

See discussions, stats, and author profiles for this publication at: <http://www.researchgate.net/publication/254919455>

Metacognitieve activiteiten meten: Een vragenlijst vergeleken met de hardopdenkmethode

ARTICLE · JANUARY 2008

READS

290

4 AUTHORS, INCLUDING:



[Gonny L.M. Schellings](#)

Technische Universiteit Eindhoven

37 PUBLICATIONS 105 CITATIONS

[SEE PROFILE](#)



[Bernadette Van Hout-Wolters](#)

University of Amsterdam

138 PUBLICATIONS 957 CITATIONS

[SEE PROFILE](#)

Metacognitieve activiteiten meten:

Een vragenlijst vergeleken met de hardopdenkmethode.

Gonny Schellings¹, Bernadette van Hout-Wolters¹, Marcel Veenman^{1,2} & Joost Meijer¹.

¹Universiteit van Amsterdam & ²Universiteit Leiden.

Onderwijsonderzoek heeft geleid tot een diversiteit aan meetinstrumenten om metacognitie in kaart te brengen (zie bijv. Veenman, van Hout-Wolters, & Afflerbach, 2006). De verschillende instrumenten of meetmethoden kunnen echter gericht zijn op verschillende componenten van metacognitie en daardoor leiden tot verschillende resultaten (van Hout-Wolters, submitted); en indirect ook tot verschillen in theoretische bevindingen. Muis, Winne en Jamieson-Noel (2007) wijzen daarom op het belang om verschillende metacognitieve meetmethoden te evalueren in relatie tot onderwijsleertheorieën.

Vanuit de onderwijspraktijk is er vraag naar gemakkelijk bruikbare meetmethoden om metacognitieve activiteiten in kaart te brengen. Leerlingbegeleiders en docenten gebruiken bijvoorbeeld vragenlijsten of observatiegegevens om leerlingen te selecteren die extra begeleiding nodig hebben. Zowel onderzoekers als docenten vragen zich echter regelmatig af welke meetmethode voor hun situatie het meest is aan te raden. Om hen te ondersteunen bij de keuze van een geschikte meetmethode, presenteert Van Hout-Wolters (submitted) een overzicht van verschillende meetmethoden, daarbij ingaand op de methodologische sterke en zwakke punten van de afzonderlijke meetmethoden. Vooral in de onderwijspraktijk zal men vragen om efficiënte meetmethoden. Daarom wordt vaak met vragenlijsten gewerkt, zoals bijvoorbeeld de MSLQ (Pintrich, et al., 1993), de LASSI (Weinstein, 1987) en de ILS (Vermunt, 1992). Metacognitie-vragenlijsten zijn zeer gebruiksvriendelijk, kosten weinig verwerkingstijd en men kan vaak hoge interne consistenties rapporteren (bijv. Mokhtari & Reichard, 2002; Muis, et al., 2007). Een methodologisch zwak punt betreft echter de validiteit, met name de convergente validiteit (zie van Hout-Wolters, submitted; Veenman, 2005). Convergente validiteit betreft de samenhang die er bestaat tussen soortgelijke meetinstrumenten (Brinkman, 2000). Of in andere woorden: meten de metacognitie-vragenlijsten hetzelfde als andere methoden die metacognitieve activiteiten meten? Veenman (2005) presenteert een overzicht van 'multi-method' studies waarin verschillende metacognitieve meetmethoden met elkaar vergeleken worden. De correlatiematen die tussen vragenlijsten en de hardopdenk methode gevonden worden, variëren van laag tot gematigd (-.07 tot .31). De vraag is nu of we de convergente validiteit

van de vragenlijst (vaker gebruikt in de praktijk) t.o.v. de hardopdenk methode (vaker gebruikt in onderzoek) kunnen verhogen, zodat de vragenlijst gebruikt kan worden als een efficiënt en gebruiksvriendelijk alternatief voor de hardopdenkmethode.

