

CONSTRUCTIE EN VALIDATIE VAN EEN TEST VOOR HET METEN VAN INQUIRY SKILLS

Martine Horstink

Samenvatting

Dit artikel beschrijft de ontwikkeling van de Inquiry Skills Test (IST), een instrument voor het meten van vaardigheden in het ontdekkend leren (inquiry skills). De IST bestaat uit de items uit The Test of Process Skills II (TIPSII) en de onderdelen Conclusie en Interpretatie uit de Watson-Glaser Kritisch Denken Test (WGKDT). De betrouwbaarheid en validiteit van de IST is onderzocht bij 25 Nederlandse leerlingen uit 4 VWO. Bij deze leerlingen werd eerst klassikaal een testbatterij afgenomen die een beroep doet op de cognitieve capaciteiten. In een counterbalanced design startten de leerlingen vervolgens met de IST en een inquirytaak. Er werden geen significante volgorde-effecten gevonden. De interne betrouwbaarheid van de totale IST bleek voldoende te zijn, maar de subschalen bleken in dit specifieke onderzoek niet intern valide. Scores op de IST en de cognitieve capaciteitentests vertoonden beide een hoge significante correlatie met de prestaties op de inquirytaak. Een regressieanalyse met de scores op de IST en de cognitieve capaciteitentest als voorspeller en prestaties op de inquirytaak als criterium liet zien dat de IST de enige significante voorspeller voor de prestaties op de inquirytaak was. De cognitieve capaciteitentest bleek in deze regressieanalyse geen significante voorspeller te zijn. Met de constructie van de IST lijkt een instrument ontwikkeld te zijn dat in een behoefte van de Nederlandse populatie kan voorzien. Verder onderzoek is echter noodzakelijk om het instrument te verbeteren en conclusies te kunnen generaliseren. In een vervolgonderzoek zouden meer respondenten, verschillende leeftijdsgroepen en een andere inquirytaak opgenomen kunnen worden.

Inleiding¹

Kessels en Poell (2001) signaleerden dat het constructivisme de afgelopen tien jaar een belangrijke stroming is geworden in de leer-, onderwijs- en opleidingspsychologie. Ook Wells (1999) constateert dat met name binnen de wetenschappelijke en wiskundige disciplines een groeiende erkenning is ontstaan voor de constructivistische visie op leren, waarin de lerende zelf kennis verzamelt en construeert. De grote belangstelling voor het constructivisme heeft tegelijkertijd geleid tot nieuwe aandacht voor de ‘inquiry-benadering’ die door John Dewey reeds een eeuw geleden werd bepleit (Wells, 1999). Dewey introduceerde in zijn ‘theory of inquiry’ het idee dat leerlingen alleen kennis kunnen vergaren door zichzelf in situaties te brengen waarin de kennis opgedaan zou kunnen worden (Dewey, 1938/1991). Bij ‘inquiry learning’² zouden leerlingen kennis moeten vinden of ontdekken in spontane situaties, gebaseerd op eigen probleemervaringen (Franken, 2003).

Volgens Kuhn (1990) heeft het onderwijs de laatste vijftig jaar ook veel interesse getoond in het ontwikkelen van nieuwe curricula en het verbeteren van bestaande curricula om de ontwikkeling van (denk)vaardigheden te stimuleren. Het aanleren van vaardigheden werd belangrijker dan het verwerven van kennis. Stokking en Van der Schaaf (1999) meldden dat sinds eind 90-er jaren binnen de kerndoelen en eindtermen in het voortgezet onderwijs in Nederland expliciet, algemene (vakoverstijgende) vaardigheden zijn opgenomen. De aandacht voor informatieverwerkings- en onderzoeksvaardigheden is sterk toegenomen. De onderzoeksvaardigheden zijn sterk verwant met zogenoemde ‘inquiry skills’.

Leerlingen uit de bovenbouw van het havo en vwo worden voorbereid op het hoger onderwijs, waarin zij leren zelfstandig wetenschappelijk te redeneren en onderzoek te doen. De mate waarin deze leerlingen beschikken over ‘inquiry skills’ is bepalend voor hun prestaties binnen het hoger onderwijs. Meer en meer moeten docenten en leidinggevenden vaardigheden, zoals inquiry skills, van leerlingen of medewerkers beoordelen. Met name binnen het onderwijs lijkt een test voor het meten van de inquiry skills in een behoefte te kunnen voorzien. Met de test zou de effectiviteit van op inquiry learning gebaseerd onderwijs gemeten kunnen worden en kan men opleidings- en bijscholingsadviezen beter onderbouwen.

In dit onderzoek wordt allereerst nagegaan hoe inquiry learning en inquiry skills te definiëren zijn. Op basis van de vastgestelde definities wordt een instrument samengesteld voor het meten van de inquiry skills. De betrouwbaarheid en validiteit van dit instrument wordt onderzocht bij leerlingen uit 4 VWO.

Definiëren van inquiry learning

Inquiry learning vindt zijn oorsprong in het werk van Dewey. In ‘How We Think’ (1933/1989) presenteert Dewey zijn inquiry theorie voor het onderwijskundige vakgebied. Volgens Dewey bestaat het inquiry process uit vijf stappen, die niet chronologisch of los van elkaar hoeven plaats te vinden. De eerste stap betreft het zich voordoen van een probleem. Alle inquiry processen zouden beginnen met een interne twijfel: een gevoel van verstoring in het doen en laten en niet weten hoe verder te gaan. In stap twee volgt de specificatie van het probleem. Hiertoe worden data geselecteerd (data worden nooit ‘gegeven’ volgens Dewey) en gestructureerd: mogelijke relevante factoren worden in kaart gebracht. De derde stap bevat het introduceren van een veronderstelling of een hypothese als oplossing voor het probleem. Het construeren van hypothesen vereist creatief gebruik van het voorstellingsvermogen om mogelijke oplossingen te ontwikkelen. In de vierde stap kijkt de lerende naar de consequenties van de hypothese: er wordt een vergelijking met andere hypothesen gemaakt en metingen worden verricht. Na het redeneren is het tijd om de gedachten experimenteel te toetsen: stap 5. De lerende onderneemt acties om de condities uit de hypothesen te creëren om te zien of de veronderstelling inderdaad klopt en of de bedoelde consequenties plaats vinden.

De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) beschreef in 2003 onderzoekend leren als onderwijs waarbij, vaak naar aanleiding van een overkoepelende vraag

¹ Met dank aan dr. P. Wilhelm en dr. A.W. Lazonder voor de begeleiding van dit onderzoek.

² Omdat de term ‘inquiry learning’ niet eenduidig te vertalen is naar het Nederlands, wordt de Engelse term in dit artikel gehanteerd. Dit geldt tevens voor de termen ‘inquiry skills’, ‘inquirytaak’ en ‘inquiry proces’.

(bijvoorbeeld “hoe werkt het broeikaseffect?”), op een ‘experimenterende’ manier gezocht wordt naar antwoorden op algemene en/of specifieke deelvragen. Deze visie komt niet geheel overeen met de benadering van Dewey en Wells. Het verschil wordt bepaald doordat NWO het leren als doel stelt en derhalve ook overkoepelende vragen in het leven roept. Dewey en Wells doen dat bewust niet. NWO stelt wel dat er bij onderzoekend leren geen vastgestelde leerroute is en zelfs niet altijd een precies vastgesteld leerdoel of leeruitkomst (NWO, 2003). Door de inquiry benadering te willen inpassen in het onderwijs in Nederland, laat NWO één aspect van inquiry learning (namelijk dat het initiatief om te leren volledig bij de lerende ligt) verdwijnen.

Bonnstetter (1998) onderscheidt binnen het onderwijs verschillende niveaus van inquiry learning: (1) traditioneel praktijkonderwijs, (2) gestructureerde experimenten, (3) begeleid inquiry learning, (4), door de student gestuurd inquiry learning, en (5) zelfstandig onderzoekende studenten. Op het eerste niveau vindt volgens Bonnstetter geen inquiry learning plaats, omdat de docent alle beslissingen en stappen neemt. De rol van de docent neemt af, naarmate het niveau van inquiry learning toeneemt. NWO doelt met name op het derde niveau van Bonnstetter (‘guided inquiry’), waarbij de docent het onderwerp en de vragen selecteert en de materialen levert. De studenten zijn zelf verantwoordelijk voor het opstellen van een onderzoek. Wells (en ook Dewey) is juist een aanhanger van het ultieme inquiry learning (‘student research’), niveau 5 van Bonnstetter, met het volledige initiatief bij de lerende. Op dit niveau geven docenten slechts ondersteuning en begeleiding en bedenken leerlingen zelf met welk onderwerp zij aan de slag gaan.

Ook Chinn en Malhotra (2002) menen dat er verschillende vormen of niveaus van inquiry learning bestaan. Zij benadrukken echter het verschil tussen ‘authentieke wetenschappelijke inquiry’ (het onderzoek dat professionele wetenschappers uitvoeren, vergelijkbaar met het vijfde niveau van Bonnstetter) en ‘eenvoudige inquiry taken’. De eenvoudige inquiry taken komen vaak binnen het onderwijs voor. Hierbij worden de onderzoeksvraag, de variabelen en de manier van experimenteren of observeren precies voorgeschreven. Het authentieke wetenschappelijk inquiry proces is een complexe activiteit die dure apparatuur, gespecialiseerde kennis en geavanceerde technieken voor data-analyse en modellering vereist. Volgens Chinn en Malhotra moet getracht worden om relatief eenvoudige ‘school inquiry taken’ te ontwikkelen die binnen de benodigde tijd, ruimte, expertise en het budget in een klaslokaal kunnen worden uitgevoerd, maar die ondanks hun eenvoud wel de belangrijkste componenten van wetenschappelijk redeneren bevatten.

