

An isometric illustration on a blue background. A large yellow cube is in the top left. A man in a red shirt and black pants stands on a large sheet of paper, holding a giant orange pencil. The paper features various charts: a pie chart with red, yellow, and green segments, and a bar chart with red and blue bars. To the right, another man in a red shirt sits on the edge of the paper, holding a tablet displaying a pie chart. In the bottom left, a third person is partially visible, working on a laptop. The entire scene is framed by a large, dark blue, rounded rectangle.

**De verschillende
afnamemodi meten
de schakel-technische
vaardigheden even
betrouwbaar**

Stroomkringschema's examineren met practicum, computer of papier

Wat is het effect van de afnamemodus?

■ Marieke de Back

M. de Back, MSc, is als toetsdeskundige werkzaam bij Cito en stuurt de constructiegroep PIE aan.
E-mail: marieke.deback@cito.nl

■ Romy Noordhof

R. Noordhof, MSc, is als toetsdeskundige werkzaam bij Cito en betrokken bij de digitale onderdelen van het cspe. E-mail: romy.noordhof@cito.nl

■ Hans Kuhlemeier

Dr. H. Kuhlemeier is als onderzoeker werkzaam bij Cito. E-mail: hans.kuhlemeier@cito.nl

Examineren of leerlingen kennis van stroomkringschema's kunnen toepassen kan op verschillende manieren. Met een steekproef van ruim 400 leerlingen is onderzocht of de afnamemodus - in een practicum, op papier of op de computer - van inhoudelijk gelijke opdrachten verschillende resultaten oplevert.

Aanleiding tot het onderzoek

In de beroepsgerichte leerwegen van het vmbo worden tijdens het examen praktische kennis en vaardigheden getoetst. Daarom bestaat het examen (het centraal schriftelijk en praktisch examen of kortweg cspe) grotendeels uit praktijkopdrachten. Deze praktijkopdrachten brengen relatief veel organisatielast met zich mee. Voor het profielvak PIE (Produceren, Installeren en Energie) geeft vooral de organisatie van practicumopdrachten knelpunten. Dit heeft te maken met het aantal werkplekken op scholen, de kostprijs van practicummaterialen en met de intensieve

manier van beoordelen van een practicumopdracht (1 examiner op 2 leerlingen).

Het cspe PIE bevatte sinds 2015 twee practicumopdrachten: één elektropracticum en één pneumatiekpracticum. Met de opdrachtgever het College voor Toetsen en Examens (CvTE) is afgesproken om vanaf 2018 maximaal één practicumopdracht in het cspe PIE op te nemen, om zo de organisatielast voor scholen te beperken.

Om toch zowel elektrische als pneumatische schakelingen in het cspe aan bod te laten komen, heeft Cito voor het elektropracticum twee alternatieve opdrachten ontwikkeld: een computeropdracht en een opdracht op papier. Het is van belang om te achterhalen of deze modi dezelfde prestaties opleveren en dus gelijkwaardig zijn aan het gewaardeerde en vertrouwde practicum. Als ze gelijkwaardig blijken, dan kan in het cspe een computeropdracht gekozen worden als alternatief voor een practicumopdracht. Uitgangspunt is dat het deeltaken uit de examenstof betreft waarvoor meerdere afnamemodi geschikt worden geacht. In dit onderzoek is gekozen

voor het kunnen aansluiten van schakelingen op basis van stroomkringschema's, waarbij de vaardigheid de toepassing van kennis van stroomkringschema's betreft.

Onderzoeksvragen

Het onderzoek beoogt antwoord te geven op de vraag of de afnamemodus – practicum, papier of computer – van invloed is op de prestaties, de betrouwbaarheid en de rangordening van leerlingen. De drie onderzoeksvragen luiden als volgt:

1. Leidt de ene afnamemodus tot hogere prestaties dan de andere?
2. Meet de ene afnamemodus betrouwbaarder (in termen van interne consistentie) dan de andere?
3. Leiden de afnamemodi tot dezelfde rang-



**In beroepsgerichte
leerwegen van het vmbo
worden tijdens het examen
praktische kennis en
vaardigheden getoetst**

ordering van de leerlingen van zwak naar sterk vaardig?

De uitkomsten van dit onderzoek kunnen bijdragen aan het antwoord op de vraag of het gerechtvaardigd is om een enkele practicum-opdracht in het cspe te vervangen door een opdracht op de computer of op papier.

