

De rekenvaardigheid van de Nederlandse verpleegkundige

De rekenvaardigheid van verpleegkundigen is de afgelopen jaren regelmatig in het nieuws geweest. Diverse onderzoekers suggererden dat die niet toereikend zou zijn (Hoes, 2005; de Jong & Koster, 2007; DUO Market Research, 2010). Het is logisch dat de beroepsgroep zich daarover zorgen maakte en activiteiten ontplooid om de rekenvaardigheid weer op peil te brengen. Toetsing speelde steeds een centrale rol bij die pogingen. Door het gemis aan een gestandaardiseerd instrument ontwikkelden groepen vaak hun eigen toets en bijbehorende prestatiestandaard wat de helderheid van de discussie over de problematiek niet bevordert heeft. Voor Cito was dit een reden om een gestandaardiseerd toetspakket 'Verpleegkundig Rekenen' te ontwikkelen. Het hart van dit pakket is een opgavenbank waarin honderden opgaven zijn opgeslagen over rekenproblemen waarmee de verpleegkundige in de praktijk te maken kan krijgen.

Deze bijdrage gaat uitsluitend over de opgavenbank en met name over een kwaliteitsonderzoek naar de opgaven daarin. Dat onderzoek beoogt niet alleen zicht te geven op de kwaliteit van de opgaven maar ook op de rekenvaardigheid van de proefpersonen in dit onderzoek. Omdat de onderzochte proefpersonen een representatieve steekproef vormen uit de populatie geven de resultaten tevens een indicatie over de rekenvaardigheid van de Nederlandse verpleegkundige in het algemeen.

Tecla Lampe, Gerard Straetmans en Theo Eggen *

Opzet en ontwikkeling van de opgavenbank

Toetsen kunnen alleen valide scores opleveren als ze opgaven bevatten die vakinhoudelijk en toetstechnisch in orde zijn. De productie van de opgavenbank van het te ontwikkelen toetspakket was daarom een strak geregisseerd proces met veel tussentijdse kwaliteitscontroles.

Als eerste stap in de ontwikkeling van de opgavenbank hebben inhoudelijk deskundigen het begrip 'verpleegkundige rekenvaardigheid' nader gespecificeerd door het benoemen van deelgebieden waarop het verpleegkundig rekenen gericht kan zijn. De constructie werd verder gestuurd door de specificatie van het gewenste percentage opgaven waarmee elk deelgebied in een toets vertegenwoordigd zou moeten zijn volgens experts op het gebied van verpleegkundig rekenen (tabel 1). Verpleegkundigen met veel kennis van en affiniteit voor het onderwerp schreven de opgaven. Met het oog op de geautomatiseerde verwerking werden alleen meerkeuze-opgaven ontwikkeld. Een toetsdeskundige van Cito, begeleidde de constructie, screende de kwaliteit en deed suggesties voor verbetering. Het constructieteam heeft in totaal 538 opgaven geproduceerd. De aantallen per deelgebied verhouden zich tot elkaar conform de wenselijke percentages per deelgebied in de toetsen.

Opzet en resultaten van de proefafname

Een zorgvuldig constructieproces is noodzakelijk om goede opgaven te produceren, maar kan geen zekerheid geven over de vraag of op-

Tabel 1: Gewenste percentage opgaven per deelgebied in een toets Verpleegkundig Rekenen

Deelgebied	%
Vaste medicatie	25
Vloeibare medicatie	40
Transfusie	5
Infusie	15
Gassen	5
Vochtbalans	10
Totaal	100

gaven voldoende kwaliteit hebben voor het doel waarvoor ze gemaakt zijn. Daarvoor is een proefafname nodig bij een steekproef uit de doelgroep. De antwoorden van de proefpersonen worden daarbij geanalyseerd om onder andere de moeilijkheidsgraad en het discriminerend vermogen¹ van de opgaven te bepalen. Deze twee kwaliteitskenmerken van opgaven zijn minimaal nodig om toetsen samen te kunnen stellen die valide scores opleveren.

I Opzet

Om de analyses betrouwbare resultaten te laten opleveren is het noodzakelijk dat elke te testen opgave door een groot aantal proefpersonen uit de doelgroep gemaakt wordt. Zorginstellingen werden op directe (via een brief) en indirecte (persbericht) wijze benaderd met het verzoek deel te nemen aan de proefafname. De werving leverde 42 (locaties) van deelnemende instellingen op met 5.578 ingeschreven proefpersonen waarvan er uiteindelijk 2.662 uit 34 instellingen participeerden in het onderzoek. Het merendeel daarvan was tot deelname verplicht door het management van de instelling.