Dat de correlaties tussen vragenlijsten en hardopdenk-maten laag zijn kan in de eerste plaats liggen aan het feit dat de leerlingen moeite hebben met het verbaal rapporteren over de eigen uitgevoerde leeractiviteiten tijdens de taakuitvoering. Maar dit is niet de enige mogelijke verklaring. Dat vragenlijsten lage correlaties vertonen met de hardopdenkmethode kan óók liggen aan het feit dat beide methoden iets anders meten. De hardopdenkmethode is altijd gericht op de activiteiten die de leerlingen tijdens de taakuitvoering verrichten, met andere woorden, de hardopdenk methode is altijd taakspecifiek. Een vragenlijst, daarentegen, kan ook gericht zijn op een algemene studeeraanpak of op een specifieke leertaak, waarop de hardopdenkmethode nu niet was gericht. Een andere mogelijke verklaring voor de lage correlatie is dat de betreffende meetmethoden gericht zijn op inhoudelijk verschillende metacognitieve activiteiten. Rekening houdend met deze verklaringen willen we in dit onderzoek nagaan of we de convergente validiteit tussen een vragenlijst en de hardopdenkmethode kunnen verhogen. Kortom, wat is de relatie tussen vragenlijstgegevens en hardopdenkresultaten als beide meetmethoden gericht zijn op dezelfde leertaak én op dezelfde metacognitieve activiteiten?

Onderzoeksmethode.

Participanten.

In het onderzoek participeerden 16 leerlingen van een grotere scholengemeenschap uit een middelgrote stad uit het zuiden van het land (8 havo-3 en 8 vwo-3 leerlingen). Alle leerlingen kregen geschiedenisles van dezelfde geschiedenisdocent. Klassikaal werd aan de leerlingen gevraagd of zij wilden participeren aan een onderzoek. De leerlingen, die zich zelf opgaven, werden na hun inzet beloond met een cadeaubon.

Materiaal.

De geschiedenisstekst. De 16 leerlingen bestudeerden elk een geschiedenisstekst over de eerste Afrikanen in Noord-Amerika, de slavernij en de oorzaken en verloop van de Amerikaanse Burgeroorlog. De tekst bestond uit 1650 woorden verdeeld over 6 paragrafen en 7 subparagrafen. Elke (sub)paragraaf werd aangegeven met een (sub)titel. De gemiddelde zinslengte was 17.64 woorden. In de geschiedenisstekst stonden twee open vragen, die de leerlingen tijdens het hardop denken ook hardop denkend moesten beantwoorden. Deze twee vragen hadden vooral het doel de leerlingen aan te zetten tot het uitvoeren van metacognitieve activiteiten, omdat 'simpelweg de tekst lezen' zou kunnen resulteren in magere hardopdenk protocollen.

De vragenlijst. In tegenstelling tot veel andere vragenlijsten, werd de vragenlijst rechtstreeks gebaseerd op een codeerschema of taxonomie om hardopdenk protocollen te analyseren. De taxonomie was in eerder onderzoek ontwikkeld vanuit theoretische indelingen en eigen empirische resultaten (Meijer, Veenman & van Hout-Wolters, 2006). Een zeer uitgebreide en fijnschalige categorieënsysteem (het oorspronkelijke systeem bestond uit 70 afzonderlijke categorieën ingedeeld in 6 hoofdcategorieën) werd op een cyclische manier door inhoudelijke en statistische analyses teruggebracht tot een taxonomie met 4 hoofdcategorieën en 56 specifieke categorieën (Meijer, et al., 2006). Voor het huidige onderzoek werd elke categorie uit de taxonomie (waar mogelijke met behulp van protocolfragmenten uit het eerdere onderzoek van Meijer et al., 2006) omgezet in een vraag voor de vragenlijst (zie tabel 1). Omdat de stellingen in de vragenlijst herformuleringen zijn van de categorieën in de taxonomie, zijn vragenlijst en taxonomie dus gericht op dezelfde metacognitieve activiteiten. Daarnaast is de vragenlijst expliciet afgestemd op de leertaak (er werd in de vragen verwezen naar het leren van de studietekst), en werd deze na het bestuderen van de tekst afgenomen.

