

- point in L1 grammar education. *Language and Linguistics Compass*, 18(3), 1–17. <https://doi.org/10.1111/lnc3.12513>
- Stichting Leerplanontwikkeling (SLO). (2025). *Conceptexamenprogramma Nederlandse taal en literatuur - vwo (versie 2)*. <https://www.slo.nl/@24900/conceptexamenprogramma-nederlands-vwo/>
- Stoel, G. L., Drie, J. P. van, & Boxtel, C. A. M. van (2015). Teaching towards historical expertise: developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, 47(1), 49–76. <https://doi.org/10.1080/00220272.2014.968212>
- Stoel, G. L., Drie, J. P. van, & Boxtel, C. A. M. van (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, 109(3), 321–337. <https://doi.org/10.1037/edu0000143>
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Drophmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., ... Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76. <https://doi.org/10.3758/s13423-017-1323-7>
- Wellington, J. (2015). *Educational research contemporary issues and practical approaches* (2nd ed.). Continuum. <https://doi.org/10.5040/9781474236966>
- Wijnands, A. (2025). *Linguistic sources in grammar education. Fostering reflective thinking about language*. [Proefschrift, Radboud Universiteit Nijmegen]. <https://doi.org/10.54195/9789465150949>
- ASTRID WIJNANDS is hoeschooldocent Nederlands aan de Hogeschool Utrecht. Zij is aan de Radboud Docenten Academie gepromoveerd met een NWO-promotiebeurs voor leraren. Haar onderzoek richt zich op het gebruik van taalkundige bronnen in grammaticaonderwijs ter bevordering van het reflectief denken over taal.
E-mail: Astrid.Wijnands@hu.nl
- GERHARD STOEL is universitair docent gammadidactiek en vakdidacticus geschiedenis aan de Radboud Universiteit Nijmegen. Zijn onderzoek richt zich op het ondersteunen van docenten in gammavakken bij het ontwikkelen van onderwijs, waarin domein-specifiek redeneren centraal staat.
E-mail: Gerhard.Stoel@ru.nl
- MARIJN GIJSEN is universitair docent aan de Tilburg School of Humanities and Digital Sciences van de Universiteit van Tilburg. Zijn onderzoek richt zich op videobaseerd leren en cognitieve processen.
E-mail: M.H.M.Gijsen@tilburguniversity.edu
- PETER-ARNO COPPEN is emeritus hoogleraar vakdidactiek aan de Radboud Universiteit Nijmegen. Zijn onderzoek richt zich op de didactiek van het taalbeschouwingsonderwijs.
E-mail: Peter-Arno.Coppen@ru.nl

De cognitieve validiteit van meerkeuzeopgaven; De leesexamens Nederlands havo en vwo nader onderzocht

PATRICK ROOIJACKERS & CORIEN STEVENS

Meerkeuzeopgaven in het CE Nederlands havo/vwo

De centrale examens (CE's) Nederlands havo en vwo bestaan momenteel – zoals ook veel internationale leestoetsen – deels uit meerkeuzeopgaven. Deze meerkeuzeopgaven dragen echter vaak relatief weinig bij aan het differentiërend vermogen van de CE's. In deze studie onderzoeken we hiervoor één specifieke verklaring: de mogelijk gebrekkige cognitieve validiteit van meerkeuzeopgaven. De denkprocessen die meerkeuzeopgaven bij lezers uitlokken, zijn onvoldoende in lijn met de denkprocessen die lezers bij alledaagse begripstaken doorlopen (Cohen & Upton, 2006).

Havo 5- en vwo 6-leerlingen (N= 25) maakten hardop denkend een CE-onderdeel, bestaande uit meerkeuze- en open opgaven. Hun antwoordgedrag werd vervolgens met een scoreformulier en een analyse van hardop-denktopocollen in kaart gebracht. Uit de resultaten blijkt vooral dat de persoonlijke tekstinterpretatie van vwo-leerlingen regelmatig botst met de bewoordingen in de antwoordopties van meerkeuzeopgaven, waardoor vwo-leerlingen regelmatig niet het juiste antwoord kiezen; voor havo-leerlingen kan dit slechts zeer beperkt worden aangetoond. Implicaties van de resultaten voor de constructie van de CE's Nederlands havo/vwo worden besproken.

Al sinds decennia bestaan de Centrale Examens (CE's) Nederlands havo/vwo deels uit meerkeuzeopgaven, aansluitend bij de internationale consensus dat meerkeuzeopgaven, net als open opgaven, een valide instrument vormen om tekstbegrip te meten (bijvoorbeeld Leslie & Caldwell, 2017). Momenteel bestaan deze CE's voor 20-30% uit meerkeuzeopgaven, voor 20-30% uit semigesloten opgaven¹ en voor 40-50% uit open opgaven. Vooral door de open opgaven behoren de CE's Nederlands havo/vwo tot de meest correctie-intensieve examens in het voortgezet onderwijs (Essen et al., 2023).

Uit een recente enquête blijkt dat docenten Nederlands graag minder open opgaven willen in het CE Nederlands (CvTE, 2025). Dat zou ervoor pleiten vaker meerkeuzeopgaven of semigesloten opgaven op te nemen. Beoordelaars zijn hieraan immers beduidend minder correctietijd kwijt dan aan open opgaven.

Echter, de meerkeuzeopgaven hebben binnen deze CE's een relatief beperkte statistische meerwaarde: ze dragen vaak weinig bij aan het differentiërend vermogen van de toets.

schoolsoort	tijdvak	gem. <i>rir</i>		
		meerkeuzeopgaven	semigesloten opgaven	open opgaven
havo	2023-1	0,11	0,16	0,19
	2024-1	0,10	0,16	0,18
	2025-1	0,05	0,20	0,20
vwo	2023-1	0,09	0,17	0,18
	2024-1	0,06	0,10	0,18
	2025-1	0,14	0,18	0,18

Tabel 1. Gemiddelde item-restcorrelatie (*rir*) van drie opgavetypen in CE Nederlands havo/vwo, voor zes examens, per schoolsoort

Uit tabel 1 is af te lezen dat de meerkeuzeopgaven in de zes CE's van 2023-2025 laag correleren met de andere opgaven: de gemiddelde item-restcorrelatie (*rir*) van meerkeuzeopgaven is in alle examens lager dan die van de twee andere opgavetypen. Deze resultaten worden grotendeels weerspiegeld in de CE's van eerdere jaren. Bereikt het CE als geheel doorgaans een relatief acceptabele betrouwbaarheid ($0,65 \leq GLB \leq 0,75$), de bijdrage van de meerkeuzeopgaven hieraan is beperkt, al doen ze er evenmin afbreuk aan.

Een logische verklaring voor dit beperkte statistische functioneren is dat meerkeuzeopgaven vaak inhoudelijke constructiefouten kennen. Daarvoor ontbreken echter aanwijzingen. In de handreikingen uit 2024 en 2025 van Levende Talen Nederlands (LTN), de vakvereniging van leraren Nederlands, is slechts een enkele opmerking te vinden over de inhoudelijke juistheid van meerkeuzeopgaven (LTN, 2024, 2025), en via de Examenlijn komen slechts incidenteel klachten binnen over onvolkomenheden in dit opgavetype. Ook is het niet zo dat de afleiders in meerkeuzeopgaven vaak apert statistisch disfunctioneren. Dat roept de vraag op hoe de beperkte meerwaarde van meerkeuzeopgaven dan wel valt te verklaren.