Uiteraard is het niet mogelijk om elke proefpersoon alle 538 opgaven uit de bank te laten maken. Daarom werden deze verdeeld over 36 toetsboekjes met elk 30 opgaven. De verdeling van de opgaven werd zodanig geregeld dat de helft van de opgaven in elk toetsboekje ook was opgenomen in nog een ander toetsboekje. Dit zorgt ervoor dat alle verzamelde antwoorden gezamenlijk geanalyseerd kunnen worden. Door Cito geïnstrueerde surveillanten zagen erop toe dat de proefpersonen werkten volgens voorgeschreven toetscondities. Elke opgave werd door gemiddeld 148 proefpersonen beantwoord.

2 Representativiteit steekproef

Uitspraken over de rekenvaardigheid van de verpleegkundige op basis van een steekproef zijn alleen houdbaar als de onderzochte proefpersonen in voldoende mate representatief zijn voor de populatie van verpleegkundig personeel. De representativiteit van de steekproef is geëvalueerd door deze op bepaalde kengetallen, zoals leeftijd, opleidingsniveau en werkervaring, te vergelijken met gegevens van het Centraal Bureau voor de Statistiek (CBS) en het Landelijk Expertisecentrum Verpleging en Verzorging (LEVV). Tabel 2 laat zien dat de leeftijdsopbouw van de proefpersonen in de steekproef een bevredigende² overeenkomst vertoont met die van het CBS.

Tabel 2: Aantal personen per leeftijdscategorie, werkzaam in de lagere (para)medische beroepen in 2008, volgens CBS en de proefafname

Leeftijd	CBS		Proefafname	
	n	%	n	%
15 t/m 24 jaar	12.000	29	322	12
25 t/m 34 jaar	7.000	17	695	26
35 t/m 44 jaar	9.000	22	694	26
45 t/m 54 jaar	10.000	24	687	26
55 t/m 64 jaar	3.000	7	248	9

In tabel 3 worden gegevens over de hoogst genoten opleiding van de proefpersonen afgezet tegen de beschikbare data daarover van het LEVV (<http://www.levv.nl>, oktober 2009). Als een In service A/B/Z-opleiding beschouwd mag worden als gelijkwaardig aan een opleiding op niveau 4, dan komt het opleidingsniveau van de proefpersonen in de steekproef goed overeen met de ramingen daarover van het LEVV.

Tabel 3: Opleidingsniveau volgens LEVV-raming en de proefafname

	LEVV-raming 2008		Proefafname	
	n	%	n	%
MBO niveau 4	46.000	68,8	780	30
In service A/B/Z			960	37
HBO niveau 5	20.900	31,2	701	27
Volgt nu MBO-opl.			17	0,7
Volgt nu HBO-opl.			17	0,7
Anders			121	4,7

Ook wat betreft de werkervaring passen de ramingen van LEVV (gemiddeld 21,5 jaar voor niveau 4 verpleegkundigen en 14,8 jaar voor niveau 5 verpleegkundigen) goed bij de gegevens van de steekproef. Die laten zien dat personen met meer dan 15 jaar werkervaring twee keer zo vaak een opleiding op niveau 4 als op niveau 5 hebben.

Gegeven de overeenkomsten tussen de gegevens van de proefafname en die van CBS en LEVV lijkt het gerechtvaardigd om de steekproef te beschouwen als voldoende representatief voor de populatie van verpleegkundigen.

3 Kwaliteit van het toetsmateriaal

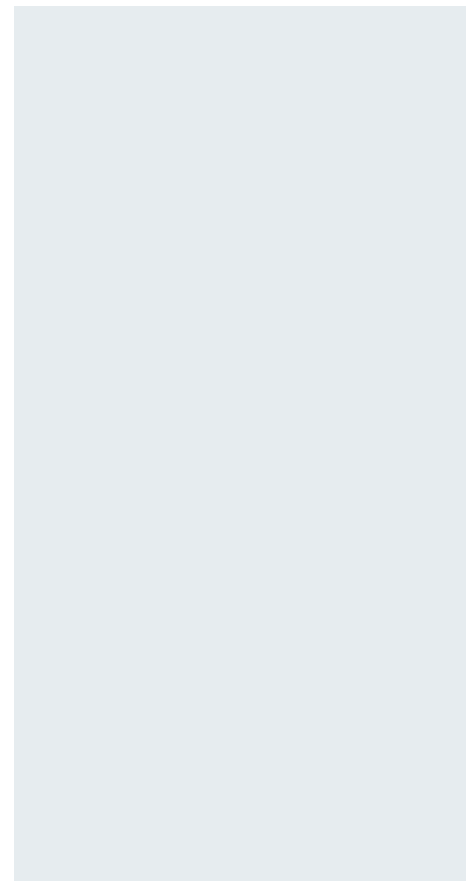
De gegevens van de proefafname zijn op twee manieren gebruikt om de kwaliteit van de opgaven mee te evalueren. De eerste manier is een analyse op basis van de klassieke testtheorie en levert naast informatie over de kwaliteit van individuele opgaven ook informatie over de kwaliteit van de toets als geheel. Alhoewel die twee kwaliteiten nauw met elkaar verband houden, zullen we hier alleen de informatie over de kwaliteit van de toetsen rapporteren. De betrouwbaarheid is een van de bekendste maten voor de toetskwaliteit. Hij geeft aan in hoeverre toetsresultaten vrij zijn van de invloed van toevallige factoren, ook wel meetfout genoemd. Hoe hoger de betrouwbaarheid des te kleiner de meetfout die gemaakt wordt. Er zijn verschillende manieren om de betrouwbaarheid te schatten. Een veel gebruikte maat is coëfficiënt alpha, die waarden kan aannemen van 1 tot 0. Hoe hoger de waarde des te hoger de betrouwbaarheid van de toetsresultaten. Tabel 4 laat voor elk toetsboekje de geschatte betrouwbaarheden zien.