Tabel 1. Constructie vragenlijst: Categorieën worden omgezet in vragen.

Label	Voorbeeld
Categorie	Planning: Vaststellen taakeisen
Protocol-fragment	Mag ik dit onderstrepen? Moet ik dit opschrijven? Moet ik nu een antwoord geven?
Vraag	Ik ging na wat er van mij werd verwacht.
Categorie	Uitvoering: Aantekeningen maken, onderstrepen.
Protocol-fragment	Ik onderstreep dat 1855 Ik ga even opschrijven dat het uhh, dat het in 1619 was
Vraag	Ik heb in de tekst onderstreept, woorden gemarkeerd of aantekeningen gemaakt.
Categorie	Monitoring: Vaststellen onbegrip.
Protocol-fragment	Uhhh. Vraag 2 weet ik even niet ... eh, ik weet niet Ik weet het niet. Ik denk dat, ja, weet ik niet.
Vraag	Tijdens het lezen, begreep ik bepaalde delen van de tekst niet.

Categorie	Elaboratie & evaluatie: Inferentie afleiden.
Protocol-fragment	Oké, dus..hij was eigenlijk helemaal niet zo geweldig Lincoln. Want hij .. Eigenlijk was het alleen maar eigen belang, oké ...
Vraag	Ik probeerde conclusies te trekken die verder gingen dan de letterlijke tekst.

Opbouw vragenlijst. De vragenlijst bestaat uit 58 vragen¹ die gereduceerd kunnen worden tot vier schalen. De schalen zijn:

1. *planning*: De activiteiten die horen bij deze schaal vinden vóór en tijdens het leren plaats. De activiteiten betreffen bijvoorbeeld: het activeren van voorkennis, bedenken waar de tekst over zou kunnen gaan, en nagaan wat de taakeisen zijn. Nadat de taakeisen zijn vastgesteld, kunnen een plan van aanpak en eigen subdoelen worden vastgesteld. Tijdens het lezen kan in de gaten gehouden worden of de subdoelen wel bereikt worden bijvoorbeeld door de eigen gedachten op een rijtje te zetten door zichzelf vragen te stellen, etc. De planning-schaal bestaat uit 16 items.
2. *uitvoering*: Na het bedenken van een plan van aanpak, moet het plan uitgevoerd worden: bepaalde gedeelten uit de tekst worden geselecteerd om te lezen, sommige tekstfragmenten kunnen overgeslagen worden bij het lezen; andere tekstgedeelten worden vaker gelezen, er worden notities gemaakt, etc. De uitvoeringsschaal bestaat uit 12 vragen.
3. *monitoring*: Monitoring vindt vooral tijdens het leren plaats. De leerling houdt zich zelf in de gaten door bijvoorbeeld tijdens het lezen onbegrip te constateren, door te merken dat er onbekende woorden in de tekst staan, door te merken dat er iets niet goed is gelezen en zichzelf daarbij te verbeteren. Deze schaal bestaat uit 14 vragen.
4. *elaboratie & evaluatie*: De activiteiten die bij deze schaal worden ingedeeld vinden grotendeels na het lezen plaats. De leerling herhaalt de gelezen tekst in eigen woorden, trekt conclusies, maakt een (uitgebreide) samenvatting, of leest de belangrijke woorden en/of titels in de tekst nog eens over. Na het lezen, wordt er ook nagegaan of de (hele) tekst begrepen is en of de gestelde doelen bereikt zijn. De elaboratie & evaluatie-schaal bestaat uit 16 items.

De respondenten gaven op een 3-punten-schaal aan of ze de activiteit bijna nooit (=1), soms (=2), of vaak (=3) hebben uitgevoerd. We maakten gebruik van een 3-punten-schaal omdat, volgens Veenman (2005), het gevaar bestaat dat bij bredere schalen (vijf of zeven punten) leerlingen eerder verschillende referentiepunten gebruiken bij de invulling.