Mogelijke bezwaren tegen meerkeuzeopgaven
Uit internationaal onderzoek zijn verschillende verklaringen te herleiden voor het mindere functioneren in leestoetsen van meerkeuzeopgaven ten opzichte van open opgaven. Er wordt wel beweerd dat meerkeuzeopgaven andere dimensies van leesvaardigheid meten dan open opgaven. De onderzoeksresultaten op dit punt zijn echter niet eenduidig. Zo vond Rodriguez (2003) duidelijke overeenkomsten tussen beide opgavetypen: als er een meerkeuzevariant en een open variant bij eenzelfde tekstplaats worden geconstrueerd, dan is de gemiddelde correlatie tussen beide varianten zeer hoog (Cordón & Day, 1996). Pearson et al. (1999) daarentegen stelden weer vast dat meerkeuzeopgaven een duidelijk lagere hoeveelheid aan 'meervoudige' en 'intertekstuele' strategieën uitlokten dan open opgaven: deelnemers verbonden bijvoorbeeld bij meerkeuzeopgaven minder vaak ideeën uit verschillende tekstdelen aan elkaar dan bij open opgaven. Ozuru et al. (2007) concludeerden op basis van de sterk uiteenlopende onderzoeksresultaten dat het effect van opgavetypen complex is en afhangt van de tekst en individuele opgaven. Veel internationale leestoetsen gebruiken daarom een mix van vraagvormen, zodat eventuele

beperkingen van meerkeuzeopgaven kunnen worden ondervangen (Brunfaut, 2021; Leslie & Caldwell, 2017). Deze werkwijze wordt ook in het CE Nederlands havo en vwo toegepast.

Twijfels over de validiteit van meerkeuzeopgaven bestaan al decennia. Farr et al. (1990) noemen drie mogelijke, deels samenhangende problemen. Ten eerste: doordat toetsmakers met meerkeuzeopgaven slechts één betekenis toekennen aan een tekst(deel) en mogelijke antwoorden aanreiken, zouden lezers zich bij het beantwoorden van meerkeuzeopgaven vaak enkel richten op de oppervlaktebetekenis van tekstdelen, waardoor ze nauwelijks diep begrip opbouwen (Bussis & Chittenden, 1987). Ten tweede: in het juiste antwoord (de 'sleutel') op een meerkeuzeopgave brengen toetsmakers de betekenis van een tekst(deel) terug tot één enkele bewering, terwijl leestheorieën vooropstellen dat lezen een constructief proces is waarbij elke lezer een persoonlijk getint begrip van het gelezene opbouwt; meerkeuzeopgaven zouden met name aan het situatiemodelbegrip van lezers onvoldoende recht doen (Kintsch, 1998). Ten derde: doordat lezers bij meerkeuzeopgaven de juiste antwoordoptie moeten selecteren, zetten zij antwoordstrategieën in die vooral als testmanagementstrategieën zijn te beschouwen: strategieën die bij een alledaagse begripstaak zelden of nooit een rol spelen. Lezers komen bij meerkeuzeopgaven vaak niet tot hun antwoord door zelfstandig een eigen antwoord op de vraagstelling te formuleren (eventueel na uitgebreide tekstraadpleging), om daarna de meerkeuzeopties te lezen en ten slotte vanuit hun eigen antwoord op de vraagstelling de best passende meerkeuzeoptie te selecteren. Lezers baseren hun antwoord vaak op het stapsgewijs falsificeren of verifiëren van de meerkeuzeopties op basis van selectieve tekstraadpleging, logica of voorkennis (Rupp et al., 2006). Een meerkeuzeopgave fungeert daardoor meer als een talige puzzel dan als

een begripstaak (Leslie & Caldwell, 2017).

De laatste twee decennia heeft men deze kritiek scherper in beeld gebracht en typeert men deze problemen inmiddels als een problematische *cognitieve validiteit* van meerkeuzeopgaven (bijvoorbeeld Cohen & Upton, 2006; O'Reilly et al., 2018). Een meerkeuzeopgave kan een complex, 'onnatuurlijk' antwoordproces uitlokken waarin de meerkeuzeopties van meet af aan centraal staan, waardoor details in formuleringen een grote impact op het antwoordproces kunnen hebben en waardoor minder vanzelfsprekende meerkeuzeopties gaandeweg het antwoordproces evolueren tot aannemelijke antwoorden.

Deze validiteitsproblemen hangen sterk samen met de eenduidige betekenisgeving aan teksten in meerkeuzeopgaven. De meeste leestheorieën veronderstellen dat lezers een tekst adequaat begrijpen wanneer ze een coherent, tekstconform 'situatiemodel' ervan hebben opgebouwd: een mentaal model waarin lezers de tekstuele wereld integreren in hun eigen voorkennis, opvattingen en ideeën (Kintsch, 1998; Magliano et al., 2018). Pas bij dit situatiemodel is er sprake van 'diep begrip'. Daartegenover staat tekstbasisbegrip: het vrij letterlijke begrip van wat een tekst expliciet stelt. Idealiter sluit toetsing van diep begrip vooral aan op situatiemodelbegrip, en niet zozeer op tekstbasisbegrip (Magliano et al., 2018). Juist voor het meten van situatiemodelbegrip lijken meerkeuzeopgaven zich minder goed te lenen dan open opgaven, doordat open opgaven meer ruimte bieden aan de interpretatie van de individuele lezer (Leslie & Caldwell, 2017).

Er zijn bovendien aanwijzingen dat cognitieve validiteitsproblemen van meerkeuzeopgaven sterker optreden naarmate leestoetsen een complexer beheersingsniveau meten en lezers over een hogere leesvaardigheid beschikken (CCSSI, 2010). Immers, naarmate teksten complexer, abstracter en gelaagder worden, neemt de persoonlijke interpretatie-

ruimte toe en naarmate lezers vaardiger worden, zullen ze deze interpretatieruimte vaker op adequate wijze benutten. Daarmee kan situatiemodelbegrip vaak botsen met de eenduidige betekenisstoekenning aan een tekst die meerkeuzeopgaven veronderstellen. Daardoor kan een meerkeuzeopgave voor lezers niet alleen een talige puzzel lijken, hun oplossing van de puzzel is persoonlijker dan het gesloten, eenduidige format van een meerkeuzeopgave veronderstelt.