Tabel 4: Overzicht van de geschatte betrouwbaarheden van de toetsboekjes

boekje	coëff. α	boekje	coëff. α	boekje	coëff. α	boekje	coëff. α
1	0,75	10	0,78	19	0,84	28	0,81
2	0,80	11	0,82	20	0,79	29	0,78
3	0,77	12	0,80	21	0,80	30	0,83
4	0,77	13	0,80	22	0,67	31	0,83
5	0,77	14	0,73	23	0,70	32	0,80
6	0,66	15	0,74	24	0,85	33	0,82
7	0,81	16	0,79	25	0,81	34	0,84
8	0,81	17	0,85	26	0,78	35	0,77
9	0,82	18	0,72	27	0,72	36	0,73

Voor de interpretatie van tabel 4 vinden we steun bij Nunnally en Bernstein (1994, p. 265), die aangeven dat een toets die wordt gebruikt voor het nemen van belangrijke beslissingen minstens een betrouwbaarheid van 0,80 moet hebben. Met belangrijke beslissingen wordt bedoeld: “beslissingen die op basis van de toetsscores worden genomen, die in principe, of op korte termijn, onomkeerbaar zijn, en die voor een belangrijk deel buiten de geteste persoon om worden genomen” (Cotan, 2010, p. 33). Ongeveer de helft van de toetsboekjes voldoet niet aan deze norm. Toch kunnen we er vertrouwen in hebben dat de toetsscores die het pakket Verpleegkundig Rekenen zal genereren voldoende betrouwbaar zullen zijn. In de eerste plaats omdat bij de samenstelling van de definitieve toetsen minder goed functionerende opgaven niet gebruikt zullen worden. In de tweede plaats omdat coëfficiënt alpha de betrouwbaarheid onderschat (Sijtsma, 2009, p. 565).

De tweede manier om de gegevens te analyseren is afkomstig uit de moderne testtheorie en wordt daar aangeduid met de naam Item-responstheorie (IRT). Zoals de naam al aangeeft staat bij deze analyses het functioneren van de individuele opgaven voorop. Eenvoudig gezegd komt het erop neer dat eerst gekeken wordt of een bepaald, hier niet nader te specificeren, wiskundig model een beschrijving en voorspelling kan geven van de antwoorden van de proefpersonen.



Vervolgens worden de moeilijkheidsgraad en de discriminatiewaarde geschat van elke opgave waarvoor het model geldig is. Opgaven die functioneren volgens het model vormen gezamenlijk een schaal. Zo'n schaal kan men zich voorstellen als een reeks opgaven die geordend zijn op moeilijkheidsgraad. Het schaalbegrip houdt in dat een kandidaat die een bepaalde opgave correct beantwoordt met een grotere kans ook correct zal antwoorden op opgaven met een lagere schaalwaarde. Echter, naarmate opgaven met hogere schaalwaarden worden aangeboden, zullen de kansen op een correct antwoord steeds verder afnemen. Een andere relevante eigenschap is dat zowel de moeilijkheidsgraad van de opgaven als de vaardigheid van de personen op dezelfde schaal afgebeeld kunnen worden. Dit klinkt wat ingewikkeld, maar de meeste mensen hebben hier ervaring mee. Bij het hoogspringen, bijvoorbeeld, worden de vaardigheid van de persoon en de lathoogte (lees: moeilijkheidsgraad) ook allebei uitgedrukt in termen van dezelfde schaal.

Een gedetailleerde bespreking van de principes van de IRT valt buiten het bestek van dit artikel. Geïnteresseerde lezers vinden een toegankelijk geschreven inleiding over de principes van IRT bij Baker (1985).