De hardopdenkmethode. De leerlingen lazen de tekst hardop en als er een gedachte bij hen opkwam moesten zij die verwoorden. In de tekst stonden geen prompts of aanwijzingen om hardop te (blijven)

denken. Als de leerlingen stil bleven en/of als de leerlingen lang zonder verbalizaties lazen, stimuleerde de testleider hen om hardop te blijven denken met steeds dezelfde aanwijzingen (zoals: “Blijf hardop denken!”, “Wat denk je nu?”). De hardopdenk sessies werden opgenomen en letterlijk uitgetypt. De protocollen werden geanalyseerd met behulp van de taxonomie (56 afzonderlijke categorieën met vier hoofdcategorieën: planning, uitvoering, monitoring en elaboratie & evaluatie). Elke opmerking van de participanten dat niet in de letterlijke tekst stond werd gecodeerd.

Procedure

De onderzoekssessies vonden individueel plaats. De leerlingen werden gevraagd de geschiedenis tekst goed te bestuderen alsof zij de tekst bestudeerden in voorbereiding op een proefwerk. Tekstbestudering moest hardop plaats vinden en werd digitaal opgenomen. De instructie over hardop denken was kort: de leerlingen moesten hardop de tekst leren en daarbij al hun gedachten verwoorden. Nadat de leerlingen aangaven klaar te zijn met het leren, kregen zij de metacognitieve vragenlijst voorgelegd. De vragenlijst werd stil ingevuld.

Resultaten

Vragenlijst.

De psychometrische gegevens van de vragenlijst moeten voorzichtig worden geïnterpreteerd, omdat de onderzoekspopulatie ($n=16$) voor de afname van een vragenlijst redelijk klein is. De interne consistentie van de hele vragenlijst is bevredigend (Cronbach's $\alpha = .87$). Opvallend is dat de interne consistentie voor de afzonderlijke schalen sterk uiteen loopt, van laag voor de uitvoering ($\alpha = .26$) en monitoring ($\alpha = .36$) schalen tot hoger voor de planning-schaal ($\alpha = .52$) en de elaboratie & evaluatie-schaal ($\alpha = .85$).

Tabel 2. Interne consistentie vragenlijst ($n=16$)

Schaal	Aantal items	Cronbach's α
Planning	16	.52
Uitvoering	12	.26
Monitoring	14	.36
Elaboratie&evaluatie	16	.85
Vragenlijst	58	.87

Hardopdenkprotocollen.

Twee beoordelaars codeerden onafhankelijk van elkaar 13 van de 16 protocollen. De codes van de twee beoordelaars werden paarsgewijs samengevat in een kruistabel en de Pearson's contingency coefficient werd vastgesteld. Deze maat wordt bijvoorbeeld door Garson (2004) geadviseerd om de associatie tussen twee nominale variabelen te bepalen bij grote kruistabellen (i.e. groot aantal rijen en kolommen). De interbeoordelaarsbetrouwbaarheid was meer dan acceptabel (Pearson contingency coefficient = .97).

In totaal werden er 2817 activiteiten geteld. Het aantal getelde activiteiten binnen een protocol kon ver uit elkaar lopen: in een mager protocol werden slechts 50 activiteiten geturfd, terwijl in een rijk protocol 384 activiteiten werden geturfd. Het gemiddeld aantal getelde activiteiten per protocol was 176.06 (sd=92.63). Het aantal getelde activiteiten per hoofdcategorie liep ook sterk uiteen (zie tabel 3). Vooral uitvoerings- (40%) en monitoring-activiteiten (30%) werden geteld, gevolgd door elaboratie & evaluatie-activiteiten (20%). Planning-activiteiten werden het minst vaak gecodeerd (10%).

Tabel 3. Aantal geturfdde activiteiten in de protocollen.

Categorieën	Aantal activiteiten	Percentage
Planning	296	10.5%
Uitvoering	1120	39.8%
Monitoring	842	29.9%
Elaboratie & evaluatie	559	19.8%
In totaal	2817	100%

Vragenlijst vergeleken met de hardopdenkprotocollen.