Onderzoekshypothesen
Deze studie heeft als doel te verkennen in hoeverre een gebrekkige cognitieve validiteit een verklaring kan vormen voor de beperkte meerwaarde van meerkeuzeopgaven in de CE's Nederlands havo en vwo. We onderzoeken daarbij gericht drie hypothesen, gerelateerd aan de drie deelaspecten van cognitieve validiteit in Farr et al. (1990):

- a. *gebrekkige-diepgangshypothese*: meerkeuzeopgaven stimuleren in vergelijking met open opgaven vooral leesgedrag dat zich richt op de oppervlaktebetekenis van tekstdelen, met een weinig diepgaand tekstbegrip als gevolg;
- b. *betekenisreductiehypothese*: meerkeuzeopgaven doen geen recht aan het situatiemodelbegrip van lezers omdat ze slechts één mogelijke interpretatie van een tekst(gedeelte) presenteren;
- c. *testmanagementhypothese*: meerkeuzeopgaven roepen specifieke testmanagementstrategieën op waarbij leerlingen zich niet richten op begripsopbouw van een tekst-deel om de best passende meerkeuzeopties te kiezen, maar waarbij ze vooral meerkeuzeopties afstrepen en tegenover elkaar afwegen met behulp van tekstraadpleging, logica en voorkennis.

Hypothese a onderzoeken we met een scoreformulier waarmee we het lees- en antwoordgedrag van leerlingen bij meerkeuze- en open opgaven in kaart brengen. Hypothesen b

en c onderzoeken we door een analyse van hardop-denkprotocollen van het antwoordgedrag van leerlingen bij meerkeuzeopgaven, waarbij we telkens meenemen in hoeverre het optreden van b en c tot de keuze van een onjuist antwoord leidt.

Methode

Algemene opzet en deelnemers
Om zicht te krijgen op het leerlinggedrag bij meerkeuzeopgaven, hebben we eindexamenleerlingen in deze studie hardop laten reflecteren op hun denkproces tijdens het beantwoorden van CE-taken. Daarbij maakten we gebruik van twee hardop-denkinstrumenten.

Om zicht te krijgen op de diepgang van het tekstbegrip bij een opgave (de *gebrekkige-diepgangshypothese*), interviewden we leerlingen bij elke opgave kort, telkens onmiddellijk nadat ze de vraagstelling van de opgave hadden gelezen, maar vóórdat ze zich tot de eigenlijke beantwoording van de vraag zetten. Dit mini-interview verrichten we bij elke open opgave en bij elke meerkeuzeopgave.

Om ook de andere twee hypothesen aan te kunnen tonen (de *betekenisreductiehypothese* en *test-managementhypothese*), lieten we leerlingen elke opgave, na het mini-interview, naar eigen ingeving beantwoorden, zonder verdere inhoudelijke sturing van de onderzoekers. Hiervan maakten we audio-opnamen. De hardop-denkprotocollen van hun antwoordgedrag bij de meerkeuzeopgaven hebben we vervolgens geanalyseerd. In tabel 2 is de inzet van de twee instrumenten in onze studie weergegeven.

Het onderzoek is afgenomen op zes middelbare scholen onder havo 5- (n = 12) en vwo 6-leerlingen (n = 13). Per school werden in overleg met docenten Nederlands vier leerlingen geselecteerd, waarbij we docenten verzochten leerlingen te selecteren die samen een representatieve dwarsdoorsnede vorm-

den van de leesvaardigheid in hun examenklassen. Uiteindelijk nam op één school nog een vijfde vwo 6-leerling deel. Het onderzoek duurde doorgaans één klokuur.

Materiaal
Als materiaal werden twee sets teksten met begripsopgaven geselecteerd uit oude CE's Nederlands havo en vwo. Daarbij golden drie selectiecriteria:

- a. de teksten zijn niet te kort of te lang (circa 900 woorden);
- b. de sets bestaan vooral uit inhoudelijke begripsopgaven en bevatten geen analytisch-functionele opgaven, bijvoorbeeld over alineafunctionalies;
- c. de sets bestaan voor een belangrijk deel uit meerkeuzeopgaven.

Bij 5-havo werd gekozen voor de set 'Ga toch eens een weekendje spelen' (CE Nederlands havo 2022-1) en bij 6-vwo voor de set 'Alles denken en voelen' (CE Nederlands vwo 2023-2).

Per set werden er in totaal zes opgaven geselecteerd: drie meerkeuze- en drie open opgaven. Van elke opgave ontwikkelde een onderzoeker vervolgens één meerkeuzevariant en één open variant. Dit materiaal werd daarop in twee rondes gescreend door drie andere onderzoekers en per ronde bijgesteld.

Met de ontwikkelde opgavenvarianten zijn uiteindelijk twee versies van beide sets samengesteld, waarbij elke versie steeds drie open opgaven en drie meerkeuzeopgaven bevatte: sommige in de oorspronkelijke en andere in de bewerkte vorm. Open opgaven

en meerkeuzeopgaven wisselden daarbij in elk opgavenboekje steeds af.

Werkwijze
Bij het onderzoek waren steeds twee uit een totaal van vier onderzoekers aanwezig, allen werkzaam als toetsdeskundigen Nederlands bij Stichting Cito. Eén onderzoeker nam het mini-interview af bij elke opgave. Beide onderzoekers vulden tijdens het gehele onderzoek per opgave een kort scoreformulier in over het denkproces van de leerlingen, tijdens het mini-interview en het hardop-denkgedeelte; de tweede onderzoeker maakte daarbij aanvullende aantekeningen.

Na een welkom en uitleg over doel en werkwijze van het onderzoek konden leerlingen kort oefenen met de werkwijze. Daarbij bleef onbenoemd dat de onderzoekers vooral geïnteresseerd waren in het antwoordgedrag van leerlingen bij meerkeuzeopgaven. Vervolgens kregen leerlingen een tekst met begripsopgaven uit een CE Nederlands havo dan wel vwo voorgelegd, waarbij ze de tekst eerst in stilte moesten lezen; daarna beantwoordden ze hardop denkend zes opgaven over deze tekst. Hier zijn audio-opnames van gemaakt.

Omdat we het antwoordgedrag bij open en meerkeuzeopgaven wilden kunnen vergelijken voor de gebrekkige-diepgangshypothese, kregen leerlingen steeds drie meerkeuze- en drie open opgaven voorgelegd. Per opgave werd leerlingen direct na het lezen van de vraagstelling, en dus nog voor het lezen van

Hypothese	ingezet analyse-instrument	onderzocht opgavetype
a) <i>gebrekkige-diepgangshypothese</i>	scoreformulier	meerkeuze- en open opgaven
b) <i>betekenisreductiehypothese</i>	hardop-denkprotocollen	meerkeuzeopgaven
c) <i>testmanagementhypothese</i>	hardop-denkprotocollen	meerkeuzeopgaven

Tabel 2. Ingezette analyse-instrumenten en onderzochte opgavetypen in het onderzoek, per onderzoekshypothese

de meerkeuze-opties, in een mini-interview gevraagd of ze:

1. de vraagstelling begrepen;
2. direct na het lezen van de vraagstelling al een antwoord in hun hoofd hadden;
3. na het lezen van de vraagstelling wisten waar ze het antwoord in de tekst moesten vinden.

Daarna startte het hardop-denkdeel van het onderzoek: leerlingen zetten zich hardop denkend tot het beantwoorden van de opgave. Ze lazen bij de meerkeuzeopgaven de meerkeuzeopties, raadpleegden indien nodig de tekst, herlazen de vraagstelling van de opgave, et cetera. Indien het hardop denken van leerlingen onvoldoende duidelijkheid gaf over de bedoeling achter hun antwoordgedrag, vroeg de onderzoeker om verduidelijking. Na afronding van het onderzoek ontving elke deelnemer een cadeaubon.