Van de 538 geteste opgaven bleken er 525 te functioneren volgens het gekozen wiskundige model en op een gemeenschappelijke schaal geplaatst te kunnen worden. Zie figuur 1 voor de verdeling van de opgaven op deze schaal, die we hierna zullen aanduiden als 'schaal voor verpleegkundige rekenvaardigheid'.

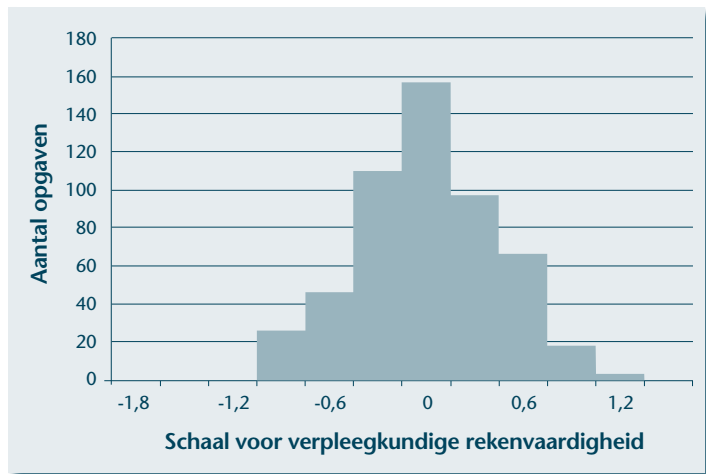
4 Verdeling van de proefpersonen op de schaal

Zoals hiervoor aangegeven kunnen op de geconstrueerde schaal niet alleen de opgaven (naar moeilijkheidsgraad) gepositioneerd worden, maar ook de proefpersonen voor wat betreft hun rekenvaardigheid, zoals gemeten met deze opgaven. Op basis van het responspatroon (de reeks met correct en foutief beantwoorde opgaven) is voor elke proefpersoon met behulp van schattingsprocedures de positie bepaald van zijn of haar vaardigheid op de geconstrueerde schaal (Verhelst, 1993). Figuur 2 laat de vaardigheidsverdeling op de schaal van de 2.662 proefpersonen zien. De verdeling is bij benadering normaal; de meesten hebben een vaardigheid rondom het groepsgemiddelde en de afwijkingen naar boven en beneden houden elkaar ongeveer in evenwicht.

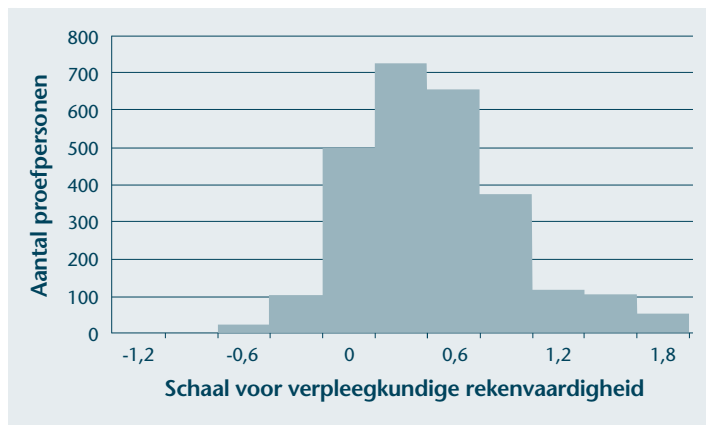
Interessant is verder om te kijken of diverse te onderscheiden subgroepen verschillend gepresteerd hebben. In tabel 5 tot en met 7 worden de gemiddelde vaardigheidsschattingen gegeven van subgroepen waarvan de samenstelling bepaald is door één achtergrondkenmerk. De resultaten zijn uitgedrukt op de geconstrueerde schaal, waarbij de schaalwaarden van de geteste personen varieerden van ongeveer -1 (zeer weinig vaardig) tot ongeveer 2 (zeer vaardig).³

Uit tabel 5 wordt duidelijk dat de gemiddelde vaardigheidsschatting van mannelijke verpleegkundigen hoger is dan die van hun vrouwelijke collega's. Statistische toetsing laat zien dat dit verschil significant is (t-test; $p < 0,001$). De effectgrootte (Cohen's d) bedraagt 0,307, wat inhoudt dat het hogere gemiddelde in de scoreverdeling van de groep met het lagere gemiddelde ongeveer samenvalt met het 62e percentiel. Een behoorlijk verschil dus. Echt verrassend is dit niet. Uit de literatuur is bekend dat mannen op rekentaken vaak significant beter presteren dan vrouwen, terwijl die verschillen veel minder vaak optreden bij andere cognitieve vaardigheden, zoals bijvoorbeeld taal (Van der Velden, 1996).

