

ADAPTIEF TOETSEN IN DE BASIS EDUCATIE

G.J.J.M. Straetmans
T.J.H.M. Eggen

Cito

ADAPTIEF TOETSEN IN DE BASIS-EDUCATIE

G.J.J.M. Straetmans

T.J.H.M. Eggen

Cito

Arnhem, april 1996

Prijs *f* 15,--

© Cito Arnhem 1996

Uit deze uitgave mogen zonder toestemming van de auteurs korte aanhalingen met volledige bronvermelding worden opgenomen.

Inhoudsopgave

Inleiding

1	Context van het Project	9
2	De 'Papieren' Plaatsingstoets	11
2.1	Afbakening van Onderwijsniveaus	11
2.2	Constructieproblemen	12
2.3	Tweefasen Toets	12
3	CAT: Computergestuurde Adaptieve Toets	15
3.1	Enkele Methodes voor Itemselectie	15
3.1.1	Schatten als rekenprocedure	16
3.1.2	Toetsen als rekenprocedure	18
3.2	Enkele Methodes voor de Beëindiging van een Toetsafname	19
4	Gecomputeriseerd Adaptief Toetsen voor het Classificeren van Kandidaten in drie Categorieën: Verslag van een Simulatiestudie	21
4.1	Inleiding	21
4.2	Context	22
4.2.1	De itembank Rekenen/Wiskunde	23
4.3	Onderzoeksvragen	24
4.4	Statistische Rekenprocedures	24
4.4.1	Statistisch schatten in het toetsalgoritme	25
4.4.2	Statistisch toetsen in het toetsalgoritme	26
4.5	Itemselectieprocedures	29
4.5.1	Startprocedure	30
4.5.2	Vijf selectieprocedures	30
4.6	Opzet Simulatiestudies	32
4.7	De Implementatie van de Toetsalgoritmen	34
4.7.1	Implementatie itemselectieprocedures	35
4.7.2	Globale inhoud programma's	35
4.8	Resultaten Simulatiestudies	36
4.8.1	De meetnauwkeurigheid bij statistisch schatten	36
4.8.2	Het functioneren van de toetsalgoritmen onder afname-omstandigheden die kenmerkend zijn voor een plaatsingstoets	38
4.8.3	Gegevens over het gebruik van de itembank	43
4.9	Conclusie	48

5	Machinale Antwoordverwerking bij 'On-Line Testing'	51
6	De Ontwikkeling van een Instructiemodule voor Kandidaten	57
7	De behoefte aan Adaptief Toetsen in de Volwasseneneducatie	81
7.1	Samenvatting	81
7.2	Inleiding	81
7.3	Methode	82
7.3.1	Respondenten	82
7.3.2	Materiaal	83
7.3.3	Resultaten	83
7.3.4	De respons	84
7.3.5	Aantal cursisten	85
7.3.6	Toetsgebruik	86
7.3.7	Herkomst toetsen	87
7.3.8	Tijdsbesteding ten behoeve van toetsontwikkeling en toetsgebruik	89
7.3.9	Opinie over het belang van de kenmerken van adaptief toetsen	94
7.3.10	Computerinfrastructuur	101
7.4	Conclusies en Discussie	105
7.4.1	Conclusies	105
7.4.2	Discussie	106
	Literatuur	109

Inleiding

In 1995 is op het Cito (Instituut voor Toetsontwikkeling te Arnhem) gewerkt aan de ontwikkeling van software en procedures die het mogelijk moeten maken in de Basiseducatie, voor het onderdeel rekenen/wiskunde, adaptief te toetsen. Het belangrijkste voordeel van adaptief toetsen ten opzichte van de meer conventionele toetsprocedures is de grotere efficiëntie: het samenstellen, afnemen en nakijken van toetsen alsmede het rapporteren van de resultaten gaan sneller bij adaptief toetsen. Deze grotere efficiëntie op het gebied van toetsen is een voorwaarde voor het realiseren van flexibel onderwijs, een voor de Basiseducatie (en voor de Volwasseneneducatie in het algemeen) noodzakelijk geacht organisatieprincipe.