Chinn en Malhotra (2002) onderscheiden vier niveaus van onderzoekstaken met cognitieve processen die bij inquiry learning voorkomen:

1. eenvoudige illustraties: studenten volgen een voorgeschreven procedure middels aanwijzingen, zij verrichten een meting en zij observeren de uitkomst. Het experiment illustreert een theoretisch principe dat in de tekst wordt uitgelegd. Hierbij gaat het om eenvoudig deductief redeneren.
2. eenvoudige observaties: studenten observeren en beschrijven aangewezen objecten volgens aanwijzingen. Zij leren eenvoudig inductief redeneren.
3. eenvoudige experimenten: studenten krijgen een onderzoeksvraag en één of twee variabelen, met één controlegroep, zij voeren één experiment uit volgens een procedure en zij generaliseren naar vergelijkbare situaties. Het gaat hierbij om eenvoudig contrastief redeneren.
4. authentieke onderzoeken: wetenschappers bedenken hun eigen onderzoeksvragen, selecteren verschillende variabelen, gebruiken meerdere controlegroepen, herhalen observaties, beoordelen generalisatiemogelijkheden, construeren theorieën en lezen andere onderzoeksrapporten.

Net als bij de niveaus van Bonnstetter neemt ook bij deze indeling de complexiteit van het inquiry learning toe met het toenemen van de niveaus. Het strikte onderscheid van Chinn en Malhotra tussen de eenvoudige inquiry taken (niveaus 1 tot en met 3) en de authentieke inquiry taken (niveau 4) verloopt bij Bonnstetter vloeiender. De overgang tussen niveau 4 en 5 bij Bonnstetter lijkt kleiner dan de overgang tussen niveau 3 en 4 bij Chinn en Malhotra. Een ander verschil is dat Bonnstetter blijft spreken over zelfstandig onderzoekende studenten, waar Chinn en Malhotra op hun hoogste niveau nadrukkelijk spreken over wetenschappers.

De auteurs lijken met hun definities van inquiry learning allemaal te doelen op hetzelfde concept: het ultieme inquiry learning zoals dat door Dewey werd beschreven. Enkele auteurs onderscheiden hierbinnen verschillende niveaus op basis van de mate van zelfstandigheid en de complexiteit van de taken.

In de literatuur wordt inquiry learning vaak in één adem genoemd met kritisch denken. Schafersman (1991) meent zelfs dat beide termen gelijk zijn: *“Critical thinking is also critical inquiry, because critical thinkers investigate problems, ask questions, pose new answers that challenge the status quo, discover new information that can be used for good or ill, question authorities and traditional beliefs, challenge received dogmas and doctrines, and often end up possessing power in society greater than their numbers”* (p. 3).

In een reactie op het definiëren van kritisch denken in termen van Bloom's taxonomie voor het opstellen van leerdoelen definieert Ennis (1993) kritisch denken als “...reasonable reflective thinking focused on deciding what to believe or do” (p. 180). Hij heeft deze definitie verder geoperationaliseerd door een tiental activiteiten te beschrijven die een persoon zou moeten uitvoeren bij het kritisch denken: beoordeel de geloofwaardigheid van bronnen; identificeer conclusies, redenen en veronderstellingen; beoordeel de kwaliteit van een argument; ontwikkel en verdedig een mening over zaken; stel geschikte verhelderende vragen; plan experimenten en beoordeel experimentele ontwerpen; definieer termen op een manier die past bij de context; wees ontvankelijk; probeer goed geïnformeerd te zijn; trek conclusies wanneer dat gerechtvaardigd is. Deze activiteiten komen grotendeels overeen met houdingen of activiteiten binnen de inquiry benadering. Het stellen van vragen, het ontvankelijk zijn voor verrassingen of ontdekkingen, het trekken van conclusies, het vormen van een mening; al deze aspecten komen ook terug in de inquiry benadering. Een verschil is dat Ennis bij zijn kritisch denken de nadruk legt op het bewust aannemen van een bepaalde houding. Kenmerkend voor inquiry learning volgens Dewey is echter dat dit spontaan plaats vindt, zonder de bewuste houding aan te nemen.

Twee andere bekende definities van kritisch denken zijn McPeck's “reflectieve scepticisme” en Paul's “strong sense” definitie (Ennis, 1993). Kritisch denken is volgens McPeck (1981) “de neiging en vaardigheid om zich bezig te houden met activiteiten met reflectief scepticisme. Deze definitie komt volgens Ennis (1993) erg negatief over. Paul (1987) definieert kritisch denken als: *“the art of thinking about your thinking while you are thinking in order to make your thinking better: more clear, more accurate, or more defensible.”* Hierbij maakt hij onderscheid tussen kritisch denken in de ‘weak sense’ en de ‘strong sense’. In het eerste geval gaat het om denken dat uitgaat van het eigen standpunt (egocentrisme, etnocentrisme) en alleen maar tot doel heeft om het eigen standpunt te behouden. In het laatste geval gaat het om denken dat ook openstaat voor andere gezichtspunten en na alles tegen elkaar afgewogen te hebben tot een weloverwogen conclusie leidt. De definitie van Paul komt overeen met de definitie van Ennis, maar legt een nog zwaardere nadruk op het bewustzijn van iemands eigen veronderstellingen en het bekijken vanuit andere oogpunten. De onderzoeksvaardigheden staan minder centraal. Hierdoor staat deze visie weer verder af van de inquiry benadering.

De operationalisering van Ennis van kritisch denken toont een groot aantal overlappende activiteiten met eerdergenoemde beschrijvingen van inquiry learning. Of inquiry learning samenhangt met kritisch denken hangt vooral af van de gehanteerde definities. Hoewel er geen duidelijke verschillen tussen kritisch denken en inquiry learning aan te wijzen zijn, lijkt het alsof de diverse auteurs er vanuit gaan dat kritisch denken bewuster plaats vindt. De door Ennis (1993) beschreven ‘houding’ en acties passen in het beeld van inquiry learning zoals geschetst door bijvoorbeeld Wells (1999). Methoden van kritisch denken vertonen veel overeenkomsten met methoden van wetenschappelijk redeneren. Hiermee hangt samen dat het ultieme inquiry learning volgens Bonnstetter (1998) en volgens Chinn en Malhotra (2002) uitgevoerd wordt door geheel zelfstandige onderzoekende studenten of wetenschappers, die daarbij wetenschappelijke redeneermethoden hanteren.

In literatuur over inquiry learning en kritisch denken wordt vaak gesproken over ‘een wetenschappelijke manier van onderzoeken’. Professionele wetenschappers zouden het ultieme inquiry proces doorlopen (Chinn & Malhotra, 2002). Volgens Schafersman (1991) is kritisch denken geheel gelijk aan wetenschappelijk redeneren. Hij baseert dit op het feit dat veel boeken over kritisch denken doelen en methoden beschrijven die gelijk zijn aan de doelen en methoden van de wetenschap. Ook Kuhn (1993)

schrijft dat ‘scientific reasoning’ een term is die veel te maken heeft met kritisch denken. Met inquiry learning en scientific reasoning wordt waarschijnlijk in veel gevallen hetzelfde bedoeld. Beide termen doelen op een kritische en onderzoekende houding met als doel het vinden van nieuwe informatie of het trekken van conclusies. Een verschil lijkt te zitten in de gestructureerde manier van denken. Bij wetenschappelijk redeneren wordt verwacht dat een voorgeschreven methode gehanteerd wordt, bij inquiry learning lijkt dit meer open gelaten te worden.

Gerelateerd aan inquiry learning is ook ‘scientific discovery learning’. Dit is een typische vorm van constructivistisch leren, gebaseerd op probleemoplossende activiteiten, waarbij het ontwerpen en implementeren van wetenschappelijke experimenten een rol spelen (Reid, Zhang, & Chen, 2003). Volgens De Jong en Van Joolingen (1998) is de kern van scientific discovery learning dat de leerling zelf zijn leren stuurt, opbouwt en bepaalt in een situatie waarin hij zelf vaststelt wat en hoe geleerd kan worden. Hierbij kunnen docenten aanmoedigen tot het formuleren van vragen, het plannen van activiteiten en het trekken en beoordelen van conclusies. Een cruciaal aspect van scientific discovery learning is het ontwerpen van experimenten die informatie verschaffen voor het beslissen over de validiteit van een hypothese (De Jong & Van Joolingen, 1998). Een verschil tussen scientific discovery learning en inquiry learning is dat bij scientific discovery learning specifiek over leerlingen en docenten gesproken wordt. Het leren wordt gezien als een doel en problemen worden aangedragen. Bij authentiek inquiry learning is dat niet het geval. Wanneer gekeken wordt naar het toepassen van inquiry learning binnen het onderwijs (vaak op niveau 4 van Bonnstetter), dan lijken beide vormen van leren slechts overeenkomsten te vertonen.

Ook tussen definities van (scientific) discovery learning en inquiry learning zijn geen duidelijke verschillen. Beide termen lijken elkaar daardoor te overlappen, waarbij scientific discovery learning wellicht meer gericht is op de natuurwetenschappen en inquiry learning meer op het leren in het algemeen. De samenhang tussen beide termen is niet verrassend, aangezien de termen discovery learning en inquiry learning in het Nederlands beiden vertaald (kunnen) worden met ontdekkend/onderzoekend leren.