Figuur 1: Verdeling van de opgaven op de schaal voor verpleegkundige rekenvaardigheid. Aantal opgaven is 525.

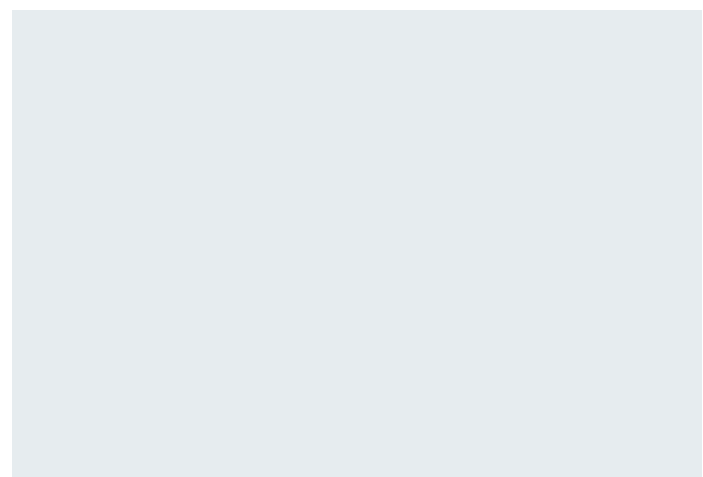


Figuur 2: Verdeling van de proefpersonen op de schaal voor verpleegkundige rekenvaardigheid. N = 2662.



Tabel 5: Gemiddelde vaardigheidsschattingen naar geslacht

Geslacht	Aantal proefpersonen ⁴	Gemiddelde vaardigheid
Man	282	0,609
Vrouw	2295	0,475



Ook de verschillen tussen de drie vooropleidingen zijn fors en in de verwachte richting (zie Tabel 6). Dit is een steuntje in de rug voor de veronderstelling dat met deze opgaven verschillen in (verpleegkundige reken-) vaardigheid tussen personen gemeten kunnen worden. Alle verschillen in gemiddelde vaardigheid tussen de drie vooropleidingen zijn statistisch significant (t-test; $p < 0,001$). Effectgroottes (Cohen's d) waren middelmatig (MBO vs In-service: $d = 0,304$; In-service vs HBO: $d = 0,293$; MBO vs HBO: $d = 0,59$).

Tabel 6: Gemiddelde vaardigheidsschattingen naar geslacht

Vooropleiding	Aantal proefpersonen ^a	Gemiddelde vaardigheid
MBO (niveau 4) MBO-V	780	0,382
In-service A/B/Z	960	0,503
HBO (niveau 5) HBO-V	701	0,629
Volgt MBO-opleiding	17	0,078
Volgt HBO-opleiding	17	0,365
Anders	121	0,268

De werkervaring die een verpleegkundige heeft is volgens de resultaten van de proefafname niet van invloed op de rekenprestaties. Tabel 7 laat zien dat geen van de verschillen tussen de gemiddelde vaardigheden van de onderscheiden categorieën significant was, behalve als vergeleken werd met de gemiddelde vaardigheid van de categorie 'geen werkervaring'.

Tabel 7: Gemiddelde vaardigheidsschattingen naar relevante werkervaring

Aantal jaren werkervaring	Aantal proefpersonen	Gemiddelde vaardigheid
Geen	88	0,286
< 1 jaar	126	0,494
1 t/m 3 jaar	334	0,542
4 t/m 5 jaar	196	0,524
6 t/m 10 jaar	444	0,511
11 t/m 15 jaar	305	0,494
> 15 jaar	1151	0,467

Verder is gekeken naar de prestaties van drie groepen (aangeduid als A, B en C) met combinaties van kenmerken die het meest voorkomen. In groep A zitten proefpersonen van het vrouwelijke geslacht die een In-service opleiding genoten hebben en meer dan 15 jaar ervaring hebben. Ze maken 24% uit van de steekproef. Omdat de steekproef voldoende representatief is kan de gemiddelde vaardigheid van deze groep opgevat worden als kenmerkend voor de modale verpleegkundige in Nederland. De gemiddelde geschatte vaardigheid van deze groep is 0,502. De op één na meest voorkomende combinatie (7,3%) wordt gevormd door vrouwelijke verpleegkundigen met een MBO-opleiding en een werkervaring van 6 tot en met 10 jaar. Deze groep (B) scoort een gemiddelde van 0,404. In groep C ten slotte zit 5,5% van de steekproef en bevat vrouwen met een HBO-opleiding en meer dan 15 jaar ervaring. De gemiddelde vaardigheid van deze groep bedraagt 0,551. De verschillen tussen de gemiddelden van groep A en B en van groep B en C zijn statistisch significant (t-test; $p < 0,01$), die van groep A en C niet.

