-----------|
| <input type="checkbox"/> | Meetlint    | <input type="checkbox"/> | Stopwatch    | <input type="checkbox"/> | Bomen     |
| <input type="checkbox"/> | Grote raket | <input type="checkbox"/> | Camera       | <input type="checkbox"/> | Fietspomp |
| <input type="checkbox"/> | Gras        | <input type="checkbox"/> | Kleine raket | <input type="checkbox"/> | Papier    |
| <input type="checkbox"/> | Eierwekker  | <input type="checkbox"/> | Pen          | <input type="checkbox"/> | Computer  |

**Vraag 3B**

Schrijf alle losse materialen op die de professor nodig heeft om het experiment uit te voeren.

.....

.....

.....

.....

.....

.....

**Vraag 3C**

Welke informatie uit het filmpje is belangrijk om te bepalen of de grote of de kleine raket langer in de lucht blijft? Kies het beste antwoord.

- Het aantal meters dat de raketten in de lucht zijn
- Het aantal slagen dat de professor heeft gepompt voordat beide raketten de lucht in gingen
- De lengte van de punt van beide raketten in centimeters
- De afstand die de raket heeft afgelegd vanaf de lanceerplek (in meters)
- De tijd in seconden dat de raketten in de lucht zijn geweest

**Vraag 3D**

Schrijf een zo volledig mogelijke conclusie op.

.....

.....

.....

.....

**Bijlage 2: Instrument nieuwsgierigheid naar wetenschap**

**Het uitvoeren van een wetenschappelijk experiment**

Naam leerling:.....

Klas leerling:..... Datum.....

**Onderdeel 1: Vul de volgende zinnen aan:**

- 1. Het materiaal ziet eruit als.....
- 2. Het materiaal ruikt als.....
- 3. Het materiaal voelt als.....

**Onderdeel 2: Geef antwoord op de volgende drie vragen:**

- 1. Ben je nieuwsgierig naar wat er gaat gebeuren? Ja / nee (omcirkel jouw antwoord)
- 2. Wat denk je dat er gaat gebeuren?

.....  
.....  
.....

- 3. Leg uit waarom je dat denkt:

.....  
.....  
.....  
.....

**Onderdeel 3: Vul de volgende zinnen aan:**

- 1. Het materiaal ziet eruit als.....
- 2. Het materiaal ruikt als.....
- 3. Het materiaal voelt als.....

**Onderdeel 4: Vul de volgende zinnen aan:**

- 1. Kun je vertellen wat er is gebeurd in het experiment?

.....  
.....  
.....

.....

2. Welke gevoelens kreeg je tijdens het doen van het experiment?

.....  
.....  
.....  
.....

3. Was je nieuwsgierig naar wat er zou gebeuren tijdens het experiment? Ja of nee? (omcirkel jouw antwoord)

3a. Als je 'Ja' hebt ingevuld: Waar was je precies nieuwsgierig naar?

.....  
.....  
.....

4. Wil je meer van dit soort wetenschappelijke experimenten zien? Ja of nee (omcirkel jouw antwoord)

4a. Als je 'Ja' hebt ingevuld: Kun je een voorbeeld geven van een wetenschappelijk experiment dat je zou willen zien?

.....  
.....  
.....  
.....

5. Wil je graag weten wat er in dit experiment is gebeurd? Ja of nee? (omcirkel jouw antwoord).