Samengevat kan gesteld worden dat slechts het hoogste niveau van inquiry learning, gedefinieerd door verschillende auteurs (Bonnstetter, Chinn & Malhotra) overeenkomt met inquiry learning zoals gedefinieerd door Dewey. Het aspect ‘verwondering’; het zelf bepalen wat men wil leren, lijkt vaak losgelaten te worden. In grote lijnen komen inquiry learning, kritisch denken, wetenschappelijk redeneren en scientific discovery learning met elkaar overeen. Het opstellen van hypothesen, het verrichten van experimenten en het trekken van conclusies komt in elke benadering voor.

Omdat dit onderzoek uitgevoerd wordt binnen een onderwijssetting zal hier niet gewerkt worden met het ultieme inquiry learning, omdat dit niet binnen het onderwijs wordt toegepast. De definitie van inquiry learning die wel gehanteerd wordt, is gebaseerd op het vierde niveau van Bonnstetter (1998): binnen het onderwijs op dit niveau stellen docenten het onderwerp vast en de hoofdvragen op, waarna de leerlingen geheel vrij gelaten worden wat betreft het opstellen van hypothesen, het verwerven van informatie, het ontwerpen van experimenten etc. Het opdoen van ervaringen (en het hierdoor uitbreiden van de kennis) wordt gezien als het doel van inquiry learning (zie ook Wells, 1999).

Definiëren van inquiry skills

Hogan en Fisherkeller (2000) benoemen vier fasen ofwel onderzoeksstappen voor het wetenschappelijke inquiry proces: hypothesen vormen, experimenten ontwerpen, uitkomsten interpreteren en de resultaten communiceren. Indien ‘vaardigheden’ worden opvat als de vereiste cognitieve bekwaamheden voor het inquiry process, dan kunnen onderzoeksstappen niet worden gelijkgeschakeld aan vaardigheden. Verschillende onderzoeksstappen kunnen immers (deels) dezelfde cognitieve activiteiten veronderstellen en voor eenzelfde stap kunnen meerdere cognitieve activiteiten nodig zijn (Stokking & Van der Schaaf, 1999). Voor het ontwikkelen van een test die vaardigheden gaat meten, lijkt het van belang om de achterliggende cognitieve activiteiten te ontrafelen. Bij het beoordelen van medewerkers of leerlingen zijn echter vaak slechts de genomen stappen zichtbaar. Het in kaart brengen van de achterliggende cognitieve vaardigheden is dan helaas vaak niet (meer) mogelijk. Als men ‘vaardigheden’ opvat als betekenisvolle taken, in de context van een groter geheel, dan kunnen de onderzoeksstappen wel gezien worden als de vaardigheden die moeten worden geleerd om een onderzoek

te kunnen opzetten en uitvoeren (Stokking & Van der Schaaf, 1999). Voor het vervolg van dit onderzoek wordt uitgegaan van deze opvatting.

Volgens Stokking en Van der Schaaf (1999) betreft het complexe domein ‘onderzoeksvaardigheden’ zowel kennis als vaardigheden en bestrijkt het een aantal componenten. Deze auteurs noemen vakinhoudelijke kennis, kennis over en inzicht in onderzoek, diverse deelvaardigheden (waaronder het formuleren van een onderzoekbare vraagstelling, opstellen van een werkplan en het rapporteren over het onderzoek), de vaardigheid in het bewaken van de voortgang van een onderzoek en de vaardigheid om achteraf te reflecteren op het uitgevoerde onderzoek.

Volgens Stokking en Van der Schaaf (1999) ontwikkelen vaardigheden zich deels op basis van kennis en vergt het flexibel benutten van vaardigheden eveneens kennis. Het aanleren van het opzetten en uitvoeren van een onderzoek, kan volgens hen niet zonder aandacht voor ‘kennis’ over onderzoek doen. Dit zou betekenen dat inquiry skills niet alleen uit vaardigheden bestaan, maar ook voor een gedeelte uit kennis. De opgedane kennis afgeleid van het trachten te vinden van een oplossing is contextgebonden en alleen bruikbaar in de specifieke situatie (Franken, 2003). Volgens deze auteurs zijn inquiry skills dus niet volledig domeinonafhankelijk. Zij geven aan dat een gedeelte van de onderzoeksvaardigheden ontwikkeld wordt door het verwerven domeinafhankelijke kennis. Volledig domeinafhankelijke vaardigheden zijn niet middels algemene testen te meten. Omdat Stokking en van der Schaaf ook andere (domeinonafhankelijke) deelvaardigheden noemen, betekent deze conclusie nog niet dat het meten van inquiry skills niet mogelijk zou zijn. Dit onderzoek richt zich derhalve op de domeinonafhankelijke aspecten van inquiry learning, met het besef dat er ook domeinafhankelijke componenten zullen zijn.

Hallonen, Bosack, Clay en McCarthy (2003) onderscheiden acht domeinen met specifieke inquiry skills, die zij rangschikken in volgorde van plaatsvinden of belangrijkheid:

1. Beschrijvende vaardigheden: observatie, interpretatie en meetvaardigheden.
2. Conceptuele vaardigheden: herkennen en toepassen van concepten en theorieën, evaluatie, synthese en generalisatievaardigheden.
3. Probleemoplossende vaardigheden: herkennen, evalueren en generaliseren van onderzoeksmethoden, afwijkingen ontdekken, statistisch redeneren.
4. Ethische redeneervaardigheden: bewustheid, evalueren en aanhangen van ethische standaarden.
5. Wetenschappelijke waarden en houding: enthousiast voor onderzoek, objectiviteit/subjectiviteit, wantrouwend, sceptisch, tolerant voor dubbelzinnigheid.
6. Communicatieve vaardigheden: argumenteervaardigheden, informatieverwervende vaardigheden, conventionele uitdrukkingsvaardigheden.
7. Samenwerkingsvaardigheden: procesmanagement, vaardigheden in projectmatig werken, leiderschap, brainstormen, streven naar overeenstemming.
8. Vermogen om zichzelf te beoordelen: zelf-regulatie, reflectie op zichzelf.

Hiermee kiezen de auteurs voor een breed perspectief. De acht domeinen vertonen veel overeenkomsten met de door Stokking en Van der Schaaf (1999) genoemde componenten. Een verschil met Hogan en Fisherkeller (2000) is dat zij juist meer spreken over onderzoeksstappen in plaats van over vaardigheden. Deze onderzoeksstappen zijn overigens wel terug te vinden in beschrijvingen van de domeinen van Hallonen et al.

‘Science process skills’, genoemd door Burns, Okey en Wise (1985) vertegenwoordigen de vaardigheden waarmee rationeel en logisch gedacht kan worden in de wetenschap. Het betreft geïntegreerde procesvaardigheden zoals het definiëren van variabelen, hypothesen opstellen, figuren interpreteren en zuivere onderzoeken ontwerpen. Door middel van deze ‘process skills’ kunnen studenten reageren op informatie en zodoende oplossingen voor problemen bedenken (Burns, Okey, & Wise, 1985). In de beschrijving van deze vaardigheden is duidelijk de koppeling met wetenschappelijk redeneren en het uitvoeren van wetenschappelijk onderzoek te zien. Volgens Bonnstetter (1998) vormen de ‘process skills’ het gereedschap voor inquiry learning. De skills omvatten een breed scala van vaardigheden die studenten en wetenschappers gebruiken om informatie te verzamelen, patronen en relaties te onderzoeken en resultaten te communiceren. Bonnstetter verbindt door deze conclusie inquiry learning dus opnieuw met wetenschappelijk redeneren. De ‘process skills’ bestaan volgens Bonnstetter uit observeren, classificeren,

meten, voorspellen, interfereren, samenvatten, communiceren, data verzamelen, data analyseren, conclusies trekken, modellen bouwen, bewijzen interpreteren en experimenteren.

De Jong en Njoo (in de Jong & van Joolingen, 1998) brengen een onderscheid aan tussen *transformatieve processen* (bijvoorbeeld hypothesen opstellen, experimenten ontwerpen, observeren, data verzamelen analyseren en interpreteren, resultaten toepassen en voorspellingen doen op basis van de resultaten) en *regulatieve processen* (bijvoorbeeld plannen en monitoren). De regulatieve processen zijn nodig om het ontdekkingsproces te sturen. De transformatieve processen leveren direct kennis op. Zij beschrijven de stappen die toegepast worden bij het wetenschappelijk redeneren. Uit de opsomming van Hallonen et al. (2003) blijkt dat inquiry skills beide soorten processen bevatten. De drie eerstgenoemde domeinen in de rangschikking van Hallonen et al. (2003) –beschrijvende, conceptuele en probleemoplossende vaardigheden– bestaan uit transformatieve vaardigheden. Hieruit kan opgemaakt worden dat Hallonen et al. menen dat inquiry learning met name uit transformatieve processen bestaat.

In de beschreven definities van inquiry skills komen de volgende activiteiten vaak overeen: variabelen identificeren/definieren; hypothesen opstellen en toetsen; interfereren en voorspellen; experimenten ontwerpen; data analyseren en interpreteren; conclusies trekken; generaliseren van gegevens; rapporteren en communiceren van de resultaten. Enkele auteurs beschrijven ook samenwerkingsvaardigheden, maar dit aspect wordt niet altijd genoemd.

Kijkend naar het onderscheid van De Jong en Njoo, dan blijken de transformatieve processen de stappen van het wetenschappelijk redeneren te beschrijven. Deze stappen komen overeen met de stappen binnen het inquiry proces. De regulerende processen zijn minder specifiek voor inquiry learning van toepassing en komen ook bij andere vormen van leren of handelen voor. De rangschikking van Hallonen et al. sluit aan op deze conclusie. Voor het definitief van inquiry skills worden de volgende vier stappen/vaardigheden gehanteerd (1) het definiëren van variabelen, (2) het opstellen van hypothesen, (3) het ontwerpen van experimenten en (4) het beoordelen van gegevens / het trekken van conclusies.