Gebrekkige-diepgangshypothese: observatiemiddelen en data-analyse

Om eventuele verschillen in diepgang in het antwoordproces aan te kunnen tonen tussen meerkeuzeopgaven en open opgaven (de *gebrekkige-diepgangshypothese*), maakten we allereerst per opgave gebruik van een scoreformulier uit Rooijackers (2023), met vierpunts-Likert-schalen. Met dit scoreformulier brachten we voor beide soorten opgaven de kwaliteit in kaart van het begripsproces van leerlingen tijdens het beantwoorden van de opgaven.

Allereerst werd het begrip van de vraagstelling bij de leerling gescoord, evenals het tekstbegrip dat de leerling onmiddellijk ná het lezen van de vraagstelling rapporteerde in het mini-interview. We brachten drie zaken in kaart:

1. In hoeverre kon de leerling de vraagstelling op een juiste wijze parafraseren?
2. In hoeverre kon de leerling, zonder nadere tekstraadpleging, al een antwoord formuleren op de vraag?

3. In hoeverre bleek de leerling al te weten waar het antwoord in de tekst was te vinden?

Vervolgens werd ook het antwoordproces gescoord tijdens het hardop-denkgedeelte, het antwoordproces zoals dat uiteindelijk uitmondde in het formuleren van het antwoord. We brachten daarbij drie zaken in kaart:

4. In hoeverre bleek de leerling in staat de tekstpassage te lokaliseren waar het antwoord was te vinden?
5. In hoeverre bleek uit het hardop denken van de leerling tijdens het antwoordproces vooral tekstconform situatiemodelbegrip of vooral tekstbasisbegrip? Met andere woorden, in hoeverre wist de leerling het gelezene in eigen woorden treffend weer te geven (situatiemodelbegrip) of herhaalde de leerling juist vooral de bewoordingen van de tekst (tekstbasisbegrip)?
6. Raadpleegde de leerling kort of juist zeer uitgebreid de tekst?

Voor deze laatste drie observatiepunten geven inzicht in de diepgang van tekstverwerking door leerlingen.

Tijdens het onderzoek noteerden beide onderzoekers hun scores op de Likert-schalen; direct na afloop van elke sessie vergeleken zij deze scores. Eventuele verschillen werden besproken, waarna een gezamenlijke score werd bepaald.

Betekenisreductie- en testmanagementhypothese: observatiemiddelen en data-analyse

Om de twee andere hypothesen te toetsen, analyseerden we de hardop-denktopprotocollen bij de meerkeuzeopgaven, zodat het tekstbegrip van leerlingen, hun ingezette antwoordstrategieën en hun rationale achter het antwoordgedrag konden worden meegenomen. Het antwoordgedrag bij open opgaven werd buiten beschouwing gelaten, omdat de *betekenisreductie- en testmanagementhypothese* hierbij niet waarneembaar zijn.

Beide onderzoekers maakten tijdens het

onderzoek een aantekening als ze in het leerlingcommentaar bij een meerkeuzeopgave antwoordgedrag waarnamen dat relevant was voor een van beide hypothesen. Voor de *betekenisreductie-hypothese*: namen de onderzoekers een botsing waar tussen het situatiemodelbegrip van de leerling en de tekstinterpretatie door de toetsmakers in de meerkeuzeopties, waardoor de sleutel mogelijk werd afgewezen? Voor de *testmanagement-hypothese*: liet de leerling in de ogen van de onderzoekers de keuze voor het goede dan wel foute antwoord vooral bepalen door het spel van verificatie en falsificatie van de meerkeuzeopties (oftewel testmanagementstrategieën), waardoor de sleutel op basis van selectieve tekstraadpleging, logica en/of voorkennis, werd afgewezen of een afleider zeer serieus werd overwogen? Na afloop van elke sessie besprak de observerende onderzoeker het leerlingcommentaar dat in diens ogen relevant was voor een van beide hypothesen. Indien de andere onderzoeker op basis van eigen waarnemingen de relevantie hiervan bevestigde, werd leerlingcommentaar gemarkeerd als 'relevant'. Voor de verwerking van de resultaten werd dit relevante antwoordgedrag in de weken na het onderzoek door een onderzoeker teruggeluisterd en uitgeschreven. Daarna werd het door twee onderzoekers geanalyseerd en verwerkt.

Resultaten

Gebrekkige-diepgangshypothese: resultaten scoreformulier

Er werden in totaal 150 opgaven gescoord via het scoreformulier (open opgaven: 75; meerkeuzeopgaven: 75). Zien we bij de beantwoording van meerkeuzeopgaven aanwijzingen voor oppervlakkiger tekstbegrip dan bij open opgaven (*gebrekkige-diepgangshypothese*)?

Voor de verwerking van de resultaten

gebruikten we variantieanalyse (ANOVA), waarbij opgavetype telkens als vast effect is meegenomen.

De gescoorde scoreformulieren leverden geen significante verschillen op tussen het antwoordgedrag bij meerkeuzeopgaven en dat bij open opgaven² (zie tabel 3, p. 22). We kunnen dus niet aantonen dat meerkeuzeopgaven aanzetten tot een relatief oppervlakkig tekstbegrip in vergelijking met open opgaven (*gebrekkige-diepgangshypothese*).

Betekenisreductie- en testmanagement-hypothese: resultaten hardop-denktopprotocollen

In totaal is tijdens het onderzoek 75 maal het antwoordproces bij een meerkeuzeopgave geobserveerd (havo: 36 maal; vwo: 39 maal). De meest opvallende resultaten komen we tegen bij vwo-leerlingen: elf keer gaven zeven verschillende vwo-leerlingen blijk van een hardop-denktopproces dat de *betekenisreductiehypothese* of de *testmanagementhypothese* onderbouwt. In tien van deze elf gevallen kozen leerlingen een afleider als antwoord. Bij de havo-leerlingen daarentegen werden minder relevante verschijnselen geobserveerd: slechts één havo-leerling gaf bij drie meerkeuzeopgaven blijk van een relevant hardop-denktopproces. In twee van deze drie gevallen koos de leerling een afleider als antwoord.

Vanwege de beknoptheid zullen we hier enkel een synthese geven van de resultaten, waarbij we via enkele kenmerkende voorbeelden ingaan op de vaak complexe interactie tussen leerlingcommentaar, opgave en brontekst. Voor een volledige analyse verwijzen we naar onze onderzoeksrapportage (Rooijackers & Stevens, 2025). Voor een goede navolgbaarheid van onderstaande bespreking is bovendien gedetailleerde kennis van de besproken teksten en opgaven een pre; op Examenblad.nl zijn de betreffende onderdelen van de havo- en vwo-CE's eenvoudig raadpleegbaar.

De betekenisreductiehypothese
Bij betekenisreductie is de sleutel slechts een mogelijke interpretatie die botst met de interpretatie van de leerling. Zeven keer gaven vier vwo-leerlingen van hardop-denkgedrag blijk waarmee de betekenisreductiehypothese kan worden onderbouwd. Bij de havo-leerling trad dit tweemaal op.