Eisen aan de constructie

Om de afnamemodi practicum, papier en computer methodologisch zuiver te kunnen vergelijken moeten de opdrachten aan ten minste twee basisvoorwaarden voldoen (Traub & MacRury, 1990). De te vergelijken modi van dezelfde opdracht moeten in de eerste plaats stamequivalent zijn. Dit wil zeggen dat de verwoording van de vraag- en probleemstelling en de instructie voor de leerlingen zoveel mogelijk gelijk moeten zijn. Naast stamequivalent moeten de modi ook scoringsequivalent zijn. Als de practicummodus bijvoorbeeld dichotoom (0-1) gescoord wordt, moet dat bij de papieren en de computermodus ook zo gebeuren. Sterker nog, het correctievoorschrift moet bij voorkeur volledig identiek zijn. Afnamemodi die niet stam- en scoringsequivalent zijn, zijn ongeschikt ter beantwoording van de vraag naar de gelijkwaardigheid van de drie afnamemodi. Immers, het effect van de afnamemodus is dan niet te onderscheiden van eventuele andere effecten.

De constructie van stam- en scoringsequivalente opgaven

Voor dit onderzoek is een toets van zes opdrachten ontwikkeld waarbij de leerling de bedrading van een schakel- of meetopstelling moet aanbrengen. Van deze opdrachten zijn steeds drie afnamemodi:

1. in een practicum
2. op papier
3. op de computer

De opdrachten zijn voor alle drie de afnamemodi identiek geformuleerd. Het enige verschil tussen de opdrachten is de afnamemodus: in de practicummodus moeten de leerlingen fysieke meetsnoeren in aansluitpunten steken,

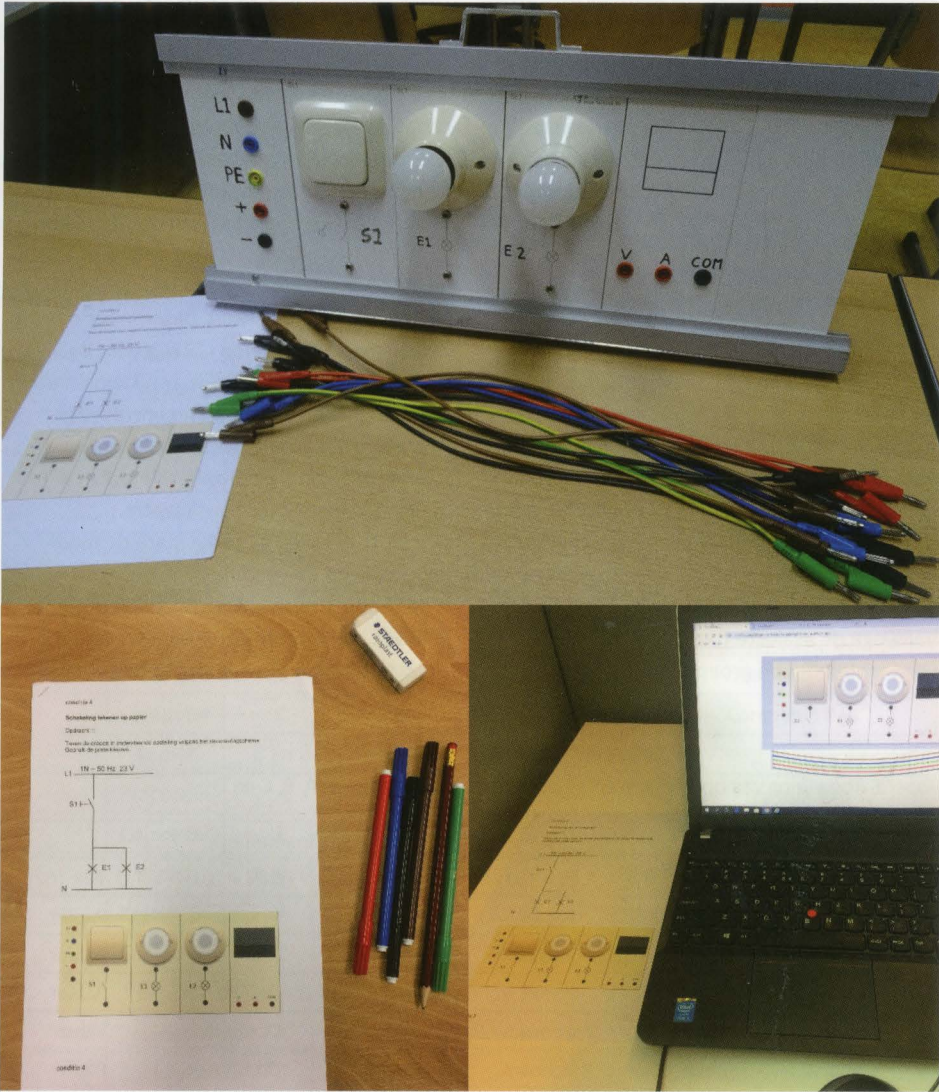
in de papiermodus moeten zij de draden met stiften tekenen en in de computermodus moeten zij gesimuleerde draden met de muis naar de juiste plek slepen. Ter illustratie toont Figuur 1 de drie afnamemodi van de eerste opdracht uit de toets.