Deze vier vaardigheden sluiten aan bij de process skills van Burns, Okey en Wise (1985) en Bonnstetter (1998) en de transformatieve processen van De Jong en Njoo (1998). De vaardigheden komen bovendien overeen met vier van de vijf stappen die Dewey presenteert als de stappen in een effectief inquiry proces. De eerste stap van Dewey (het zich voordoen van een probleem/de interne twijfel) ontbreekt in deze keuze. Dit hangt samen met de gekozen definitie van inquiry learning. Binnen het onderwijs krijgen de leerlingen te maken met een centrale vraagstelling of een vastgelegd onderwerp, zodat er geen sprake is van een inquiry proces waarbij gestart wordt met een interne twijfel of vraag.

Construeren van een test voor het meten van inquiry skills

Volgens Lavinghouzes (1997) kunnen inquiry skills van studenten het best gemeten worden door laboratoriumexperimenten en mondelinge presentaties. Hierdoor krijgt de student namelijk de kans om uit te leggen wat hij gedaan heeft en om te demonstreren dat het onderwerp begrepen is. Het beoordelen van vaardigheden door observatie van laboratoriosituaties kan echter moeilijk en tijdrovend zijn (Burns et al., 1985). Om deze reden richt dit onderzoek zich op de ontwikkeling van een domeinonafhankelijke test, met het besef dat ook domeinspecifieke aspecten een rol spelen in het inquiry proces.

Voor het testen van het vermogen om kritisch te redeneren of problemen op te lossen en voor het meten van ‘process skills’ zijn diverse tests ontwikkeld. Ennis (1993) wijst erop dat bij het ontwikkelen en gebruiken van een test voor kritisch denken het doel van de test helder moet zijn. Gezien de grote verwantschap tussen kritisch denken en inquiry learning is zijn opsomming met testdoelen ook hanteerbaar voor tests voor het meten van inquiry skills. Ennis maakt onderscheid tussen tests die één aspect van kritisch denken meten en testen die meerdere aspecten beogen te meten. Dit onderzoek richt zich op de tweede categorie, omdat gekozen is voor een viertal inquiry skills. De tests op het gebied van kritisch redeneren zijn volgens Ennis allemaal gebaseerd op algemene inhoud (niet onderwerpspecifiek). Lavinghousez (1997) noemt wel enkele onderwerpspecifieke tests, zoals de Biology Readiness Scale uit 1972. In dit onderzoek wordt slechts gekeken naar tests met niet-curriculum gerelateerde inhoud.

Op basis van het literatuuronderzoek zijn zeven tests op het gebied van critical thinking en proces skills geselecteerd. Elke test wordt beoordeeld op de volgende criteria:

- de aansluiting op de vier gekozen inquiry skills (het definiëren van variabelen, het opstellen van hypothesen, het ontwerpen van experimenten en het beoordelen van gegevens/het trekken van conclusies);
- de geschiktheid voor de doelgroep (Tweede Fase leerlingen);
- de meetpretentie (het niveau van inquiry learning vaststellen);
- de afnametijd: men heeft voor toepassing binnen het onderwijs met name behoefte aan een test die binnen één lesuur af te nemen is.
- de betrouwbaarheid (Cronbach's α is minimaal .70) en validiteit (andere auteurs noemen het instrument valide).

Op basis van testbeschrijvingen en geraadpleegde literatuur wordt elke test op deze aspecten als voldoende of onvoldoende beoordeeld. Indien een test inhoudelijk onvoldoende aansluit op de gekozen definitie worden de overige aspecten niet verder beoordeeld. Tabel 1 bevat een overzicht van de beoordelingsresultaten.

Volgens Young (1980) is de test die het meest gebruikt wordt in onderzoeken op het gebied van kritisch denken de Watson-Glaser Critical Thinking Appraisal (WGCTA, Watson & Glaser, 1964). Deze klassieke test is gebaseerd op de vijf vaardigheden die volgens Dressel en Mayhew (1954) bij critical thinking horen. De WGCTA bevat vijf subtests in twee parallelle vormen met elk 80 items (multiple choice en juist/onjuist). De test meet de aspecten inferentie/conclusie (inductie, van het bijzondere naar het algemene), assumptie (beoordelen van mogelijke veronderstellingen), deductie (van het algemene naar het bijzondere), interpretatie (grondige analyse van informatie) en evaluatie van argumenten (de sterkte van argumenten beoordelen). Op basis van de namen en beschrijvingen lijken deze subschalen niet direct aan te sluiten op de vier gedefinieerde inquiry skills. Nadere beschouwing toont echter aan dat bepaalde subschalen raakvlakken hebben met het beoordelen van gegevens/het trekken van conclusies'. De meeste overeenstemming hebben de subschalen 'Conclusie' en 'Interpretatie'. De subschaal 'Conclusie' doet een beroep op de vaardigheid waarmee de validiteit van conclusies beoordeeld wordt. Bij de subschaal 'Interpretatie' moet 'het bewijs gewogen' worden waarna besloten moet worden of de generalisatie of de conclusie op basis van de gegeven informatie valide is. Van versie C van de WGCTA is in 1997 een experimentele Nederlandstalige bewerking uitgegeven door Swets & Zeitlinger BV (thans: Harcourt): de Watson-Glaser Kritisch Denken Test (WGKDT, Van Zanten, Dekker & Berkhout, 1997). De WGKDT is geschikt voor personen op hbo/wo-niveau, met een leeftijd van minimaal 16 jaar. De beoogde doelgroep in dit onderzoek zit net op de grens van deze doelgroep. De test wordt met name gebruikt voor het selecteren van managers, het voorspellen van iemands vaardigheden in bepaalde functiegroepen en als meetinstrument binnen trainingsprogramma's op het gebied van kritisch denken. Dit laatste doel sluit enigszins aan bij de te ontwikkelen test. De WGKDT kent een acceptabele afnametijd van 40 minuten. De betrouwbaarheid van de WGCTA is goed (Cronbach's $\alpha = .81$). Er is veel onderzoek gedaan naar de validiteit van deze test. Er is een samenhang gevonden met intelligentie (een correlatie van circa 0.50), waarbij er meestal een sterker verband met de verbale dan met de numerieke component was.

Peter en Noreen Facione ontwikkelden, op basis van de definitie van de ideale kritische denker van de American Psychological Association, de California Critical Thinking Skills Test (CCTST, 1990) en de California Critical Thinking Dispositions Inventory (CCTDI, 1992). Hierbij wordt onderscheid gemaakt tussen de vaardigheden en de neiging of houding tot kritisch denken. Volgens Facione worden de vaardigheden niet gebruikt wanneer de neiging (houding) hiertoe niet bestaat en zonder de vaardigheden wordt de houding tot kritisch denken niet volledig aangeboord.

De CCTDI is een test waarin zeven houdingsaspecten zijn opgenomen, waaronder nieuwgierigheid en het willen zoeken naar de waarheid. Hoewel de houdingsaspecten verwant zijn aan inquiry learning, komen zij niet overeen met de vier gedefinieerde vaardigheden. Deze test is daarom niet verder beoordeeld.

De vaardigheden die met de CCTST gemeten worden, zijn: analyseren, evalueren, gevolgtrekkingen, deductief redeneren en inductief redeneren. De voor de vier gedefinieerde variabelen relevante subschalen

zijn analyseren van informatie en het maken van gevolgtrekkingen. Er bestaan drie versies van de CCTST: Form 2000, Form A (uitgegeven in 1990) en Form B (uitgegeven in 1992). Elke versie bevat 34 multiple choice items. Er kunnen subschaalscores worden berekend. De test kan afgenomen worden bij volwassenen en studenten. Ennis (1993) geeft aan dat de test wellicht ook bij de hogere niveaus van het voortgezet onderwijs gebruikt kan worden, wat aansluit op de beoogde doelgroep. De CCTST is bedoeld voor het meten van zowel individuele vaardigheden als groepsvaardigheden op het gebied van kritisch denken en redeneren. Dit sluit aan bij de criteria. De afnametijd van 45 minuten is acceptabel. Voor het meten van de convergente validiteit zijn diverse testen gehanteerd, waaronder de Watson-Glasertest en de SAT-scores. Met name Form 2000 is valide en betrouwbaar gebleken (Facione, Facione & Sanchez, 1994). Brody, Koenigseder, Ishee en Williams (2001) plaatsen met name bij de analyse subschaal vraagtekens bij de betrouwbaarheid en bij de mogelijkheid om individuele mogelijkheden te meten. De KR-20 waarden bij Form A en Form B (Facione et al., 1994) duiden ook op een lage tot middelmatige betrouwbaarheid. Bij Form 2000 variëren de KR-20 alfa's van .78 tot .84, afhankelijk van de testcontext.

De Test of Enquiry Skills (TES, Fraser, 1979) is gebaseerd op instructiemethoden die inquiry learning en individuele ontwikkeling stimuleren. Het is een serie tests die onderwerpafhankelijke onderzoeksvaardigheden (relevant voor de natuurwetenschappen, sociale wetenschappen, geschiedenis en aardrijkskunde) meet. De vaardigheden die aan bod komen, zijn: het gebruiken van referentiemateriaal (bibliotheek, index, inhoudsopgave), het interpreteren en gebruiken van informatie en het wetenschappelijk, kritisch redeneren. Door haar contentafhankelijke karakter is deze test niet geschikt voor het ontwerpen van een algemene test voor het meten van inquiry skills. Bovendien kent deze test met 87 items een afnametijd van 180 minuten, waardoor hij ook praktisch niet in aanmerking komt. De overige beoordelingscriteria zijn derhalve niet in beeld gebracht.