■ Voldoende rekenvaardigheid: Wat is het en wie heeft het?

De resultaten uit de proefafname laten zien hoe de vaardigheden van de proefpersonen verdeeld zijn over de geconstrueerde schaal voor verpleegkundige rekenvaardigheid.

Voor de interpretatie daarvan, maar ook voor de interpretatie van de toetsresultaten die het toetspakket Verpleegkundig Rekenen in de operationele fase zal opleveren, is het noodzakelijk om een punt op de schaal te bepalen dat de minimale rekenvaardigheid representeert waarover een verpleegkundige moet beschikken om de opgedragen taken op het gebied van medicatie adequaat uit te kunnen voeren. In toetsjargon wordt dan gesproken over een cesuurscore.

Aan de hand van zo'n cesuurscore zijn de toetsscores van individuele kandidaten heel eenvoudig te interpreteren. Zolang de geschatte vaardigheid van een persoon kleiner is dan de cesuurscore schiet de vaardigheid tekort en is de persoon gezakt voor de toets. In alle andere gevallen is de persoon geslaagd. De vastgestelde cesuurscore kan gebruikt worden voor elke toets die uit het toetspakket Verpleegkundig Rekenen wordt samengesteld, mits de kandidaat deel uitmaakt van de doelgroep.

Het ontwikkelen en vaststellen van een valide cesuurscore is een van de lastigste aspecten van toetsconstructie. In de afgelopen jaren zijn diverse methodieken ontwikkeld om dit op verantwoorde wijze te doen (Cizek, 2006, p.238). Twee daarvan, de 'Bookmark methode' en de 'Angoff methode', werden ingezet om een cesuurscore te ontwikkelen voor het toetspakket Verpleegkundig Rekenen. Concreet hield dit in dat zes vakinhoudelijke experts in een speciaal georganiseerde sessie onder leiding van een toetsdeskundige, onafhankelijk van elkaar, een representatieve toets van 40 opgaven uit de opgavenbank evalueerden. Deze experts vertegenwoordigden verschillende belangengroeperingen, zoals een beroepsvereniging, (academisch) ziekenhuis, opleiding en hun expertise op het gebied van verpleegkundig rekenen wordt in brede kring erkend. Hun taak was om de volgende twee opdrachten uit te voeren:

1. Voor elke opgave diende te worden aangegeven of die correct beantwoord moest kunnen worden door een fictieve persoon die over juist voldoende verpleegkundige rekenvaardigheid beschikt;
2. In de reeks van 40 van eenvoudig naar

moelijk geordende opgaven moest een markering (bookmark) worden gezet bij de opgave die de persoon met juist voldoende verpleegkundige rekenvaardigheid nog correct zou kunnen beantwoorden.

De toepassing van beide methodes leidde na twee achtereenvolgende beoordelings- en discussierondes tot een minimaal te behalen toetsscore van (afgerond) 31 uit 40 punten, wat op de geconstrueerde schaal overeenkomt met een schaalwaarde van 0,784. Dat is bijna één standaarddeviatie boven de gemiddelde vaardigheidsschatting van de personen die deelnamen aan de proefafname. Ter vergelijking: Het 3F-referentieniveau voor rekenen, dat volgens de commissie Meijerink aan het einde van een MBO- of HAVO-opleiding behaald moet zijn, ligt op de geconstrueerde schaal bij schaalwaarde 0,471. Dat is iets lager dan de gemiddelde vaardigheid van de deelnemers aan de proefafname, die 0,487 bedraagt. Toepassing van de cesuurscore op de behaalde resultaten bij de proefafname leidt tot de conclusie dat bijna 78 procent van de proefpersonen onvoldoende rekenvaardigheid bezit. Het 3F-niveau wordt door 48,5 procent van de proefpersonen niet gehaald.

De omvang van de steekproef, de verdeling van de proefpersonen over meer dan 30 verschillende instellingen en de als voldoende ingeschatte representativiteit doen vermoeden dat dit percentage in de populatie niet veel anders zal zijn. Tabel 8 geeft het percentage personen met onvoldoende rekenvaardigheid voor enkele subgroepen.