Figuur 2 toont een deel van het correctievoorschrift, namelijk het deel horend bij de eerste opdracht. Het correctievoorschrift was voor

elke opdracht voor alle drie de afnamemodi steeds identiek.

nummer en omschrijving beoordelingsaspect		score
OPDRACHT 1		
1	Er loopt een draad (bruin) van L1 naar S1.	1
2	Er lopen twee draden (zwart) van S1 naar E1 en E2.	1
3	Er lopen twee draden (blauw) van N naar E1 en E2.	1
4	Overall zijn de juiste kleuren draden gebruikt.	1

Figuur 2. Deel van het correctievoorschrift met de vier beoordelingsaspecten van de eerste opdracht



Figuur 1. De eerste opdracht in drie afnamemodi: practicum, papier en computer

Toewijzing van leerlingen aan groepen met verschillende afnamemodi van opdrachten

Aan het onderzoek namen 406 leerlingen van negentien scholen voor vmbo deel. Allen volgden het profielvak Produceren, Installeren en Energie (PIE) en zouden binnen enkele weken tot maanden het profielvakexamen PIE maken. Van de leerlingen zat 25% in het derde leerjaar en 75% in het vierde leerjaar, hetgeen van belang is omdat zowel derde- als vierdejaars leerlingen het profielvakexamen mogen afronden. De verdeling naar leerweg was 36% basisberoepsgerichte leerweg en 64% kaderberoepsgerichte leerweg. De leerlingen zijn afzonderlijk op basis van toeval toegewezen aan vier groepen: A, B, C en D. Elke leerling maakte zes opdrachten met in totaal 33 dichotoom (0-1) gescoorde beoordelingsaspecten. Tabel 1 laat zien hoe de opdrachten met bijbehorende beoordelingsaspecten en de afnamemodi aan de vier groepen zijn toegewezen. Zo kreeg groep A eerst de opdrachten 1, 2 en 3 in de papiermodus voorgelegd en vervolgens de opdrachten 4, 5 en 6 in de practicummodus en zo maakte groep D alle zes opdrachten op papier. De verdeling van de leerlingen over de afnamemodi was binnen elke opdracht 50% papier, 25% practicum en 25% computer. Dit ontwerp maakt het mogelijk de afnamemodi zowel binnen als tussen groepen met elkaar te vergelijken.

Vanwege de aselechte toewijzing mag men verwachten dat de groepen A, B, C en D gelijk-

waardig zijn, maar honderd procent zeker is dat natuurlijk niet. Om die reden zijn we nagegaan in hoeverre de groepen verschilden voor wat betreft de gemiddelde prestaties op de gemeenschappelijke papieren opdrachten en de verdeling naar leerjaar (3 of 4) en leerweg (bb of kb). Geen van de acht uitgevoerde toetsingen (hier niet gerapporteerd) gaf significantie te zien op 5%-niveau. De groepen mogen om die reden inderdaad als gelijkwaardig worden beschouwd.

Statistische analyse

De verschillen tussen de gemiddelde prestaties voor de onderscheiden modi zijn op significantie getoetst met behulp van variantieanalyse. Behalve de gemiddelde prestaties en de standaarddeviatie rapporteren we ook de effectgrootte. Daarin is het verschil tussen twee gemiddelden uitgedrukt als een proportie van de standaarddeviatie. Bij de evaluatie van de grootte van de verschillen baseren we ons op de vuistregel van Cohen (1977), waarbij 0,20, 0,50 en 0,80 respectievelijk een klein, middelgroot en groot verschil vertegenwoordigt. De betrouwbaarheid van de onderscheiden modi is bepaald met Cronbach's alpha, een maat voor de interne consistentie. Om uit te drukken in welke mate de modi de leerlingen op vergelijkbare wijze ordenen is de product-moment correlatiecoëfficiënt gebruikt. De waarde kan variëren tussen -1 en +1. Daarbij staat -1 voor een volstrekt nega-