'The Integrated Process Skills Test' (TIPS) van Okey en Dillashaw (1980) werd ontwikkeld vanuit een behoefte aan een test voor het meten van niet-curriculum specifieke procesvaardigheden bij scholieren in het voortgezet onderwijs (grade 7 tot 12 in de Verenigde Staten). De test beoogt vaardigheden die bij rationeel en logisch redeneren in de wetenschappen gebruikt worden te meten. Door de ontwikkeling van TIPSII (Okey, Wise, & Burns, 1982) is het totale aantal beschikbare items uitgebreid. TIPS I en TIPS II hebben dezelfde moeilijkheidsgraad en kunnen gezien worden als twee zeer gelijkwaardige tests. Zowel TIPS I als TIPS II bevatten 36 items in de volgende vijf subtests: variabelen identificeren, operationeel definiëren, hypothesen opstellen, data en grafieken interpreteren en onderzoeken ontwerpen. De vaardigheden die gemeten worden door de vijf subschalen van de TIPS sluiten allen aan op de vier gekozen vaardigheden. Deze test dekt de vier vaardigheden geheel: variabelen identificeren en operationeel definiëren komen overeen met het definiëren van variabelen, data en grafieken interpreteren sluit aan bij het beoordelen van gegevens/het trekken van conclusies en hypothesen opstellen en onderzoeken ontwerpen komen zelfs letterlijk overeen. Deze test is speciaal gericht op leerlingen in het voortgezet onderwijs en binnen één lesuur af te nemen. High school studenten, vergelijkbaar met de beoogde doelgroep in dit onderzoek, hadden slechts 35 tot 30 minuten nodig om de test te completeren (Burns et al., 1985). Met een Cronbach's alfa van .86 is de betrouwbaarheid van de TIPS II goed te noemen. De splittest correlatiecoëfficiënten tussen TIPS II en de originele TIPS bedragen 0.86 en 0.90. Over de TIPS II melden Burns et al (1985) dat dit een nieuw betrouwbaar instrument is voor het beoordelen van de vaardigheden van leerlingen in het voortgezet onderwijs op het gebied van geïntegreerde science process skills. TIPS II kan ook gebruikt worden in onderwijsprogramma's voor docenten of voor het evalueren van curricula of leermaterialen gericht op procesvaardigheden (Burns et al., 1985).

De Cornell Critical Thinking Tests (CCTT, Ennis & Millman, 1985) kennen ook twee vormen. Deze test kan vanaf 12-13 jarige leeftijd worden afgenomen en bevat de twee niveaus: X en Z. Niveau X bevat de subschalen inductie, deductie, geloofwaardigheid en het identificeren van vooronderstellingen. In niveau Z worden aan de schalen uit niveau X de schalen denkfouten, definities, voorspellen en experimenten plannen toegevoegd. Het onderdeel experimenten plannen komt overeen met de variabele het ontwerpen van experimenten. De overige onderdelen sluiten niet direct aan op de gedefinieerde inquiry skills. De CCTT kunnen gebruikt worden om kritische redeneervaardigheden aan te leren, om de prestaties

van leerlingen op (Amerikaanse) examens te voorspellen, in cursussen over kritisch denken of tijdens loopbaantrainingen binnen het bedrijfsleven. De tests bevatten 71-multiple-choiceitems die in vijftig minuten beantwoord moeten worden. De betrouwbaarheid van het instrument varieert bij verschillende populaties tussen .87 en .91. Stein, Haynes en Unterstein (2003) melden expliciet dat er nog geen validiteitonderzoeken zijn verricht.

De Critical Reasoning Test (CRT, Smith, & Whetton, 1992) bestaat uit een verbaal en een non-verbaal gedeelte. De testpersoon moet zich bij het verbale gedeelte inleven in de situatie van managertrainees en krijgt 56 items over een fictief bedrijf. De items van dit gedeelte zijn onderverdeeld in drie delen: analyse van informatie, evaluatie (beoordelen van argumenten op objectieve gronden) en assumptie (beoordelen van mogelijke veronderstellingen). De score op het numerieke gedeelte geeft inzicht in het kritisch denkvermogen met betrekking tot non-verbaal materiaal. De totale afname duurt circa één uur. Deze test kan bij volwassenen afgenomen worden en wordt met name gebruikt bij het selecteren en beoordelen van medewerkers in (toekomstige) managementfuncties. De meetpretentie en het specifieke karakter van deze test sluiten derhalve niet aan op de te ontwikkelen test. Om deze reden zijn ook bij deze test de overige beoordelingscriteria niet beoordeeld.

Tabel 1
Scoring van de onderzochte tests op de beoordelingscriteria

Beoordelingscriteria	WGKDT	CCTDI	CCTST	TES	TIPSII	CCTT	CRT
Definiëren van variabelen	-	-	+	-	+	-	
Opstellen van hypothesen	-	-	-	-	+	-	
Ontwerpen van experimenten	-	-	-	-	+	+	
Trekken van conclusies	+	-	+	-	+	-	
Geschikt voor doelgroep	+		+		+		-
Meetpretentie	-			-	+		-
Afnametijd	+		+	-	+	+	
Betrouwbaarheid	+		+		+	+	
Validiteit	+		+		+	-	

Noot 1: WGKDT = Watson Glaser Kritisch Denken Test, CCTDI = California Critical Thinking Dispositions Inventory, CCTST = California Critical Thinking Skills Test, TES = Test of Enquiry Skills, TIPSII = The Integrated Process Skills Test II, CCTT = Cornell Critical Thinking Tests, CRT = Critical Reasoning Test;

Noot 2: - = de test scoort onvoldoende op dit criterium; + = de test scoort voldoende op dit criterium.

Samenstelling instrument

Bij het bepalen van de geschiktheid van de bestaande tests krijgt de aansluiting op de gedefinieerde inquiry skills prioriteit. Vanuit dit oogpunt is de TIPSII de meest geschikte test, doordat de vier inquiry skills allen door deze test gemeten worden. Hoewel ook de CCTST bij veel beoordelingscriteria positief naar voren kwam, sluit deze test niet volledig aan op de vier inquiry skills. Daarnaast kent de CCTST een lagere betrouwbaarheid en een langere afnametijd dan de TIPSII. Om deze redenen is gekozen voor de TIPSII als basis voor de constructie van een nieuwe test.

De TIPSII is vanuit het Engels naar het Nederlands vertaald. In de TIPSII komen met name de vaardigheden uit de eerste drie fasen in het inquiry proces (het definiëren van variabelen, het opstellen van hypothesen en het ontwerpen van experimenten) terug. De vierde vaardigheid, het beoordelen van gegevens/het trekken van conclusies, komt in de TIPSII vooral tot uiting in vragen met grafieken. Omdat dit een beperking opleverde bij het samenstellen van een instrument dat de vier inquiry skills volledig behelst, is besloten om het meetinstrument uit te breiden met onderdelen waarin deze vierde vaardigheid op basis van verbale informatie aan bod komt. Binnen de onderzochte tests werd een oplossing gevonden in de onderdelen 'Conclusie' en 'Interpretatie' uit de WGKDT. In deze twee subtests moet de respondent verbale informatie verwerken en beoordelen. Beide subtests bevatten meerdere uiteenzettingen van feiten.

Op basis van de gegeven informatie moet beoordeeld worden of de getrokken conclusies of interpretaties logisch voortvloeien uit de feiten.

Onderzoeksvraag

Op basis van het literatuuronderzoek zijn definities voor inquiry learning en inquiry skills gekozen. Vervolgens werd een test samengesteld die de inquiry skills beoogt te meten. Het doel van het uitgevoerde experiment is nagaan of de nieuwe test, de Inquiry Skills Test (IST), inderdaad in staat is om inquiry skills te meten. Om de validiteit van de test te bepalen, wordt bij een groep leerlingen van 4VWO naast de test een leertaak afgenomen die een beroep doet op inquiry skills. Bovendien wordt als controlemaat een cognitieve capaciteitentestbatterij afgenomen. De cognitieve capaciteiten vormen over het algemeen een goede voorspeller voor inquiry learning, omdat de inquiry skills een sterk beroep lijken te doen op de cognitieve vaardigheden (Veenman, Wilhelm, & Beishuizen, 2004). Middels een meervoudige regressieanalyse wordt nagegaan welke test het best in staat is om de prestaties op de inquirytaak te voorspellen.

Methode

Respondenten

Aan het onderzoek hebben 25 leerlingen uit een vierde klas van het VWO van één scholengemeenschap deelgenomen. De klas bestond uit 11 meisjes en 14 jongens. De gemiddelde leeftijd van de respondenten was 16.16 jaar ($SD = .47$, range = 15-17). Zij hadden allemaal de Nederlandse taal als moedertaal. De scholengemeenschap is geselecteerd op basis van haar bereidheid te willen deelnemen aan het onderzoek. De leerlingen werden voorafgaand aan het onderzoek op de hoogte gebracht van de (verplichte) deelname. Het experiment verving twee uitgevallen lesuren.

Instrumenten³

De samenstelling van de nieuwe test (Inquiry Skills Test, IST) is reeds in de inleiding beschreven. De IST is in een toetsboekje opgenomen en wordt als een papieren test afgenomen.