De totaalcorrelatie tussen vragenlijst en hardopdenk protocollen is zeer bevredigend, namelijk, $r=.512$ ($p=.051$) en duidelijk hoger dan in eerder onderzoek. Op het oog lijkt deze correlatie gematigd, maar de onderzoekspopulatie is klein, waardoor alleen zeer sterke samenhangen gevonden kunnen worden.

In tegenstelling tot in veel ander onderzoek, kunnen we hier ook de correlaties op schaalniveau berekenen. We kijken hierbij alleen naar de schalen met redelijke tot hoge interne consistenties (zie tabel 4): de planning-schaal en de elaboratie & evaluatieschaal. Hierbij vinden we wel een correlatie voor de elaboratie & evaluatie-schaal, maar geen correlatie voor de planning-schaal. De leerlingen lijken zich beter bewust van de uitgevoerde elaboratie & evaluatie-activiteiten dan van planning-activiteiten.

Tabel 4. Correlaties tussen vragenlijst en protocollen op schaalniveau.

	Planning	Elaboratie & evaluatie
Planning ($\alpha=.52$)	.067	
Elaboratie & evaluatie ($\alpha=.85$)		.546

Conclusie

De onderzoeksvraag betrof de relatie tussen vragenlijst- en hardopdenkresultaten, terwijl beide meetmethoden gericht zijn op dezelfde leertaak én op dezelfde metacognitieve activiteiten. In dit onderzoek wordt een hogere totaalcorrelatie tussen de instrumenten ($r=.512$) gevonden dan in veel ander onderzoek, waar de correlatie varieerde van $-.07$ tot $.31$. Omdat de instrumenten beide gericht zijn op dezelfde leertaak, i.e. het leren van een geschiedenis tekst, en omdat beide instrumenten zich richten op dezelfde metacognitieve activiteiten die eerder al gevonden zijn in hardopdenk protocollen verzameld tijdens het leren van een tekst, is de convergente validiteit tussen beide instrumenten verhoogd.

Eerder onderzoek presenteert vaak alleen de totaalcorrelaties tussen verschillende meetinstrumenten, maar in dit onderzoek kunnen wij ook kijken naar correlaties op schaalniveau. De schalen in de vragenlijst zijn immers dezelfde als de schalen in de taxonomie. Voor de elaboratie & evaluatie-schaal vinden we wel een behoorlijke correlatie, maar voor de planning-schaal niet. Leerlingen weten kennelijk beter welke elaboratie- en evaluatieactiviteiten ze hebben verricht, dan dat ze bij het invullen van de vragenlijst zich iets kunnen herinneren van hun planning-activiteiten. Het kan echter ook zijn dat bij de antwoorden op de laatste categorie sociale wenselijkheid een rol speelde: leerlingen geven op de vragenlijst wel aan dat zij planning-activiteiten hebben uitgevoerd, maar tijdens het coderen van de protocollen werden deze activiteiten het minst vaak geteld (10%).

Verder onderzoek naar meetinstrumenten, die gebruik maken van zelf-rapportage van leerlingen (bijvoorbeeld vragenlijsten) om metacognitieve activiteiten in kaart te brengen, zou rekening moeten houden met onze bevinding dat leerlingen over de ene soort van metacognitieve activiteiten gemakkelijker (en meer valide) kunnen rapporteren dan over andere activiteiten-soorten. Ook op item-niveau hebben we indicaties gevonden dat leerlingen over één item meer valide rapporteren dan over een ander item.

Het is jammer dat de interne consistenties van de uitvoering- en monitoring schalen van de vragenlijst laag bleken te zijn. In correlationele analyses op schaalniveau zijn deze schalen dan ook buiten beschouwing gelaten. In vervolgonderzoek zal aan dit punt zeker aandacht besteed moeten worden. Misschien zien de onderzoekers die de taxonomie hebben opgesteld toch een andere samenhang/structuur in de activiteiten op deze schalen, dan de respondenten op vragenlijst. Een

aanvullend interview met de respondenten of hen de vragenlijst hardop denkend laten invullen zou op dit gebied verhelderend kunnen werken. Daarnaast zou het niet onverstandig zijn om de vragenlijst in een grotere respondenten-groep af te nemen, zodat er meer waarde gehecht kan worden aan de gevonden psychometrische gegevens.