Over het algemeen geldt: een klein detail in de formulering van de sleutel bleek een (onverwacht) grote impact te kunnen hebben op het antwoordproces van leerlingen. Het patroon was hierbij bij vrijwel alle leerlingen identiek. De examenmakers hebben een sleutel opgesteld die zij niet zozeer op tekstbasis-

niveau formuleren (namelijk: door letterlijke tekstpassages over te nemen), maar vooral op situatiemodelniveau: ze parafraseren tekstpassages en voegen daaraan logische inferenties toe; impliciete denkstappen op basis van hun eigen situatiemodelbegrip. Echter, doordat leerlingen hun eigen accent leggen in de gelezen tekst, blijken juist deze inferenties van de examenmakers voor leerlingen reden om de sleutel als fout te beschouwen, terwijl leerlingen wel de tekst afdoende leken te begrijpen, zoals bleek uit hun hardop denken.

We geven een drietal voorbeelden. Meerkeuzeopgave 3 in de vwo-set (oorspron-

kelijk opgave 15 in CSE 2023-2) vraagt waarom de filosoof McLuhan als een voorspeller van het internet geldt. De sleutel is: 'McLuhan beschouwde elektronische media als een verbetering van ons brein, wat bij uitstek een kenmerk is van internet', waarbij de woordgroep 'bij uitstek' een inferentie van de toetsmakers is. Op situatiemodelniveau lijkt deze sleutel inderdaad een correcte samenvatting van alinea 7: de tekst stelt namelijk expliciet als 'belangrijkste punt' dat 'elektronische media een verbetering van ons brein [zijn]', en, zo vullen de toetsmakers aan, dat geldt bij uitstek voor het internet. Die gedachtegang is in lijn met alinea 7, en bovendien wordt vanaf alinea 8 benadrukt dat het internet een nieuw menselijk 'superbrein' voortbrengt: 'bij uitstek' lijkt dus een logische inferentie. Echter, binnen alinea 7 is de denkstap 'bij uitstek' geen noodzakelijke inferentie. En dat plaatst proefpersoon 16 voor problemen:

Dan ga ik nog een keer terug naar A. 'beschouwde elektronische media als een verbetering van ons brein, wat bij uitstek een kenmerk is van het internet'. Op zich, tot aan die komma lijkt dat een best goed antwoord, maar dat dat bij uitstek [met nadruk] een kenmerk is van het internet, lijkt me ook niet echt terug te komen in alinea 7.

Proefpersoon 24 signaleert hetzelfde probleem: 'Ja, volgens mij is het volgens de tekst niet 'bij uitstek' een kenmerk van internet'. Beide leerlingen lijken de tekstpassage afdoende te begrijpen, maar kiezen om deze reden niet de sleutel, maar een afleider die raakt aan de informatie in de tekst maar voorbijgaat aan 'het belangrijkste punt' in de alinea. De interpretatie van de toetsmakers botst hier dus met die van deze leerlingen.

Een vergelijkbaar proces zien we bij meerkeuzeopgave 2 in de vwo-set (oorspronkelijk opgave 14 in CSE 2023-2). Op situatiemodelniveau problematiseert de sleutel de bewering in de tekst dat de filosoof McLuhan

'ook wel' als een voorspeller van het internet wordt gezien. Alinea 2 geeft daarvoor enkel als argument dat de filosoof McLuhan in de jaren zestig een krachtige typering 'van elektronische media' gaf, 'van radio en tv'. De examenmakers formuleren daarom als kritiek dat McLuhan niet het internet voorspelt, want internet behoort niet tot de elektronische media, zo impliceert de tekst. Maar precies die laatste denkstap problematiseert proefpersoon 21:

Onderzoeker: Waarom twijfel je nu (...) over D? Bij D zit ik na te denken: elektronische media, hoort dat bij [internet] ... Internet hoort er niet bij, maar radio en tv ja, op zich in principe wel ... maar op radio en tv komen wel dingen voorbij die je anders ook op internet zou kunnen vinden ... dus toch een beetje wel ... ja, wel ...

Uiteindelijk kiest ook deze leerling daarom een afleider. Ook hier vormt een logische maar niet noodzakelijke denkstap vanuit de toetsmakers dus een reden om niet de sleutel maar een afleider te kiezen die de centrale gedachtegang nauwelijks recht doet.

Niet-noodzakelijke inferenties vanuit de leerling zelf kunnen evenzeer een probleem creëren bij het identificeren van de sleutel. Proefpersoon 5, de havo-leerling, parafraseert bij meerkeuzeopgave 2 uit de havo-set (oorspronkelijk opgave 11 in CSE 2022-1) in eerste instantie de hoofdgedachte van alinea 4 zeer accuraat en komt dan verraderlijk dicht bij de formulering van de sleutel. ('Dus, het nut [van spelen] is vooral dat het nutteloos is, dus dat het geen zin heeft'. Vergelijk de sleutel A: 'Het nut van sport en spel schuilt vooral in het nutteloze en vrijblijvende ervan.'). Maar zodra ze de meerkeuzeopties leest, valt sleutel A snel voor haar af. Het moet volgens haar over maatschappelijk nut gaan, wat een verdedigbare maar niet noodzakelijke inferentie is bij alinea 4. En sleutel A bevat dit maatschappelijke element

vraag in scoreformulier	schaal	meerkeuze gem. (sd)	open gem. (sd)	p
<i>Begrip direct na lezen vraagstelling</i>				
1. In welke mate begrijpt de leerling de vraagstelling van de opgave?	correct: 1 incorrect: 4	1,76 (0,91)	1,65 (0,88)	0,45
2. In welke mate heeft de leerling direct na het lezen van de vraagstelling al een antwoord in het hoofd?	aanwezig: 1 afwezig: 4	3,44 (0,90)	3,14 (0,13)	0,07
3. In welke mate weet de leerling na het lezen van de vraagstelling al waar het antwoord is te vinden in de tekst?	nauwkeurig: 1 afwezig: 4	2,74 (0,93)	2,68 (1,04)	0,81
<i>Antwoordproces na lezen vraagstelling</i>				
4. In welke mate kan de leerling na het lezen van de vraagstelling direct het antwoord in de tekst lokaliseren?	vrijwel direct: 1 niet of met veel moeite: 4	2,19 (0,91)	2,03 (0,94)	0,30
5. Van welk begripsniveau geeft de leerling vooral blij: situatiemodelbegrip of vooral tekstbasisbegrip?	situatiemodel: 1 tekstbasis: 4	2,31 (0,65)	2,27 (0,73)	0,76
6. Hoe intensief raadpleegt de leerling de tekst?	niet: 1 omstandig: 4	2,44 (0,72)	2,47 (0,65)	0,77

Tabel 3. Gemiddelde antwoordscore en standaarddeviatie (sd) per opgave in scoreformulier, per opgavetype

niet, terwijl afleider B dat wel doet:

Nee. [Leest A nog eens.] Ja ik zie vooral dat het gaat echt over die maatschappij, wat die filosoof zegt, dus daarom zou ik A afstrepen. Het komt dichtbij, maar ik vind B beter passend. A is denk ik niet per se fout, maar het is niet het volledige antwoord. Dus ik zou gaan voor B. [Schrijft B op het antwoordblad.]