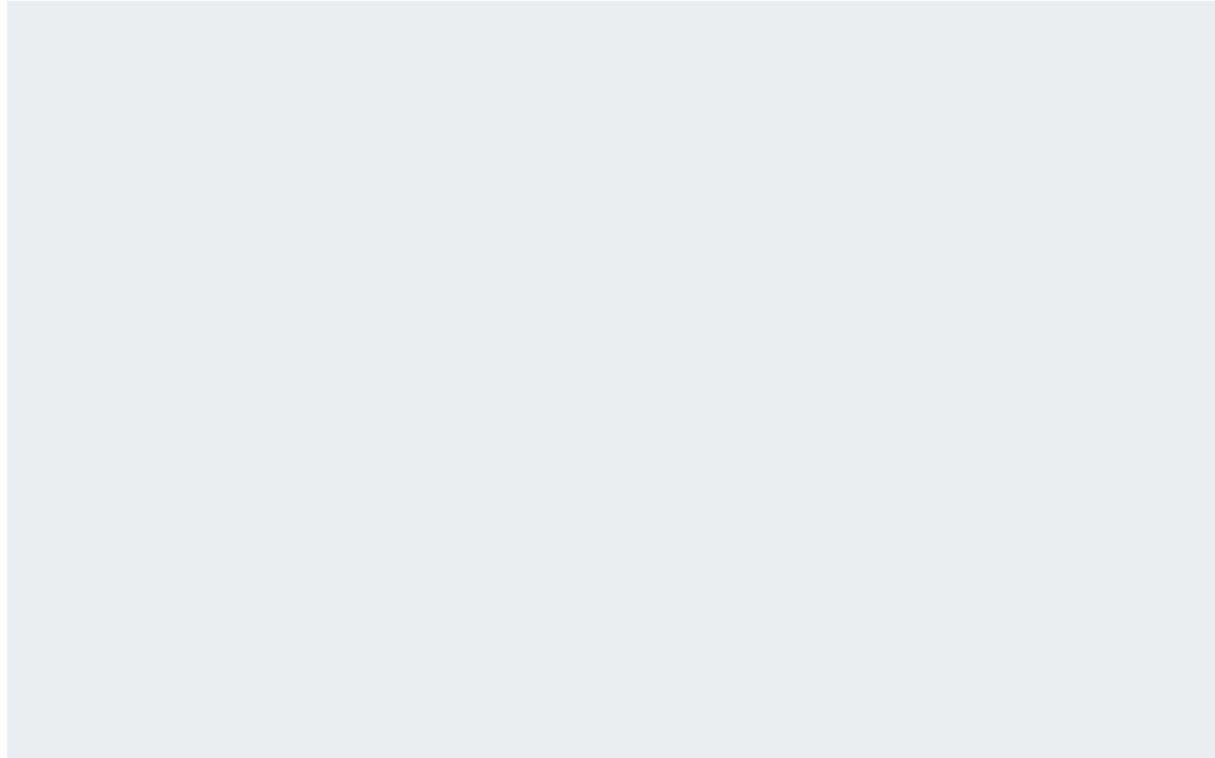
Tabel 8: Percentage personen met onvoldoende rekenvaardigheid bij subgroepen

Geslacht	Man	66,7
	Vrouw	78,6
Vooropleiding	MBO-V	87,3
	In-service A/B/Z	77,3
	HBO-V	66,2
Werkervaring	Weinig (enkele maanden tot 4 jaar)	75,7
	Redelijk (vier of vijf jaar)	79,6
	Veel (zes jaar of meer)	77,6

■ Conclusie en discussie

Dat de rekenvaardigheid van de Nederlandse verpleegkundige onder de maat is, werd in de afgelopen jaren al vaker gesuggereerd. Uit diverse onderzoeken zou gebleken zijn dat verpleegkundigen op dit terrein onvoldoende vaardig zijn om hen met een gerust hart medicaties te laten samenstellen of controleren (de Jong & Koster, 2007; DUO Market Research, 2007; DUO Market Research, 2010). Vanuit methodologisch gezichtspunt schieten deze onderzoeken echter tekort voor het doen van stellige uitspraken over de rekenvaardigheid van de verpleegkundige en helemaal als die generaliserend bedoeld zijn. In de eerste plaats omdat het aantal onderzochte proefpersonen meestal te klein is om representatief te kunnen zijn voor de populatie van Nederlandse verpleegkundigen. In de tweede plaats vanwege de onbekende kwaliteit van de rekentoets zelf, waarop uitspraken over de rekenvaardigheid berusten. In de derde plaats omdat de proefpersonen de rekentoets vaak niet gemaakt hebben onder nauwkeurig voorgeschreven (en gehandhaafde) toetscondities. En ten slotte door het ontbreken van een kritische score op grond waarvan de toetsprestaties geëvalueerd kunnen worden in termen van onvoldoende of voldoende rekenvaardigheid.

Onderhavige studie kent deze tekortkomingen niet. Er vanuit gaande dat de vastgestelde cesuur een correcte representatie is van de minimaal vereiste rekenvaardigheid, moet geconstateerd worden dat vier van de vijf verpleegkundigen in Nederland over onvoldoende rekenvaardigheid beschikt.