Dit rapport doet verslag van de werkzaamheden die in 1995 zijn uitgevoerd ten behoeve van de ontwikkeling van een adaptief toetspakket voor rekenen/wiskunde. De primaire doelstelling van het project was niet om te komen tot een verkoopprodukt, maar om ervaring op te doen met de ontwikkeling van dergelijke gecompliceerde instrumenten en om na te gaan in hoeverre de genoemde voordelen werkelijk kunnen worden waargenomen in een praktische gebruikssituatie. Activiteiten in het kader van het laatstgenoemde doel staan gepland voor 1996. Dit verslag gaat vooral in op de activiteiten die zijn uitgevoerd om de instrumenten en toetsprocedures te ontwikkelen voor adaptief toetsen. Tevens wordt verslag gedaan van een onderzoek onder instellingen in de Volwasseneneducatie dat tot doel had het gebruik van toetsen voor het onderdeel rekenen/wiskunde te inventariseren, de interesse voor adaptief toetsen te peilen en de computer-infrastructuur in kaart te brengen.

1 Context van het Project

De Basiseducatie (BE) is een sinds 1987 bestaande vorm van volwasseneneducatie. De BE is onder andere voortgekomen uit het projectenbeleid (Open School, Alfabetisering) dat de overheid in de jaren zeventig voerde om de permanente educatie gestalte te geven. De BE richt zich op volwassenen, al dan niet met de Nederlandse nationaliteit, die het primair onderwijs of de eerste fase van voortgezet onderwijs niet doorlopen hebben of daarvan te weinig hebben opgestoken. Dit betekent dat de cursisten van de BE een grote heterogeniteit laten zien met betrekking tot het niveau en de aard van de leerbehoefte.

De BE kent geen formeel vastgesteld curriculum: de deelnemers stellen in overleg met educatief werkers hun eigen programma samen. Dat geldt niet alleen voor de inhoud en didactiek maar ook voor het beoogde eindniveau, het tempo, de frequentie van de lessen, de plaats en de tijd.

Taal (Nederlands; Nederlands als tweede taal; alfabetisering in de eigen taal; Fries; Engels), rekenen/wiskunde en sociale kennis en vaardigheden vormen de drie hoofdelementen van de BE. Volgens de Rijksregeling moeten deze hoofdelementen in samenhang worden aangeboden, alhoewel accenten gelegd kunnen worden.

De BE heeft globaal twee opleidingstrajecten waarvan de een zich richt op maatschappelijke redzaamheid en de ander op professionele redzaamheid. Het traject maatschappelijke redzaamheid beoogt volwassenen beter te laten functioneren in hun eigen situatie. Professionele redzaamheid is een opleidingstraject waarin het cursus aanbod is afgestemd op doorstroming naar vervolgonopleidingen of maatschappelijke activiteiten.

In een instelling als de BE, waar cursisten met zeer verschillende achtergronden naar toe komen om binnen afzienbare termijn zeer uiteenlopende doelen te bereiken, bestaat grote behoefte aan toetsing. Toetsing vervult daarbij verschillende functies.

Bij binnenkomst moet worden vastgesteld wat een cursist al weet van een bepaald vakgebied. Op grond van die informatie wordt hij toegewezen aan een van de niveaus (voor de meeste gebieden worden drie à vier niveaus onderscheiden) waarop het onderwijs verzorgd wordt.

Tijdens het onderwijsproces moet de voortgang bewaakt worden, zodat bij stagnering tijdig passende maatregelen getroffen kunnen worden. Dit betekent dat er regelmatig getoetst moet worden in hoeverre de doelstellingen van het onderwijs bereikt zijn.

Onder invloed van de sociaal-economische ontwikkelingen leggen de cursisten steeds meer nadruk op de maatschappelijke relevantie van het cursusaanbod. Daardoor groeit de behoefte aan voorgestructureerde leerplannen en daarop afgestemd leermateriaal (Dijkman, 1993). De overheid speelt hierop in door te streven naar certificaten voor de cursussen die worden aangeboden. Sinds kort worden daarom eindtermen vastgesteld voor het BE-cursusaanbod. Om te bepalen welke cursisten de eindtermen bereikt hebben, zijn afsluitende toetsen nodig.