Hier maakt de leerling op basis van het eigen situatiemodelbegrip dus zelf een logische maar niet noodzakelijke inferentie bij de tekst, die de toetsmakers in de sleutel niet hebben gevolgd. Dat antwoord B om andere inhoudelijke redenen de sleutel niet kan zijn, doet er in de redenering van de leerling niet toe: het element dat in de sleutel naar haar idee ontbrak, vindt zij wel terug in de afleider.

De eenduidige betekenistoekenning aan teksten die meerkeuzeopgaven veronderstellen, botst dus in deze gevallen met de eigen interpretatieruimte van leerlingen, vooral van vwo-leerlingen. We zien deze botsing op twee manieren ontstaan. Ten eerste: doordat de examenmakers aan de sleutel denkstapen toevoegen die logisch maar niet noodzakelijk zijn, wijzen leerlingen de sleutel af omdat zij deze inferentie niet noodzakelijk vinden. Ten tweede: doordat leerlingen bij het lezen van de tekstpassage ook zelf logische maar niet noodzakelijke inferenties kunnen opbouwen, wijzen ze de sleutel af omdat deze niet aan hun eigen inferentie recht doet. Daardoor kiezen de leerlingen in deze gevallen dus doorgaans een afleider, ondanks vaak afdoende begrip van de relevante tekstpassage.

Testmanagementhypothese

Bij testmanagement gaan leerlingen uit van de gegeven opties en selecteren zij inhoudselementen uit de tekst om te bepalen welke de goede optie is. Vier keer gaven drie vwo-leerlingen van hardop-denkgedrag blijk waar-

mee de *testmanagementhypothese* kan worden onderbouwd. De havoleerling gaf er eenmaal van blijk. In drie gevallen kiezen leerlingen daarbij uiteindelijk het onjuiste antwoord.

Bij deze vier gevallen zagen we een antwoordpatroon conform deze hypothese. Leerlingen zijn niet bezig de meerkeuzeoptie te kiezen die het best aansluit bij hun eigen tekstbegrip van de passage die in de opgave bevestigd wordt. De deelnemers richten zich op het bevestigen of weerleggen van de meerkeuzeopties, met behulp van gerichte, selectieve tekstraadpleging. In eerste instantie onjuiste weergaven van de tekstinhoud werden in deze gevallen daardoor gaandeweg verdedigbare interpretaties, omdat ze met tekstelementen als 'logisch' konden worden onderbouwd.

We werken één voorbeeld uitgebreid uit en één voorbeeld beknopt. Meerkeuzeopgave 3 in de vwo-set (oorspronkelijk opgave 14 in CSE 2023-2) bevat één afleider die een tekstueel onjuiste interpretatie geeft van de gedachtegang van alinea 7: de filosoof McLuhan zou 'een ontwikkeling' zien in de geschiedenis van 'menselijke hulpmiddelen' en van daaruit zou voor hem het internet 'logisch volgen'. Er is geen enkele expliciete verwoording van deze gedachtegang in de alinea, en ook de veronderstelde chronologische ontwikkeling in deze afleider wordt in de tekst niet expliciet uitgewerkt. Bij eerste lezing van de meerkeuzeoptie wijzen proefpersonen 16 en 24 deze afleider dan ook af (proefpersoon 16: 'onwaarschijnlijk').

Maar na uitgebreid de tekst geraadpleegd te hebben, komen beide vwo-leerlingen toch weer bij deze afleider uit. Er kan namelijk een interpretatie worden verdedigd waarbij antwoordoptie D wél als het goede antwoord is te beschouwen. Al stelt de tekst het niet expliciet, uit de voorbeelden van menselijke hulpmiddelen van McLuhan kan wel een ontwikkeling worden afgeleid die in het internet uitmondt. Proefpersoon 24 stelt daarom na

tamelijk nauwkeurige tekstraadpleging dat hij 'uiteindelijk' (!) afleider D kiest:

Want ze [de tekst en antwoordoptie D] zeggen beide precies hetzelfde: het zijn hulpmiddelen en als je gaat kijken naar de verbeteringen die het allemaal heeft gebracht, dan kom je uiteindelijk uit bij het internet. Ik zou daarom uiteindelijk zeggen D.

Ook proefpersoon 16 kiest na uitgebreide tekstraadpleging afleider D ('op zich wel logisch'). Doordat deze leerlingen de meerkeuzeopties als stellingen onderbouwen of weerleggen met tekstelementen, wordt deze afleider, die in strikte zin de tekstinhoud geen recht doet, een aanvaardbaar antwoord.

Een ander voorbeeld is vwo-meerkeuzeopgave 4 in de vwo-set (oorspronkelijk opgave 16 in CSE 2023-2), waarin afleider D stelt dat we dankzij het metaverse allemaal 'in een digitale maatschappij zullen gaan leven'. Deze interpretatie van alinea 6 wordt nergens expliciet gemaakt, terwijl sleutel A ('beduidend minder reëel contact met anderen hebben') apert juist is. Alinea 6 benadrukt namelijk dat we door het metaverse 'minder de deur' uitgaan, dat gasten voortaan 'virtueel' aan tafel zitten, et cetera. Proefpersoon 20 stelt na eerste lezing dan ook meteen vast dat A het meest logische antwoord is. Echter, na tekstraadpleging kiest deze toch voor afleider D. Er is namelijk een causale keten te construeren, die tot afleider D leidt:

A klopt, eerder werd gezegd dat ... (kijkt in tekst) D, in een digitale maatschappij leven, dat is eigenlijk wel wat er wordt gezegd ... want je bent een projectie in elkaars leven, dus ... je leeft online met elkaar, dus ik ben bij antwoord D. Die wordt het ...

De leerling kiest hier dus voor een interpretatie die de tekst nergens expliciet maakt, maar die vanuit tekstelementen wel kan

worden afgeleid.

Hier zien we dus twee voorbeelden van specifieke testmanagementstrategieën van meerkeuzeopgaven. Een meerkeuzeopgave nodigt leerlingen niet enkel ertoe uit op de juiste wijze een tekst(deel) te begrijpen, maar stuurt door meerkeuzeopties als 'stellingen' aan te reiken in sterke mate de begripsopbouw. Daarmee kunnen uitspraken die in strikte zin niet accuraat de gedachtegang van een tekstpassage weergeven, alsnog zeer verleidelijke afleiders worden.

Discussie

Kanttekeningen en discussie

Vooropgesteld: er zijn de nodige kanttekeningen te plaatsen bij dit onderzoek. Zo is het aantal geobserveerde items beperkt. De steekproef is daarnaast relatief klein, wat het ontbreken van resultaten bij vooral de scoreformulieren kan verklaren. Daarbij is de vraag of de inzet van een meer fijnmazige onderzoeksmethode dan het scoren van scoreformulieren (bijvoorbeeld eye-trackonderzoek) wél resultaten had kunnen blootleggen, zoals het geval was in Pearson et al. (1999) en Roelofs et al. (2026). Bovendien hebben we ons in dit onderzoek maar op één mogelijke verklaring gericht voor de geringe meerwaarde van meerkeuzeopgaven in de CE's: een gebrekkige cognitieve validiteit. In hoeverre andere verklaringen, en zo ja, welke verklaringen hierbij een rol spelen, blijft onbelicht.