Toch zijn er enkele kanttekeningen te plaatsen bij deze conclusie. Ten eerste bij de totstandkoming van de cesuur. Zes vakinhoudelijke experts, die weliswaar verschillende organisaties vertegenwoordigen, is niet veel om een prestatiestandaard te bepalen die landelijk moet functioneren. Bij dergelijke aantallen is er een risico dat een ander panel tot een lagere of hogere cesuurscore komt. In de tweede plaats bij de representativiteit van de toetsopgaven voor de rekentaken waarmee de verpleegkundige in de praktijk geconfronteerd kan worden.

Het feit dat de opgaven gemaakt zijn door verpleegkundigen met affiniteit en deskundigheid op het gebied van verpleegkundig rekenen sluit niet uit dat (een deel van) de opgaven bij (een deel van) de kandidaten een beroep doet op andere cognitieve processen dan de rekentaken in de beroepspraktijk. In dergelijke gevallen zou het bijvoorbeeld kunnen voorkomen dat iemand in de praktijk de juiste doseringen toedient, maar in de meer abstracte context van een rekenopgave niet tot de juiste oplossing komt, of net andersom. Onderzoek dat hier meer duidelijkheid over kan geven is gewenst.

Noten

- ¹ Het discriminerend vermogen geeft aan in welke mate een opgave onderscheid kan maken tussen beheersers en niet-beheersers van de betreffende leerstof.
- ² Bij de categorieën '15 t/m 24 jaar' en '25 t/m 34 jaar' wijken de relatieve frequenties van de CBS-gegevens behoorlijk af van die van de proefafname. Echter als beide categorieën samengenomen worden is de overeenkomst redelijk te noemen. Deze samenvoeging is acceptabel gelet op het feit dat de gemiddelde rekenvaardigheden van beide leeftijdscategorieën nagenoeg gelijk zijn (15 t/m 24 jr: 0,605; 25 t/m 34 jr: 0,604).
- ³ De getallen die we op deze schaal gebruiken, hebben geen absolute betekenis. Vergelijkbaar met bijvoorbeeld het meten van temperatuur zijn het nulpunt en de meeteenheid vrij te kiezen. Zoals gebruikelijk (Verhelst, 1993) is er gekozen voor een schaal waarbij de gemiddelde moeilijkheidsgraad van de opgaven gelijk is aan 0.
- ⁴ Door ontbrekende gegevens is de som van het aantal mannen en vrouwen afwijkend van het totale aantal personen dat aan de proefafname heeft meegedaan. Soortgelijke afwijkingen doen zich ook voor in tabel 6 en 7.

Literatuur

- Baker, F.B. (1985). *The basics of item response theory*. Portsmouth (NH): Heinemann.
- Cizek, G.J. (2006). Standard setting. In S.M. Downing & T. M. Haladyna (Eds.), *Handbook of test development*. (pp. 225-258). Mahwah (NJ): Lawrence Erlbaum Associates.
- COTAN (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- DUO Market Research (2007). *Rapportage rekenvaardigheid verpleegkundigen 2007*. Utrecht: DUO Market Research.
- DUO Market Research (2010). *Rapportage rekenvaardigheid verpleegkundigen 2010*. Utrecht: DUO Market Research.
- Hoes, L. (2005). *Verpleegkundige hoofdbreken: Rekenvaardigheid voor verbetering vatbaar*. Nursing, december, 32-35.
- Jong, C.W. de, & Koster, A.P. (2007). *Rekenvaardigheid van verpleegkundigen. Een cross-sectioneel onderzoek in vier ziekenhuizen naar de competentie in de verschillende dimensies*. Doctoraalscriptie Gezondheidswetenschappen. Maastricht: Universiteit van Maastricht.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Sijtsma, K. (2009). *Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven*. De Psycholoog, november, 561-567.
- Van der Velden, L.F.J. (1996). *Context, visie, aanpak en effectiviteit. De bestrijding van achterstanden van Nederlandse leerlingen in het basisonderwijs*. Academisch proefschrift. Groningen: R.U.
- Verhelst, N.D. (1993). *Itemresponstheorie*. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de Praktijk*. Arnhem: Cito, 83-178.

* **Drs. Tecla T.M. Lampe** is werkzaam bij Cito Arnhem. * **Dr. Gerard J.J.M. Straetmans** is lector Assessment bij Saxion en daarnaast werkzaam bij Cito Arnhem. * **Prof. Dr. Ir. Theo J.H.M. Eggen** is werkzaam bij Cito en hoogleraar bij Universiteit Twente.