The Integrated Process Skills Test II (TIPSII) is letterlijk vertaald, de inhoud bevatte geen specifieke cultuurgebonden aspecten die een alternatieve vertaling noodzakelijk maakten. De test bevat vijf subschalen: variabelen identificeren (12 items), hypothesen opstellen (9 items), operationeel definiëren (6 items), onderzoeken opzetten (3 items) en grafieken en data interpreteren (6 items). De TIPSII bestaat uit 36 meerkeuzeitems met ieder vier antwoordmogelijkheden, waarvan steeds één antwoord juist is. In Figuur 1 is één vraag uit de TIPSII opgenomen. Wegens lay-outtechnische redenen is een wijziging doorgevoerd in de volgorde van de vragen 34, 35 en 36.

21. Een manager van groentekassen wil de productie van tomatenplanten versnellen. Zij plant tomatenzaad in verschillende bakken. Haar hypothese is dat hoe meer vochtigheid het zaad ontvangt, des te sneller het zaad ontspruit. Hoe kan zij deze hypothese testen?

 - A. Tellen hoeveel dagen het duurt totdat het zaad ontspruit wanneer het verschillende hoeveelheden water ontvangt.
 - B. Eén dag na elke bewatering de hoogte van de tomatenplanten meten.
 - C. De hoeveelheid water die de planten in de verschillende bakken gebruiken meten.
 - D. Het aantal tomatenzaadjes dat in elke bak geplant is tellen.

Figuur 1. Een item uit de vertaalde TIPSII

Het onderdeel Conclusie van de Watson Glaser Kritisch Denken Test (WGKDT) omvat drie uiteenzettingen van feiten met in totaal zestien conclusies die door de respondent beoordeeld moeten worden. De vijf antwoordcategorieën zijn: waar, waarschijnlijk waar, onvoldoende informatie, waarschijnlijk onwaar en onwaar. Het onderdeel Interpretatie bevat zes uiteenzettingen met in totaal

³ Alle gebruikte instrumenten zijn op te vragen bij de auteur.

zestien interpretaties die als juist of onjuist beoordeeld moeten worden. In Figuur 2 is een gedeelte uit het onderdeel Interpretatie opgenomen.

Uiteenzetting: Een verkoper van Dermatrix Lotion verkondigde dat zijn product in een handomdraai spierpijn zou verlichten, door in de pijnlijke lichaamsdelen door te dringen. De verkoper bracht tien druppels lotion op een dik stuk schoenleer aan, dat de lotion al snel opnam.		
Mogelijke interpretaties:	Juist	Onjuist
1. De verkoper toonde de genezende werking van het product aan.		
2. Het was de bedoeling van de verkoper te suggereren dat als de lotion door een dik stuk schoenleer heen kon dringen, die ook door zou kunnen dringen in pijnlijke spieren.		
3. De demonstratie van de verkoper was een bewijs voor zijn bewering dat de lotion spierpijn kan verlichten.		

Figuur 2. Een uiteenzetting met drie interpretaties uit het onderdeel Interpretatie van de WGKDT

De testbatterij voor cognitieve capaciteiten bestond uit vijf subtests, te weten: Lineaire syllogismen, Cijferreeksen, Verborgene figuren, Woordenschat en Woordanalgieën. De subtests zijn gekozen op basis van empirisch vastgestelde verbanden met prestaties op inquiry learning taken (Wilhelm, 2001) en op de praktische mogelijkheden om in korte tijd klassikaal een maat te krijgen voor de cognitieve vaardigheden. Bij het bepalen van een keuze voor de subtests, is in navolging van Veenman et al (2004) gekozen voor de subtests lineaire syllogismen (Abstract Syllogisms Task - Conclusions, Elshout, 1976); cijferreeksen (Number Series Test, Elshout, 1976) en verborgen figuren (Hidden Figures Test, Flanagan, 1951). Dit betreffen subtests waarvan bekend is dat zij een samenhang vertonen met inquiry taken. Volgens Sternberg (1985) hangt Woordenschat het hoogst samen met de algemene intelligentie. Derhalve werd een Woordenschat subtest (zie Veenman, 1994) aan de testbatterij toegevoegd. Bovendien werd er een subtest Woordanalgieën (zie Veenman, 1994) aan de testbatterij toegevoegd om de verhouding tussen verbale en performale onderdelen in stand te houden. Van de subtest Woordanalgieën is ook aangetoond dat hij correleert met een taak waarbij een beroep gedaan wordt op inquiry skills (Wilhelm, 2001). Er werd een toetsboekje geconstrueerd met de instructies, de vijf subtests en de antwoordformulieren.

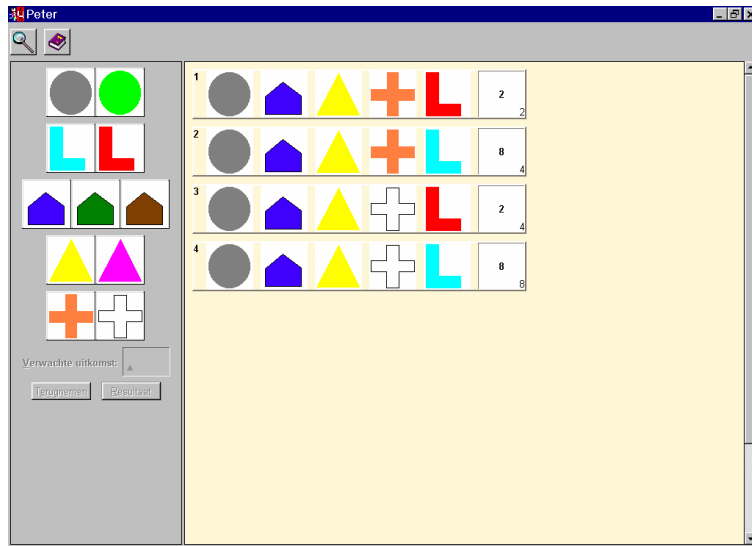
Tevens werd bij elke respondent via een computer een inquirytaak afgenomen. Hiervoor werd gebruik gemaakt van FILE (Flexible Inquiry Learning Environment, Hulshof, Wilhelm, Beishuizen & Van Rijn, 2005). Er is gekozen voor een abstracte, domeinonafhankelijke inquirytaak om effecten van voorkennis zoveel mogelijk uit te sluiten.

In de inquirytaak moesten de respondenten telkens vijf figuren (cirkel: grijs of lichtgroen, letter L: lichtblauw of rood, huisje: blauw, donkergroen of bruin, driehoek: geel of roze en kruis: oranje of wit) als inputvariabelen uit het linkerframe selecteren (zie Figuur 3). De respondent moest zelf de waarde van de geselecteerde variabelen voorspellen en kreeg vervolgens de werkelijke waarde én de verwachte waarde van de output in beeld. In het rechterframe vormden de experimenten rijen. Middels experimenteren met de inputvariabelen moesten de respondenten uitzoeken waarom het ene groepje figuren meer punten opleverde dan het andere. De inquirytaak kon volledig met de muis worden bediend.

In de FILE-taak waren variabelen opgenomen die een positieve invloed hadden, variabelen die geen invloed hadden en variabelen die een interactie met elkaar hadden (zie Tabel 2). De variabele “kruis” en “driehoek” hadden geen invloed op de uitkomst. Bij de variabele “huisje” was de blauwe waarde minder punten waard dan de groene en bruine waarde, die evenveel waard waren. Bij een groene waarde op de variabele “cirkel” had de variabele “letter L” geen invloed, maar bij een grijze waarde op de variabele “cirkel” gaf de blauwe waarde op de variabele “letter L” meer punten dan de rode. Bij een blauwe waarde op de variabele “letter L” leidde de grijze waarde bij de variabele “cirkel” tot een hogere score en bij een rode waarde op de variabele “letter L” leidde de groene waarde bij de variabele “cirkel” tot meer punten.

De respondenten hadden de mogelijkheid om reeds uitgevoerde experimenten te selecteren en apart te bekijken. Tevens kon de instructie op elk moment worden nagelezen. De respondenten konden zoveel

experimenten uitvoeren als zij zelf wilden. Op een antwoordformulier konden de leerlingen hun bevindingen per variabele en eventuele overige bevindingen noteren.



Figuur 3. De abstracte FILE inquirytaak

Onderzoeksdisein

Omdat de verschillende onderdelen (IST en inquiry skills) in het onderzoek wegens testeffecten elkaars resultaten wellicht zouden beïnvloeden, werd op verschillende manieren gecontroleerd voor volgorde-effecten. Allereerst werden twee versies van de IST ontwikkeld: versie A met eerst de TIPSII items en vervolgens de onderdelen Conclusie en Interpretatie en versie B met eerst de twee WGKDT onderdelen en vervolgens de TIPSII items. Bovendien werd onderscheid gemaakt in de volgorde van het aanbieden van de IST en de inquirytaak.

Door de beschreven onderzoeksopzet ontstonden vier aselekt gekozen groepen respondenten. Na de klassikale cognitieve capaciteitentests maakten 6 leerlingen eerst de inquirytaak en vervolgens versie A van de IST (groep 1), 6 leerlingen maakten eerst de inquirytaak en vervolgens versie B van de IST (groep 2), 6 leerlingen maakten eerst versie A van de IST en vervolgens de inquirytaak (groep 3) en 7 leerlingen maakten eerst versie B van de IST en vervolgens de inquirytaak.

Procedure

Het experiment vond plaats in één sessie in een computerlokaal binnen de scholengemeenschap. Alle onderdelen werden klassikaal afgenomen. De leerlingen werden willekeurig in het lokaal gepositioneerd, waarbij zoveel mogelijk één plaats tussen twee leerlingen werd vrijgehouden om spieken en overleg te voorkomen.

Eerst werden de vijf subtests uit de cognitieve capaciteitentestbatterij afgenomen. De respondenten kregen een tijdsduur van zeven minuten voor elke subtest. De keuze voor deze afnametijd werd met name bepaald door de totale beschikbare tijd. Voorafgaand aan elke subtest kregen de respondenten de tijd om de instructie te lezen. De proefleider gaf telkens het begin en het einde van de beschikbare tijd aan. Aan het einde van dit onderdeel werden alle toetsboekjes ingenomen.