We lopen kort de drie onderzoekshypotheseën langs. Allereerst de *gebrekkige-diepgangshypothese*: we konden niet aantonen dat leerlingen de meerkeuzeopgaven oppervlakkiger beantwoorden dan open opgaven. Hierin dient echter verdisconteerd dat we leerlingen expliciet uitnodigden om stapsgewijs en nauwkeurig na te denken over hun antwoordgedrag. Dat kan een uitvoeriger, gelaagder antwoordproces bij meerkeuzeopgaven

hebben gestimuleerd. Daarnaast zou met name bij deze hypothese een fijnmazigere onderzoeksmethode wel degelijk resultaten kunnen opleveren.

De tweede hypothese, de *betekenisreductiehypothese*, konden we wel aantonen. De eenduidige betekenis-toekenning aan teksten die meerkeuzeopgaven veronderstellen, botst in een behoorlijk aantal gevallen met de eigen interpretatieruimte van leerlingen, vooral van vwo-leerlingen. Situatiemodelbegrip, zowel dat van de toetsmakers als dat van de leerlingen zelf, lijkt bij enkele meerkeuzeopgaven een belangrijke stoorzender om tot de keuze van de sleutel te komen. Ondanks dat leerlingen vaak een voldoende accurate weergave van de tekstinhoud leken op te bouwen, leidde deze stoorzender ertoe dat leerlingen in het merendeel van deze gevallen de sleutel afwezen.

Ook de derde hypothese, de *testmanagementhypothese*, konden we aantonen, zij het in een beperkt aantal gevallen. In een beperkt aantal gevallen wezen leerlingen eerst nog op basis van hun eigen tekstbegrip een afleider af die de gedachtegang in de tekst niet correct samenvatte, om diezelfde afleider later in het antwoordproces te verdedigen als een correcte interpretatie.

Conclusie en aanbevelingen

Al met al lijken deze resultaten de stelling te onderbouwen dat de beperkte statistische meerwaarde van meerkeuzeopgaven in de CE's Nederlands havo en vwo in elk geval deels kan worden verklaard door een gebrekkige cognitieve validiteit. Het format van meerkeuzeopgaven stimuleert bij lezers vaak antwoordgedrag dat zeer specifiek op het meerkeuzeformat is toegesneden. Meerkeuzeopgaven nodigen lezers vaak niet uit om zelfstandig een mentale representatie van de tekstinhoud op te bouwen, terwijl

dat wel in de ideale leestoets centraal zou moeten staan (bijvoorbeeld Magliano et al., 2018). In die zin lijken de meerkeuzeopgaven inderdaad vaak meer op een talige puzzel die lezers moeten oplossen, dan op een begripstaak (Leslie & Caldwell, 2017). Vooral de vwo 6-leerlingen kiezen op basis van dit antwoordgedrag uiteindelijk een ander antwoord dan de sleutel.

Opvallend is dat vooral het antwoordgedrag van de vwo-leerlingen relevante informatie opleverde. Mogelijk komt dit doordat vwo 6-leerlingen hun antwoordproces vaak scherper kunnen toelichten dan havo 5-leerlingen; naarmate lezers vaardiger worden, kunnen ze hun leesgedrag immers ook nauwkeuriger uitleggen (McNamara & Magliano, 2009). Echter, doordat de cognitieve-validiteitsproblemen vooral lijken te ontstaan rondom situatiemodelbegrip, lijkt het vooral een aanwijzing dat meerkeuzeopgaven meer cognitieve validiteitsproblemen opleveren naarmate het getoetste leesvaardigheidsniveau hoger is en leerlingen over een hoge leesvaardigheid beschikken (CCSSI, 2020).

Wat betekenen deze conclusies nu voor de meerkeuzeopgaven, vooral in het CE Nederlands vwo (en mogelijk havo)? Er zijn twee aanbevelingen te geven. Ten eerste: de sleutel in een meerkeuzeopgave moet geen elementen bevatten die kunnen afwijken van de persoonlijke, tekstconforme interpretatie van lezers. Alleen noodzakelijke inferenties kunnen in de sleutel worden geëxpliciteerd; logische maar niet noodzakelijke inferenties bij een tekstpassage moeten worden vermeden (zoals de woorden 'bij uitstek' in vwo-opgave 3). Ten tweede: de afleiders in een meerkeuzeopgaven moeten geen beweringen bevatten die goede lezers met de tekst overtuigend kunnen beargumenteren, ook al worden ze in strikte zin niet door de tekst gesteld. Afleiders dienen apert onjuiste beweringen of onjuiste inferenties bij de tekstpassage te zijn.³

Echter, het systematisch doorvoeren van deze aanbevelingen is lastig. Zeker in het vwo-examen moeten toetsmakers meerkeuzeopgaven vanuit situatiemodelbegrip kunnen formuleren om aan het complexe leesvaardigheidsniveau 4F recht te doen; de vraag of een inferentie bijvoorbeeld niet alleen logisch maar ook noodzakelijk is, is daarbij niet altijd eenvoudig te beantwoorden. Aanvullend hardop-denkenonderzoek onder leerlingen zou hierbij behulpzaam kunnen zijn, maar dergelijk onderzoek is lastig in te bouwen in de complexe constructieprocedure van deze CE's en is kosten- en tijdsintensief.

Met de vraagvernieuwing in de CE's Nederlands havo en vwo vanaf 2024 wordt expliciet beoogd meer ruimte te geven aan situatiemodelbegrip, voor havo (niveau 3F) en voor vwo (niveau 4F) (Evers-Vermeul et al., 2022). De uitkomsten van dit onderzoek werpen de vraag op of meerkeuzeopgaven een bruikbaar middel zijn om dit doel te realiseren. Dat geldt vooral voor de CE's Nederlands vwo, waar situatiemodelbegrip, meer dan in het havo-examen, in het leeuwendeel van de opgaven een rol speelt.

Hoe kan het bestaande examen dan de valkuilen van de bestaande meerkeuzeopgave vermijden? Het opnemen van méér open opgaven lijkt niet wenselijk gezien de hoge correctielast van het bestaande examen. Het ligt daarom – als eerste mogelijkheid – voor de hand méér semigesloten opgaven op te nemen zoals citeeropgaven en stellingenopgaven. Deze opgaven kunnen beoordelaars immers snel nakijken en ze functioneren statistisch vaak beter dan meerkeuzeopgaven (zie tabel 1).

Een tweede mogelijkheid is een gedeeltelijk andere inrichting van meerkeuzeopgaven in het CE. Enkele Angelsaksische L2-toetsen (Educational Testing Service, 2025) werken met zogenaamde *negative factual information*-meerkeuzeopgaven: leerlingen kiezen hierbij de antwoordoptie die als enige onjuist is.