Tabel 1. Toewijzing van opdrachten en afnamemodi aan de groepen A, B, C en D

Opdracht	Beoordelingsaspecten	Groep en afnamemodi			
		A (N = 101)	B (N = 102)	C (N = 101)	D (N = 102)
1	1-4	Papier	Computer	Practicum	Papier
2	5-9	Papier	Computer	Practicum	Papier
3	10-14	Papier	Computer	Practicum	Papier
4	15-20	Practicum	Papier	Computer	Papier
5	21-26	Practicum	Papier	Computer	Papier
6	27-33	Practicum	Papier	Computer	Papier

tieve samenhang, 0 voor een totaal gebrek aan samenhang en +1 voor een perfect positieve samenhang (waarbij het voor de rangorde-ning van de leerlingen niet uitmaakt of zij de opdrachten in de ene of de andere afnamemo-dus maken).

Geen noemenswaardige verschillen tussen modi in gemiddelde prestaties

Tabel 2 toont onder meer de gemiddelde prestaties per opdracht per modus met de bijbehorende standaarddeviatie. Omwille van de vergelijkbaarheid zijn de scores omgezet naar een percentage van de maximumscore. Getoetst is of de verschillen tussen de gemid-delden voor de telkens drie modi van een opdracht statistisch significant zijn (op 5%-ni-veau). Slechts één van de twaalf getoetste verschillen tussen de modi gaf significantie te

zien. Het betrof het verschil tussen de gemid-delden voor de papieren en practicum modus van de eerste opdracht. De practicummodus van deze opdracht blijkt beter gemaakt dan de papieren modus: 74% versus 66% goede ant-woorden. De bijbehorende effectgrootte van 0,29 ligt iets boven de 0,20 waarbij er volgens de vuistregel van Cohen (1977) sprake is van een klein verschil. Een nadere bestudering van de eerste opdracht bracht geen inhoudelijke verklaringen aan het licht, zodat we de betere prestaties op de papieren opdracht aan toeval-lige steekproeffluctuaties moeten toeschrij-ven. Ook bij de derde opdracht is er gezien de effectgrootte van 0,20 sprake van een klein verschil in het voordeel van de practicumop-dracht. Gezien de afwezigheid van significantie moet dit verschil echter als verwaarloosbaar worden beschouwd. Al met al kunnen we niet

Tabel 2. Gemiddelden, standaarddeviaties en effectgroottes voor de verschillen tussen de modi per opdracht

Opdracht	Modus	Groep	N	Gem	Stddev	Contrast	Effectgrootte
1	Papier	A en D	203	65,52	29,74		
	Practicum	B	100	74,00	29,07	Prac min Pap*	0,29
	Computer	C	101	70,30	30,76	Comp min Pap	0,16
2	Papier	A en D	202	66,93	27,49		
	Practicum	B	100	71,80	30,50	Prac min Pap	0,17
	Computer	C	101	71,09	29,59	Comp min Pap	0,15
3	Papier	A en D	202	59,41	28,14		
	Practicum	B	100	65,20	28,80	Prac min Pap	0,20
	Computer	C	102	57,45	32,05	Comp min Pap	-0,06
4	Papier	B en D	203	61,66	26,96		
	Practicum	A	99	59,43	24,99	Prac min Pap	-0,09
	Computer	C	100	62,67	28,34	Comp min Pap	0,04
5	Papier	B en D	202	55,45	28,90		
	Practicum	A	97	52,23	28,83	Prac min Pap	-0,11
	Computer	C	101	56,27	29,52	Comp min Pap	0,03
6	Papier	B en D	202	52,76	28,84		
	Practicum	A	95	56,54	25,00	Prac min Pap	0,14
	Computer	C	99	56,71	26,25	Comp min Pap	0,14

**p* < 0,05

vaststellen dat de ene modus gemiddeld genomen tot duidelijk hogere of lagere prestaties leidt dan de andere.