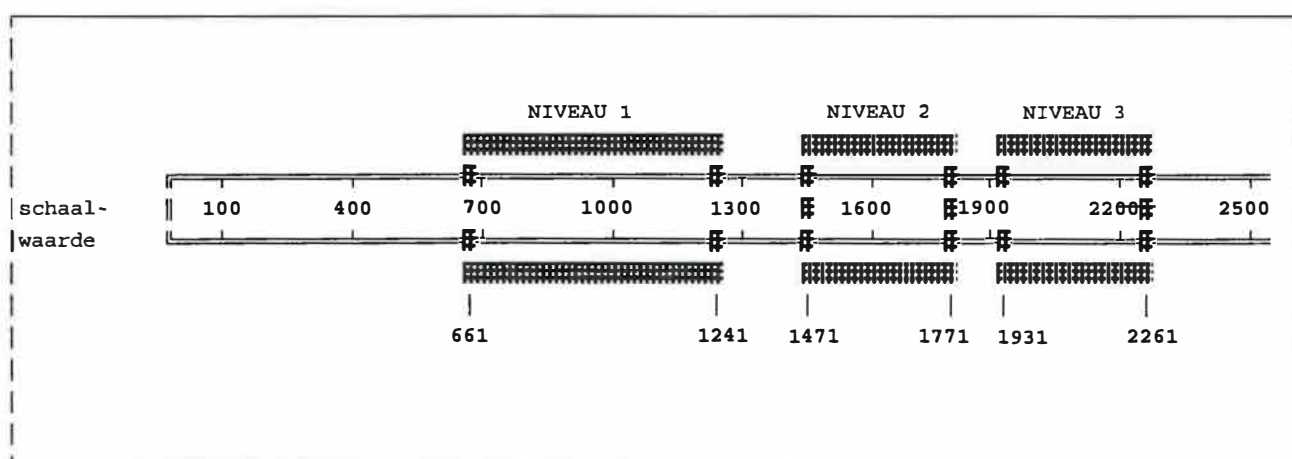
Het construeren van goede toetsopgaven kost veel tijd en vergt naast vakinhoudelijke expertise ook veel deskundigheid en ervaring op het gebied van toetsconstructie. Het aantal mensen dat beschikt over deze combinatie van deskundigheden en daarnaast tijd vrij kan maken voor de constructie, is relatief gering. Dit dwingt tot bezinning op een meer efficiënt gebruik van geproduceerde toetsopgaven. Waarom zouden goede toetsopgaven niet meerdere malen gebruikt mogen worden? Het is zonde van de tijd als elke instelling voor BE op eigen gelegenheid min of meer dezelfde toetsen produceert. Efficiënter is het als op centraal niveau een grote verzameling toetsopgaven beschikbaar is waaruit toetsen kunnen worden samengesteld voor plaatsing, voortgangsbewaking en certificering. Om in deze behoefte te voorzien werkt het Cito sinds enige tijd aan de ontwikkeling van itembanken voor de onderdelen rekenen/wiskunde, Nederlands als tweede taal en Nederlands als moedertaal. De ontwikkelingen bij rekenen/wiskunde lopen voor op die van de andere onderdelen. Bij het verschijnen van dit rapport bestaat de itembank rekenen/wiskunde uit 250 items (Straetmans, 1993; Straetmans, 1994a) en wordt er gewerkt aan een uitbreiding met nog eens 450 items. Daarnaast is subsidie aangevraagd voor een uitbreiding van de itembank met items die de doelstellingen van rekenen/wiskunde in hogere (meer beroepsgerichte) opleidingen van de volwasseneneducatie kunnen meten.

2 De 'Papieren' Plaatsingstoets

2.1 Afbakening van Onderwijsniveaus

Uit de itembank is ondermeer een plaatsingstoets (plaatsingstoets in het jargon van de volwasseneneducatie) samengesteld. De plaatsingstoets wordt afgenomen aan instro- mende cursisten met het doel deze te verwijzen naar het meest geschikte onderwijs- niveau voor rekenen/wiskunde. Het rekenonderwijs in de BE is cyclisch van opbouw. Dat wil zeggen dat de meeste onderwerpen op elk niveau aan de orde komen, maar dat op hogere niveaus dieper op elk onderwerp wordt ingegaan.

De plaatsingstoets is samengesteld uit een geschaalde itembank. De items in zo'n bank zijn via een mathematisch model (zie paragraaf 4.2.1 voor een specificatie van het gebruikte model) aan elkaar gerelateerd. Toetsen die uit een geschaalde itembank zijn samengesteld en worden afgenomen aan kandidaten leveren scores op die omgezet kunnen worden in vaardigheidsschattingen op dezelfde schaal als waarop de items qua moeilijkheidsgraad geplaatst zijn. Om op grond van een vaardigheidsschatting te kunnen beslissen of een cursist naar niveau 1, 2 of 3 verwezen moet worden, is het noodzakelijk dat ook deze niveaus op de rekenvaardigheidsschaal afgebeeld zijn. Dat wil zeggen dat duidelijk moet zijn tussen welke schaalwaarden de onderscheiden niveaus liggen. In Figuur 1 is de afbakening van de niveaus grafisch weergegeven.