Daarmee zou de problematiek van situatiemodelbegrip als stoorzender kunnen worden ondervangen: je onderkent immers dat er meerdere interpretaties van een tekstpassage mogelijk zijn en je vraagt leerlingen nu te zien welke interpretatie niet strookt met de tekstpassage. Vervolgonderzoek hiernaar is wenselijk.⁴

NOTEN

1. Semigesloten opgaven zijn opgaven die geen meerkeuzeopgaven zijn maar waarop de antwoordmogelijkheden zeer beperkt zijn. Denk aan citeeropgaven of sorteer-taken waarbij leerlingen bijvoorbeeld een aantal gegeven beweringen in de juiste causale volgorde moeten plaatsen.
2. We rapporteren hier enkel op basis van variantie-analyse, waaruit geen significante resultaten volgen. We zijn ons ervan bewust dat vanwege de genestheid van de data multiniveau-analyse in ons onderzoek de voorkeur verdient. Echter, de standaarden voor het vinden van significante data liggen bij multiniveau-analyse nog beduidend hoger dan bij variantie-analyse. Omwille van arbeidsintensiviteit (én de bondigheid van dit artikel) is daarom ervan afgezien een nadere multiniveau-analyse door te voeren.
3. Bezien vanuit het specifieke antwoordgedrag van meerkeuzeopgaven (!) lijken sommige afleiders in meerkeuzeopgaven in het onderzoeksmateriaal nu suboptimaal gekozen. Zo was afleider D in vwo-meerkeuzeopgave 4 een te goed verdedigbare interpretatie voor leerlingen; ze parafraseert weliswaar onvoldoende de centrale gedachtegang, ze reikt een vervolgstap erbij aan die in lijn met de tekst lijkt.
4. In veel handboeken wordt aangeraden dergelijk negatief geformuleerde meerkeuzeopgaven te vermijden (Alderson, 2000). De vraag is echter of binnen de context van een toets die vooral complex situa-

tiemodelbegrip meet, dit vraagtype juist niet goed zou kunnen functioneren. De opname van dit type opgaven in bijvoorbeeld de TOEFL-iBT (Educational Testing Service, 2025) – een toets waar veel onderzoek aan ten grondslag ligt – lijkt hiervoor te pleiten.

LITERATUUR

- Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press
- Brunfaut, T. (2021). *Assessing reading*. In G. Fulcher & L. Harding (Eds.), *The Routledge handbook of language testing* (2nd ed., pp. 254–267). Routledge.
- Bussis, A. M., & Chittenden, E. A. (1987). Research currents: What the reading tests neglect. *Language Arts*, 64(3), 302–308.
- CCSSI (Common Core State Standards Initiative). (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. <https://corestandards.org>
- Cohen, A. D., & Upton, T. A. (2006). Strategies in responding to the new TOEFL reading tasks. *ETS Research Report Series*, 2006(1), i–162.
- Cordón, L. A., & Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology*, 88(2), 288–295.
- CvTE. (2025). *Terugblik 2024; Centraal Examen Nederlands Eerste Tijdvak*. CvTE. https://examenblad.nl/system/files/2025-06/terugblik-nederlands-2024-vwo-vi_def.pdf
- Educational Testing Service (2025). *TOEFL iBT® test*. <https://www.ets.org/toefl>
- Essen, A., Ploeg, S. van der, & Vermeulen, J. (2023). *Correctielast centrale examens vo; indeling van centrale examens vmbo-gl/tl, havo en vwo naar correctie-intensiteit*. CAOP & Oberon. <https://cvte.nl/documenten/rapporten/2023/10/09/onderzoek-correctielast-centrale-examens-vo>
- Evers-Vermeul, J., Heesters, K., Schuurs, U.,

- & Rooijackers, P. (2022). *Opbrengsten van het project 'Uitbreiding van vraagtypen voor het leesvaardigheidsexamen Nederlands havo en vwo'*. CvTE/Cito. https://www.examenblad.nl/system/files/2022/2022_05_09_verslag_opbrengsten_uitbreiding_vraagtypen_leesvaardigheid_ce_NL_hv.pdf
- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27(3), 209–226.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
- Leslie, L., & Caldwell, J.S. (2017). Assessments of reading comprehension: Challenges and directions. In S. E. Israel (Ed.), *Handbook of research on reading comprehension* (pp. 219–240). Routledge.
- LTN. (2024). *Handreikingen examenbesprekingen 2024*. <https://lerarennederlands.nl/handreikingen-examenbesprekingen-2024/>
- LTN. (2025). *Handreikingen examenbesprekingen 2025*. <https://lerarennederlands.nl/handreikingen-examenbesprekingen-2025/>
- Magliano, J. P., McCrudden, M. T., Rouet, J.-F., & Sabatini, J. (2018). The modern reader: Should changes to how we read affect research and theory? In M.F. Schober, D.N. Rapp, & M.A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 343–361). Routledge.
- McNamara, D. S., & Magliano, J. P. (2009). Self-explanation and metacognition: The dynamics of reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 60–81). Routledge.
- O'Reilly, T., Feng, D. G., Sabatini, D. J., Wang, D. Z., & Gorin, D. J. (2018). How do people read the passages during a reading comprehension test? The effect of reading purpose on text processing behavior. *Educational Assessment*, 23(4), 277–295.
- Ozuru, Y., Best, R., Bell, C., Witherspoon,

- A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399–438.
- Pearson, P. D., Garavaglia, D., Lycke, K., Roberts, E., Danridge, J., & Hamm, D. (1999). *The impact of item format on the depth of students' cognitive engagement*. American Institutes for Research. Technical Report.
- Rodriguez, M. C. (2003). Construct equivalence of multiple-choice and constructed-response items: A random effects synthesis of correlations. *Journal of Educational Measurement*, 40(2), 163–184.
- Roelofs, E., Farrokhnia, M., & Haeri, M. A. (2026). *Concept map analysis of reading comprehension assessment: Tracing situation models in multiple-choice and constructed-response items*. *Studies in Educational Evaluation*, 89, 101599.
- Rooijackers, P. (2023). *Oog voor diep begrip: Onderzoek naar het tekstbegrip van vwo-leerlingen* [Dissertatie, Universiteit Utrecht]. Stichting Cito. <https://doi.org/10.33540/1643>
- Rooijackers, P., & Stevens, C. (2025). *De cognitieve validiteit van meerkeuzevragen in de leesexamens Nederlands havo/vwo; een onderzoeksverslag*. Stichting Cito. [Intern onderzoeksverslag]
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23(4), 441–474.
- SLO. (2025). *Toelichtingsdocument conceptexamenprogramma's Nederlandse taal en literatuur voor vmbo, havo en vwo*. <https://www.slo.nl/@24902/toelichtingsdocument-nederlands-versie2/>

PATRICK ROOIJACKERS is leesonderzoeker bij Stichting Cito en verantwoordelijk voor de constructie van de vwo-examens Nederlands. Hij werkt daarnaast als vakdidacticus Nederlands aan de Universitaire Lerarenopleiding Tilburg.
E-mail: patrick.rooijackers@cito.nl

CORIEN STEVENS is toetsdeskundige bij Stichting Cito en verantwoordelijk voor de constructie van de havo-examens Nederlands.
E-mail: corien.stevens@cito.nl