Geen verschillen tussen modi qua interne consistentie

Tabel 3 geeft inzicht in de verschillen tussen de drie modi voor wat betreft de betrouwbaarheid van de meting in termen van interne consistentie (Cronbach's alfa). Gegeven het gehanteerde afnameontwerp met vier wederzijds exclusieve groepen kandidaten kan de betrouwbaarheid alleen binnen elke groep bepaald worden. De betrouwbaarheid van de modi is daarom berekend binnen de twee aangeboden modi per groep, dit wil zeggen de interne consistentie over a) de 14 beoordelingsaspecten van de opdrachten 1, 2 en 3 en over b) de 19 beoordelingsaspecten van de opdrachten 4, 5 en 6 (zie Tabel 2). Ter vergelijking toont de rechterkolom in tabel 3 de betrouwbaarheid na verlenging van de toets

tot 40 opgaven met de formule van Spearman-Brown.

Wat onmiddellijk opvalt, is dat alle betrouwbaarheden hoog zijn (d.w.z. $> 0,80$) en bovendien qua grootte weinig van elkaar verschillen ($0,84 < \alpha < 0,89$). Kijken we naar de 95%-betrouwbaarheidsintervallen rond de schattingen van de betrouwbaarheid dan blijken deze elkaar in alle acht gevallen te overlappen. De conclusie is dat de ene modus even betrouwbaar meet als de andere.

Modi rangordenen de leerlingen ongeveer even goed

Verschillen tussen modi kunnen niet alleen worden uitgedrukt in termen van gemiddelde prestaties, spreiding en interne consistentie, maar ook in de wijze waarop de modi de leerlingen op basis van hun prestaties ordenen van zwak naar sterk vaardig. Tabel 4 geeft een indicatie van de mate waarin de drie modi tot een vergelijkbare rangordening van de leerlingen

Tabel 3. Verschillen tussen modi qua betrouwbaarheid in termen van interne consistentie (Cronbach's alfa)

Opdracht	Beoordelingsaspecten	Groep	Modus	Cronbach's alfa	95%-betrouwbaarheids-interval		Cronbach's alfa 40 opgaven
					Ondergrens	Bovengrens	
1, 2 en 3	1-14	A	Papier	0,84	0,78	0,88	0,94
		B	Practicum	0,88	0,84	0,91	0,95
		C	Computer	0,86	0,82	0,90	0,95
		D	Papier	0,87	0,83	0,91	0,95
4, 5 en 6	15-33	A	Practicum	0,85	0,79	0,89	0,92
		B	Papier	0,89	0,85	0,92	0,94
		C	Computer	0,88	0,84	0,91	0,94
		D	Papier	0,88	0,85	0,91	0,94

leiden. Voor de groepen A, B en C zijn de correlaties tussen de telkens twee afnamemodi berekend. Daarbij heeft de als eerste genoemde modus betrekking op de beoordelingsaspecten 1 tot en met 14 van de opdrachten 1, 2 en 3 en de als tweede genoemde modus op de beoordelingsaspecten 15 tot en met 33 van de opdrachten 4, 5 en 6. Ter vergelijking toont de onderste rij van tabel 4 de correlatie tussen de samenhang van de prestaties in groep D op de twee onderdelen van de papieren modus (opdracht 1, 2 en 3 en opdracht 4, 5 en 6). Binnen de groepen A, B en C zien we een sterke en positieve samenhang tussen de gemiddelde prestaties op de telkens twee modi ($0,80 < r < 0,85$). Kennelijk maakt het voor de rangordening van de leerlingen weinig verschil of daartoe de ene of de andere modus wordt gebruikt. Onder de aanname dat de modus van invloed is op de rangordening, zou men verwachten dat de correlatie tussen de beide delen van de papieren modus in groep D hoger is dan die tussen de beide delen van de verschillende modi binnen de groepen A, B en C. De correlaties tussen de beide delen van dezelfde papieren modus in groep D is met 0,83 echter van dezelfde orde van grootte (als die tussen verschillende modi in de groepen A, B en C). Dit vormt een extra aanwijzing dat de modi weinig van elkaar verschillen als het erom gaat de leerlingen op basis van hun prestaties te ordenen van zwak naar sterk vaardig.