Figuur 1

Afbakening van de onderwijsniveaus op de rekenvaardigheidsschaal

De afbakening van de niveaus is gebaseerd op het principe van de beheersing van de doelstellingen. Het criterium voor beheersing is bepaald op 70%. Een cursist bijvoorbeeld met een geschatte vaardigheid van 1471 (net niveau 2) zal ten minste 7 van de 10 opgaven correct beantwoorden die strikt toevallig getrokken zijn uit een verzameling opgaven van niveau 1.

2.2 Constructieproblemen

Een belangrijke randvoorwaarde voor de constructie van de plaatsingstoets rekenen/wiskunde had betrekking op de lengte van de toets. Rekening houdend met de beperkt beschikbare tijd voor de plaatsing van nieuwe cursisten werd het maximum aantal items, in overleg met de opdrachtgever, bepaald op 25. Echter, een direct gevolg van het gebruik van korte toetsen is dat de meetnauwkeurigheid relatief beperkt blijft.

Een ander probleem bij de constructie van de plaatsingstoets was de onbekendheid met de kennis en vaardigheid van instromende cursisten op het betreffende vakgebied. Bij plaatsingstoetsen ten behoeve van het reguliere onderwijs kan een toetsconstructeur er vaak van uitgaan dat potentiële kandidaten grosso modo hetzelfde onderwijstraject gevolgd hebben. In de volwasseneneducatie is de onderwijsachtergrond van instromende cursisten zeer divers en bovendien vaak niet relevant omdat de onderwezen kennis en vaardigheden grotendeels vergeten zijn. Voor de toetsconstructeur betekent dit dat hij min of meer 'blind' de moeilijkheidsgraad van een plaatsingstoets moet bepalen. Voor de zekerheid zal hij vaak besluiten de moeilijkheidsgraad van de items te laten variëren van zeer eenvoudig tot zeer moeilijk. Maar hoe efficiënt is zo'n toets? Als het niveau van de toets veel hoger is dan de vaardigheid van de cursist, dan levert de toets relatief weinig informatie op over de vaardigheid van die cursist. Er kan dan alleen geconstateerd worden dat de toets te moeilijk was voor hem/haar. Als het niveau van de toets veel lager is dan de vaardigheid van de cursist, dan is er een soortgelijk probleem: bijna alle opgaven zullen goed gemaakt worden, maar een precieze indicatie over de vaardigheid valt daaruit niet af te leiden.

2.3 Tweefasen Toets

Een oplossing voor beide hierboven besproken problemen werd gevonden in de zogeheten tweefasen toets (Lord, 1971). Een tweefasen toets is een adaptieve toets waarbij de afname

zich voltrekt in twee fasen. Voor de plaatsingstoets rekenen/ wiskunde is dit idee als volgt uitgewerkt. In de eerste fase wordt een voor alle kandidaten identieke '*sorteertoets*' afgenomen, bestaande uit 15 items met een gemiddelde moeilijkheidsgraad die is afgestemd op de gemiddelde rekenvaardigheid van BE-cursisten. In de tweede fase krijgt een kandidaat, afhankelijk van de behaalde score op de sorteertoets één van de drie verschillende '*meettoetsen*' te maken. De drie meettoetsen bestaan elk uit 10 items maar verschillen van elkaar in de gemiddelde moeilijkheidsgraad. Kandidaten die op de sorteertoets de lagere scores behalen krijgen de eenvoudigste meettoets te maken; kandidaten met de hogere scores krijgen de moeilijkste meettoets voorgelegd; de overige kandidaten krijgen een meettoets die qua moeilijkheidsgraad overeenkomt met die van de sorteertoets. Meer gedetailleerde informatie over de plaatsingstoets rekenen/wiskunde is te vinden bij Straetmans (1994b).

Het gebruik van tweefasen toetsen kent de volgende nadelen:

- de afname is als gevolg van het noodzakelijke, tussentijdse beoordelingsmoment erg bewerkelijk;
- er is maar één adaptatie-moment, namelijk na beantwoording van het vijftiende item. Als dan een verkeerde beslissing wordt genomen (te eenvoudige of te moeilijke meet- toets) dan zal dit vrijwel zeker leiden tot een lagere meetnauwkeurigheid.