Na dit eerste onderdeel startte de helft van de respondenten met de inquirytaak en de andere helft met de IST. Zowel voor de inquirytaak als voor de IST was geen tijdslimiet vastgesteld. Alle instructies stonden op papier. De respondenten meldden zich bij de proefleider op het moment dat zij met het laatste onderdeel wilden starten.

Scoring

Met behulp van scoringsmallen is bij de vijf subtests voor het meten van de cognitieve capaciteiten telkens één punt aan het juiste antwoord toegekend. Per respondent is het aantal juiste antwoorden per subtest berekend. Vervolgens is een totaalscore voor de cognitieve capaciteiten verkregen door het aantal juiste antwoorden bij de verschillende subtests bij elkaar op te tellen. Bij de IST is een gelijke scoringsmethode gehanteerd.

Voor de beoordeling van de antwoordformulieren bij de inquirytaak is een scoringsprotocol opgesteld waarin de relaties tussen de variabelen als uitgangspunt dienden (zie Tabel 2). Voor het volledig beschrijven van een gevonden relatie kreeg de respondent 2 punten. Bij de variabelen met interactie werd 1 punt gegeven indien de respondent wel de relatie, maar niet de bijbehorende restrictie noemde. Voor de conclusie “de blauwe L is meer waard dan de rode L” werd bijvoorbeeld 1 punt gegeven. Indien de respondent hierbij bijvoorbeeld had vermeld: “indien de groene cirkel gekozen wordt”, scoorde hij 2 punten. Foute antwoorden leverden geen punten op. Bij het scoringsprotocol is een lijst samengesteld met termen en bewoordingen van de respondenten en het aantal punten dat hiervoor werd gegeven. De helft van de antwoordformulieren is ter controle door een onafhankelijke beoordelaar met behulp van het protocol en de lijst met bewoordingen gescoord. Dit leverde een Cohen’s kappa van .904 op, wat duidt op een zeer hoge mate van overeenstemming.

Tabel 2

Scoringsprotocol voor de antwoordformulieren bij de abstracte FILE inquirytaak

Relatie tussen de symbolen uit de inquirytaak	Punten bij correct	Punten bij noemen relatie zonder restrictie	Punten bij fout
rode L = blauwe L (als cirkel = groen)	2	1	0
rode L < blauwe L (als cirkel = grijs)	2	1	0
groene cirkel > grijze cirkel (als L = rood)	2	1	0
grijze cirkel > groene cirkel (als L = blauw)	2	1	0
blauw huisje < groen huisje	2	-	0
blauw huisje < bruin huisje	2	-	0
groen huisje = bruin huisje	2	-	0
paarse driehoek = gele driehoek	2	-	0
oranje kruis = wit kruis	2	-	0

Resultaten

Prestaties respondenten

De scores van de respondenten op de cognitieve capaciteitentest, de IST en de inquirytaak zijn weergegeven in Tabel 3. De scores op de cognitieve capaciteitentest, de inquirytaak en de IST waren allen normaal verdeeld met Kolmogorov-Smirnov waarden van respectievelijk $Z = .675$ ($p = .752$), $Z = 1.044$ ($p = .226$) en $Z = .471$ ($p = .980$). De verkregen data voor de ontdekkende inquirytaak leverden één uitschieter (score 18) op. De score van 18 punten week meer dan tweemaal de standaarddeviatie ($SD = 1.7$) af van de gemiddelde score ($M = 12.0$). Door het kleine aantal respondenten had de uitschieter tevens een sterke invloed op de correlaties tussen de inquirytaak en de andere onderdelen. Derhalve is besloten om deze score buiten de analyses te houden.

Omdat één (dyslectische) respondent niet alle onderdelen van de IST binnen de beschikbare tijd kon afronden, is de score van deze leerling buiten beschouwing gelaten. Hierdoor was het totale aantal respondenten bij de IST ($n = 24$) kleiner dan het aantal respondenten op de overige onderdelen ($n = 25$).

Een éénwegvariatie-analyse leverde geen significante volgordeverschillen tussen de vier groepen op. De groepsindeling had geen effect op de scores op de inquirytaak ($F(3,21) = 1.17$, $p = .343$) of op de scores op de IST ($F(3,20) = 1.31$, $p = .298$).

Tabel 3

Resultaten van de respondenten op de verschillende onderdelen

Subtest	n	Maximaal te behalen aantal punten	Gemiddelde	Standaard-deviatie	Laagste score	Hoogste score
Cognitieve capaciteiten	25	211	93.6	14.2	61	121
<i>Woordenschat</i>	25	60	12.4	5.1	5	23
<i>Lineaire syllogismen</i>	25	50	23.8	4.8	13	34
<i>Woordanalogieën</i>	25	40	31.0	4.4	22	39
<i>Cijferreeksen</i>	25	45	22.3	4.3	14	28
<i>Verborgene figuren</i>	25	16	4.1	2.2	1	9
IST	24	68	46.9	5.8	37	56
<i>TIPSII</i>	25	36	25.9	4.1	18	33
<i>WGKDT</i>	24	32	21.0	2.6	17	26
Inquirytaak	25	18	12.0	1.7	10	15

Betrouwbaarheid van de tests

Van de gehanteerde tests is de interne betrouwbaarheid nagegaan (zie Tabel 4). De interne consistentie van de cognitieve capaciteiten testbatterij bleek redelijk ($\alpha = .676$) te zijn. Ook de interne betrouwbaarheid van de IST als geheel was redelijk ($\alpha = .686$). Nunnally (1978) beschouwde .70 als een acceptabele betrouwbaarheidscoëfficiënt, maar in de literatuur worden ook lagere drempels gehanteerd. De betrouwbaarheid van de IST kwam vrijwel overeen met die van de TIPSII ($\alpha = .684$).

In tabel 4 valt op dat de subtests van de IST niet intern consistent waren. Met uitzondering van de subschaal Variabelen identificeren ($\alpha = .748$) van de TIPSII vertoonden de subschalen lage of zelfs negatieve Cronbach's alfa's. Dit betekent dat verdere analyses van de subschalen niet verricht zullen worden. Omdat de interne betrouwbaarheid van het gehele instrument voldoende hoog was, is besloten om de gegevens op dit niveau wel verder te analyseren.

Tabel 4

Betrouwbaarheid van de cognitieve capaciteitentest, de IST en haar subtests

	Aantal items	Cronbach's alfa
Cognitieve capaciteitentest	211	.676
IST	58	.686
<i>TIPSII</i>	36	.684
Variabelen identificeren	12	.748
Operationeel definiëren	6	-.073
Hypotheses opstellen	9	.316
Data en grafieken interpreteren	6	-.200
Onderzoeken ontwerpen	3	.097
<i>WGKDT</i>	32	.304
WGKDT - Conclusie	16	.394
WGKDT - Interpretatie	16	-.474

Correlatie- en regressieanalyse

Op de scores op de cognitieve capaciteitentest (bestaande uit vijf subtests), de IST (bestaande uit twee delen, die elk weer subschalen bevatten) en de abstracte inquirytaak is een correlatieanalyse uitgevoerd (zie Tabel 5). Er werd een hoge correlatie gevonden tussen de scores op de IST en de scores op de inquirytaak ($r = .616$, $p < .01$). Ook de cognitieve capaciteitentest vertoonde een significante correlatie

met de inquirytaak ($r = .499, p < .01$). De correlatie tussen de IST en de totale scores op de cognitieve capaciteitentest bleek laag en niet significant ($r = .243, p = .263$) te zijn.

De sterke correlatie tussen de IST en de scores op de inquirytaak ontstond met name door de hoge significante correlatie tussen de TIPSII en de inquirytaak ($r = .574, p < .01$). Met name de subschaal Variabelen identificeren vertoonde veel samenhang met de inquirytaak ($r = .558, p < .01$). De correlatie van de TIPSII met de inquirytaak werd met .042 verhoogd door de toevoeging van de WGKDT schalen.

Door een meervoudige regressieanalyse via de 'Enter'-methode uit te voeren met de IST en de cognitieve capaciteitentest als onafhankelijke variabelen (voorspellers) en de scores op de inquirytaak als afhankelijke variabele werd nagegaan welk percentage van de variantie in de scores op de inquirytaak werd verklaard door de gezamenlijke en de unieke voorspellers. Het totale percentage verklaarde variantie door de IST en de cognitieve capaciteitentest bedroeg 44.9%. Alleen de IST bleek significant bij te dragen aan de voorspelling van de prestaties op de inquirytaak ($\beta = .550, p < .01$), de cognitieve capaciteitentest bleek geen significante voorspeller te zijn ($\beta = .273, p = .127$). De 'Stepwise'-methode wees uit dat de IST 37.9% van de variatie in de scores op de inquirytaak verklaarde.

Tabel 5

Correlaties tussen scores op de IST, de inquirytaak en de cognitieve capaciteitentest

	Inquirytaak	Cognitieve capaciteiten	IST	TIPSII
Inquirytaak				
Cognitieve capaciteiten	.499*			
IST	.616**	.243		
TIPSII	.574**	.194	.918**	
WGKDT - Conclusie+Interpretatie	.370	.150	.764**	.445*

Noot: * $p < .05$; ** $p < .01$.