Conclusie

Het praktijkdeel van het cspe Produceren, Installeren en Energie (PIE) kende tot voor kort twee practica. Om de organisatielast voor

de scholen te verminderen werd in 2018 een van de beide practica vervangen door een papieren opdracht en in de toekomst mogelijk door een computeropdracht. Ons onderzoek had tot doel een antwoord te geven op de vraag of het vervangen van een practicum - door een papieren of een computeropdracht - gerechtvaardigd is. In het onderzoek zijn drie modi van dezelfde zes opdrachten met elkaar vergeleken. Vergeleken zijn de gemiddelde prestaties, spreiding, betrouwbaarheid en rangordening van de leerlingen. In het onderzoek hebben we niet kunnen vaststellen dat de ene modus tot hogere of lagere prestaties of tot meer of minder spreiding leidt dan de andere. Daarnaast blijkt de ene modus de schakel-technische vaardigheden even betrouwbaar te meten als de andere. De modi blijken ook niet van elkaar te verschillen als het erom gaat de leerlingen op basis van hun prestaties te ordenen van zwak naar sterk vaardig. Deze uitkomsten laten zien dat de onderzochte modi in hoge mate vergelijkbaar zijn en waarschijnlijk een beroep doen op dezelfde cognitieve vaardigheden.

Kanttekeningen en aanbevelingen

De uitkomsten van het onderzoek met de schakelopdracht zijn gunstig, want de inzet van computeropdrachten kan enorm bijdragen aan het verminderen van de organisatielast voor scholen. Er is echter een kanttekening te maken. Een cspe is de afsluiting van een beroepsgericht vak. Leerlingen die dit beroepsgerichte vak volgen, worden voorbereid op een praktische vervolgopleiding en uiteindelijk een praktisch beroep. Praktijkvaardigheid is hun

Tabel 4. Correlaties tussen modi per opdracht binnen groepen

Groep	Samenhang tussen	Correlatie
A	Papier en Practicum	0,80
B	Computer en Papier	0,85
C	Practicum en Computer	0,81
D	Papier en Papier	0,83

kracht en om die reden volgen zij de opleiding. Het cspe moet recht blijven doen aan deze leerlingen, waarbij praktische opdrachten en een authentieke context waardevol zijn. Het is dan ook niet de bedoeling om grootschalig praktijkopdrachten uit het cspe te vervangen, zelfs niet als dat ten gunste zou zijn van de organisatielast. Er moet steeds een goede afweging gemaakt worden tussen wat een afnamemodus oplevert en wat ermee verloren gaat.

De uitkomsten van dit onderzoek zijn een stimulans om door te gaan met de ontwikkeling van andere computeropdrachten, bijvoorbeeld voor andere beroepsgerichte vakken waar de organisatielast hoog is. Op basis van dit ene onderzoek durven we nog geen conclusies te trekken over een bredere scope dan PIE. Daarom verdient het aanbeveling een soortgelijk onderzoek uit te voeren na ontwikkeling van een computeropdracht voor een ander beroepsgericht vak.

Een andere aanbeveling is om ook dynamischer computersimulaties te ontwikkelen, die qua mogelijke handelingen en werking nog meer lijken op een echt practicum. Zo kan de computeropdracht PIE authentiek gemaakt worden als de leerling niet alleen de draden, maar bijvoorbeeld ook de lampen en schakelaars kan plaatsen. Of als deze computersimulatie ook werkende elementen zou bevatten die de leerling (net als in het practicum) kan

testen, zoals een brandende lamp. Uiteraard moet een dergelijke dynamischer simulatie eerst goed getest worden.

Literatuur

- Cohen, J. (1977). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Routledge.
- Traub, R.E. & MacRury, K. (1990) Antwortauswahl- vs. freie-antwort-aufgaben bei lernfolgstests. [Multiple-choice vs. free-response in the testing of scholastic achievement]. In K. Ingenkamp & R.
- S. Jager (Eds), Tests und trends 8: Jahrbuch der padagogischen diagnostic (pp. 128-169). Weinheim Germany: Beltz Verlag
- Traub, R. (1993). On the equivalence of the traits assessed by multiple-choice and constructed response tests. In Bennett, R., & Ward, W. (eds.). Construction versus choice in cognitive
- measurement (pp. 29-44). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dankwoord

Graag willen we alle leerlingen en docenten die hebben meegedaan aan het onderzoek hartelijk bedanken. Ook dank aan de scholen die zich wel hadden aangemeld, maar die we uiteindelijk niet bezocht hebben. Tot slot dank aan de vele collega's die een bijdrage hebben geleverd aan het onderzoek. ■

Gesignaleerd

Nieuwe rekentoets: een 4 halen is voldoende

Middelbare scholen en mbo's gaan zelf een rekenexamen maken voor hun eigen scholieren. Voor deze nieuwe rekentoets hoeft geen voldoende meer worden gehaald, een 4 of hoger is genoeg om het diploma te halen. De landelijk verplichte rekentoets verdwijnt definitief.

Uit: BNdeStem, 10 november 2018