De totaalcorrelatie tussen de vragenlijst- en de hardopdenk methode die in dit onderzoek is gevonden, is bemoedigend, zeker ook gezien de lage totaalcorrelaties uit eerder onderzoek. Het bevorderen van metacognitieve activiteiten is een belangrijk onderwijsdoel en daarvoor blijven betrouwbare en valide meetinstrumenten nodig; en dan het liefst instrumenten die gemakkelijk te gebruiken zijn en weinig verwerkingstijd vragen, zoals de vragenlijst. Daarnaast hebben vragenlijsten ook nog eens het voordeel dat op de resultaten binnen grote onderzoekspopulaties multivariate analyses en 'multiple scaling' toegepast kunnen worden, hetgeen belangrijk is voor de construct-validiteit. Deze vorm van validiteit wordt bepaald door te onderzoeken welke psychologische kwaliteit de vragenlijst nu precies meet, of in engere zin geformuleerd, is het begrip-zoals-gemeten wel gelijk aan het begrip-zoals-bedoeld (i.e. het theoretische begrip) (Peet, et al. 1995). De eerste stap voor een meer valide vragenlijst die is afgestemd op het leren uit studieteksten is gezet. Wellicht dat de vragenlijst na verbetering nog meer overeenkomstige resultaten met de hardopdenk methode oplevert. In dat geval kan het onderwijsveld worden verrijkt met een vragenlijstmethode, die sneller is in te zetten dan de arbeidsintensieve hardopdenk-methode, maar die toch dezelfde metacognitieve activiteiten meet.

Referenties

- Brinkman, J. (2000). *De vragenlijst*. Groningen/Houten: Wolters-Noordhoff.
- van Hout-Wolters, B.H.A.M. (submitted). *Leerstrategieën meten. Soorten meetmethoden en hun bruikbaarheid in onderwijs en onderzoek*. Ook gepresenteerd als paper op de Onderwijsresearchdagen, 2006 Vrije Universiteit Amsterdam.
- Garson, D. (2004). *Nominal Association: Phi, Contingency Coefficient, Tschuprow's T, Cramer's V, Lambda, Uncertainty Coefficient*, from <http://www2.chass.ncsu.edu/garson/pa765/assocnominal.htm>
- Meijer, J., Veenman, M.V.J. , & van Hout-Wolters, H.H.A.M. (2006). Metacognitive activities in text studying and problem solving: Development of a taxonomy. *Educational Research and Evaluation*, 12, 209-237.
- Mokhtari, K. & Reichard, C.A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology*, 94, 249-259.

- Muis, K.R., Winne, P.H., & Jamieson-Noel, D (2007). Using a multitrait- multimethod analysis to examine conceptual similarities of three self-regulated learning inventories. *British Journal of Educational Psychology*, 77, 177-195.
- van Peet, A.A.J., van den Wittenboer, G.L.H., Hox, J.J. (1995). *Toegepaste statistiek: Beschrijvende technieken*. Groningen: Wolters-Noordhoff.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). *A manual for the use of the motivated strategies for learning questionnaire*. Ann Arbor, MI: University of Michigan.
- Veenman, M.V.J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77-99). Münster: Waxmann.
- Veenman, M.V.J., van Hout-Wolters, B.H.A.M. & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning*, 1, 3-14.
- Vermunt, J.D.H.M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs. Naar procesgerichte instructie in zelfstandig denken* [Learning styles and regulation of learning in higher education –Toward process-oriented instruction in autonomous thinking]. Amsterdam: Swets & Zeitlinger.
- Weinstein, C. E. (1987). *LASSI user's manual*. Clearwater, FL: H&H Publishing.

¹ Twee categorieën uit de taxonomie, namelijk: “woorden of titels overslaan bij het lezen” en “samenvatten door titels of vetgedrukte woorden te herlezen” werden beide omgezet in twee stellingen. De vragenlijst bestaat uit 58 stellingen; de taxonomie bestaat uit 56 categorieën.