Conclusie en discussie

Na het definiëren van inquiry learning en inquiry skills werd een test samengesteld waarmee beoogd werd vier typen inquiry skills te meten, te weten het definiëren van variabelen, het opstellen van hypothesen, het ontwerpen van experimenten en het beoordelen van gegevens/het trekken van conclusies. Door de betrouwbaarheid en validiteit van de ontwikkelde test te onderzoeken, werd nagegaan of de gekozen benadering voor de constructie van de test juist was.

De vier groepen in het onderzoek bleken niet significant verschillend te scoren op de leertaak en de IST. Dit betekent dat het voor de scores niet uitmaakt of de leertaak voor of na de IST gemaakt wordt en dat de volgorde van de onderdelen binnen de IST ook niet van invloed is. In verder onderzoek kan wellicht volstaan worden met slechts één versie van de IST.

De totale IST bleek redelijk betrouwbaar te zijn. Een groot aantal subschalen van de TIPSII en de WGKDT bleken in dit specifieke onderzoek niet voldoende betrouwbaar. In literatuur over de WGTCa (Watson & Glaser, 1980a) en de TIPSII (Burns et al, 1985) werden geen vergelijkbare lage of negatieve Cronbach's alfa's gevonden. Een mogelijke verklaring voor de negatieve Cronbach's alfa's zou kunnen liggen in het kleine aantal respondenten en het kleine aantal items in enkele schalen. Wellicht hebben de leerlingen enkele items niet begrepen of was het niveau toch te moeilijk. Ook de antwoordmogelijkheden met een hoge gokkans of mogelijke domeinafhankelijke aspecten in de subtests zouden van invloed geweest kunnen zijn. De beperkte betrouwbaarheid van de subtests vereist een vervolgonderzoek, zodat ook analyses op van de subschalen mogelijk worden.

De gevonden correlatie tussen de IST en een inquirytaak lijkt erop te duiden dat met de IST inquiry skills gemeten worden. Zoals veel literatuur veronderstelt, correleren ook in dit onderzoek cognitieve capaciteiten én inquiry skills met de prestaties op een inquirytaak. Tussen de scores op de IST en de scores op de cognitieve capaciteitentests werden echter geen significante correlaties gevonden. Beide tests meten waarschijnlijk andere constructen. Bovendien blijkt uit de resultaten dat alleen de IST significant bijdraagt

aan de voorspelling van de score op de inquirytaak wanneer beide predictoren in de regressievergelijking worden opgenomen.

De IST vertoont bovendien meer samenhang met de prestaties op de gebruikte inquirytaak dan de afzonderlijke subtests waaruit de IST is opgebouwd. Door de uitbreiding van de TIPSII met de WGKDT schalen werd de correlatie met de inquirytaak met .042 verhoogd. In de betrouwbaarheid trad vrijwel geen verandering op. Inhoudelijk is er voldoende reden om de WGKDT schalen niet uit de IST te verwijderen. Ondanks het feit dat nader onderzoek vereist is, lijkt de IST een instrument te zijn dat in een (Nederlandse) behoefte kan voorzien.

Voor de ontwikkeling van de IST is het wenselijk om bij een volgende afname gebruik te maken van een grotere groep respondenten van verschillende leeftijden en op diverse scholen. Het is raadzaam om in een volgend onderzoek ook een concrete (inhoudsafhankelijke) inquirytaak te gebruiken. Nader onderzoek kan leiden tot verbetering van het instrument en tot een hogere generaliseerbaarheid van de resultaten naar andere leeftijden, scholen en wellicht andere instellingen of organisaties.

Referenties

- Bonnstetter, R.J. (1998). Inquiry: Learning from the past with an eye on the future. *Electronic Journal of Science Education*, 3 (1). Retrieved, 27 februari 2004, from <http://unr.edu/homepage/jcannon/ejse/bonnstetter.html>
- Brody, K.N., Koenigseder, L.A., Ishee, J.H. & Williams, B.G. (2001). Psychometric properties of the California Critical Thinking Tests. *Journal of Nursing Measurement*, 9 (3), pp. 309-328.
- Burns, J.C., Okey, J.R., & Wise, K.C. (1985). Development of an integrated process skill test: TIPSII. *Journal of Research in Science Teaching*, 22 (2), pp. 167-177.
- Chinn, C.A., & Malhotra, B.A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*. 86 (2), pp. 175–218.
- Dewey, J. (1933/1989). How We Think. *The Later Works. Vol. 8*. Carbondale: Southern Illinois UP. pp. 105-352.
- Dewey, J. (1938/1991). Logic: A Theory of Inquiry. *The Later Works. Vol. 12*. Carbondale: Southern Illinois UP. pp. 77-103.
- Dillashaw, F.G., & Okey, J.R. (1980). Test of integrated process skills for secondary science students. *Science Education*. 64, pp. 601-608.
- Dressel, P.L. & Mayhew, L.B. (1954). *General Education: Explorations in Evaluation*. Washington: American Council on Education.
- Ennis, R.H. (1993). Critical thinking assessment. *Theory into Practice*, 32 (3), pp. 179-186.
- Ennis, R.H. & Millman, J. (1985). *Cornell critical thinking test, level X*. Pacific Grove CA: Critical Thinking Books and Software (formerly Midwest Publications).
- Facione, N., Facione, P. & Sanchez, C. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, 33(8), pp. 345–351.

- Franken, A. (2003). *ICT in beeld: Een onderzoek naar toekomstscenario's van ICT-gebruik in het primair en secundair onderwijs*. Utrecht: ICO-ISOR/Onderwijsresearch.
- Fraser, B.J. (1979). *Test of Enquiry Skills*. Melbourne: Australian Council for Educational Research.
- Hallonen, J.S., Bosack, T., Clay, S., & McCarthy, M. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30 (3), pp. 196-208.
- Hogan, K., & Fisherkeller, J. (2000). Dialogue as data: Assessing students' scientific reasoning with interactive protocols. In J. J. Mintzes, J. H. Wandersee & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 95-127). San Diego: Academic Press.
- Hulshof, C.D., Wilhelm, P., Beishuizen, J.J. & Rijn, H. van (2005). FILE: a tool for the study of inquiry learning. *Computers in Human Behavior*, 21(6), pp. 945-956.
- Jacobs, S.S. (1999). The equivalence of forms A and B of the California Critical Thinking Skills Test. *Measurement & Evaluation in Counselling & Development*, 31 (4), pp. 211-223.
- Jong, T. de, & Van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68 (2), pp. 179-201.
- Kessels, J.W.M., & Poell, R.F. (Ed.) (2001). *Human Resource Development: Organiseren van het leren*. Groningen: Samsom.
- Kuhn, D. (Ed.) (1990). Developmental perspectives on teaching and learning thinking skills. *Contributions to Human Development* (21). New York: Karger.
- Kuhn, D. (1993). Connecting scientific and informal reasoning. *Merrill-Palmer Quarterly*, 39 (1), pp. 74-103.
- Lavinghousez, W.E. (1997). *Evaluation and assessment*. Retrieved, 9 januari 2005 from: <http://pages.xtn.net/~blavingh/evaluation.html>
- McPeck, J.H. (1981). *Critical thinking and education*. New York: St. Martins Press.
- Nederlandse Organisatie voor Wetenschappelijk Onderzoek/Programmaraad voor het Onderwijsonderzoek (2003). *Programmaboek Onderwijsonderzoek 2004-2007*. Den Haag: NWO.
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Okey, J.R., Wise, K.C., & Burns, J.C. (1982). *Integrated process skills test II*. University of Georgia, Department of science education.
- Paul, R. (1987). Critical thinking and the critical person. In *Thinking: Report on Research*. New York: Erlbaum.
- Petress, K. (2004). Critical thinking: an extended definition. *Education*, 124 (3), pp. 461-465.
- Reid, D.J., Zhang, J., & Chen, Q. (2003). Supporting scientific discovery learning in a simulation environment. *Journal of Computer Assisted Learning*, 19, pp. 9-20.

- Schafersman, S.D. (1991). *An introduction to critical thinking*. Retrieved, 27 februari 2005, from <http://www.freeinquiry.com/critical-thinking.html>.
- Smith, P., & Whetton, C. (1992) *Critical Reasoning Tests*. Windsor: NFER-Nelson.
- Stein, B.S., Haynes, A.F., & Unterstein, J. (2003). *Assessing Critical Thinking Skills*. Tennessee: Technological University.
- Sternberg, R.J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York.
- Stokking, K.M., & van der Schaaf, M.F. (1999). *Beoordelen van onderzoeksvaardigheden van leerlingen: Richtlijnen, alternatieven en achtergronden*. Utrecht: Universiteit Utrecht, Onderwijskunde, ISOR.
- Veenman, M.V.J. (1994). Intelligentie en probleemaanpak. *De Psycholoog*, 29, pp. 223-228.
- Veenman, M.V.J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction*, 14, pp. 89-109.
- Watson, G., & Glaser E.M. (1964). *Watson-Glaser Critical Thinking Appraisal*. New York: Harcourt, Brace and World.
- Watson, G., & Glaser, E.M. (1980a). *Watson-Glaser Critical Thinking Appraisal-Form A*. San Antonio-Texas: Psychological Corporation.
- Wells, G. (1999). *Dialogic inquiry in education: building on the legacy of Vygotsky*. New York: Cambridge University Press. Retrieved, 9 februari 2005, from <http://tortoise.oise.utoronto.ca/~gwells/NCTE.html>
- Wilhelm, P. (2001). *Knowledge, skills and strategies in self-directed inductive learning*. Doctoral dissertation, Leiden University.
- Young, R.E. (Ed.) (1980). Fostering critical thinking. *New directions for teaching and learning*, (3).
- Zanten, D.L. van, Dekker, R., & Berkhout, C. (1997). *Watson-Glaser kritisch denken test*. Lisse: Swets & Zeitlinger.