Adaptieve toetsen die via de computer worden afgenomen kennen dit probleem niet. Tezamen met het streven van de Volwasseneneducatie (VE) naar flexibel onderwijs en flexibel toetsen, de steeds betere computer-infrastructuur in de VE-instellingen en de beschikbaarheid van steeds meer geschaalde itembanken waren dit belangrijkste argumenten voor de ontwikkeling van een computergestuurde adaptieve plaatsingstoets rekenen/wiskunde voor de Basiseducatie.

3 CAT: Computergestuurde Adaptieve Toets

Het meest in het oog springende kenmerk van deze vorm van adaptief toetsen is dat de kandidaat 'aan de computer' een toets maakt. Een essentiële onderscheid echter heeft betrekking op het feit dat er op voorhand geen toets klaar ligt; die wordt tijdens de afname samengesteld door de computer. De responsen van de kandidaat zijn bepalend voor de keuze van de volgende aan te bieden opgave.

In zijn algemeenheid kan de werking van een CAT beschreven worden aan de hand van de volgende vier stappen:

- 1 Er wordt een item geselecteerd uit een geschaalde itembank.
- 2 De computer beoordeelt het antwoord als correct of fout.
- 3 Op basis van alle tot dan toe beschikbare responsen van een kandidaat voert de computer rekenprocedures uit die leiden tot schattingen van de vaardigheid (in het geval 'schatten' gebruikt wordt als rekenprocedure: zie par. 3.1.1) of tot een toetsingsgrootte (in het geval 'toetsen' gebruikt wordt als rekenprocedure: zie par. 3.1.2).
- 4 Op basis van de uitkomst in stap 3 wordt ofwel een nieuw item geselecteerd of de toets beëindigd.

De stappen 1 tot en met 4 worden in de literatuur samengevat onder de term 'toetsalgoritme' (Wainer et al, 1990). Bij de ontwikkeling van een CAT moeten deze stappen nader uitgewerkt worden. Voor stap 1 en 4 geldt dat er diverse methodes zijn waaruit gekozen kan worden. Veel van die methodes staan in de literatuur beschreven en/of zijn in simulatiestudies of in praktijksituaties getest. Echter de onderzoeksresultaten zijn specifiek voor de gebruikte itembank en de populatie waarvoor de bank ontwikkeld is. Dit betekent dat voor elke nieuwe toepassing simulatiestudies noodzakelijk zijn waarin de verschillende varianten van een toetsalgoritme met elkaar vergeleken worden op efficiëntie, meetnauwkeurigheid en eventuele andere criteria. Hieronder worden enkele methodes voor itemselectie en voor het beëindigen van een toetsafname besproken.

3.1 Enkele Methodes voor Itemselectie

Er zijn verschillende methodes voor de selectie van items uit een itembank. Alle zijn gebaseerd op de voorlopige, op grond van de gegeven antwoorden verkregen, kennis over

de vaardigheid van de getoetste persoon. De selectie van een item wordt gebaseerd op de uitkomsten van een statistische rekenprocedure. In het project worden twee verschillende statistische rekenprocedures gebruikt: schatten en toetsen. Achtereenvolgens worden beide kort toegelicht.

3.1.1 Schatten als rekenprocedure

Een tweetal populaire methoden die uitgaan van een schatting van de vaardigheid zijn:

- een methode die items selecteert die de grootste bijdrage leveren aan de toets- informatie gegeven de voorlopige geschatte vaardigheid van een kandidaat (maximum information of **MaI**);
- een methode die items selecteert die de Bayesiaanse a posteriori standaarddeviatie rondom de vaardigheidsschatting minimaliseert (minimum a posteriori standaarddeviatie of **MiSD**).

Het belangrijkste verschilpunt tussen beide methoden betreft het gebruik in de Bayesiaanse methode van informatie over de verdeling van theta in de doelgroep. Beide methoden leveren dezelfde resultaten als er sprake is van een uniforme (prior) verdeling van theta in de doelgroep. Over het algemeen is de Bayesiaanse methode wat nauwkeuriger kort na de start van de toets. Doordat rekening wordt gehouden met de verdeling van theta zullen de vaardigheidsschattingen minder sterk fluctueren.

Beide methoden selecteren op elk punt in de toets het beste item zonder daarbij acht te slaan op vakinhoudelijke of andere aspecten. Dit roept een aantal problemen op. Drie daarvan worden hieronder aan de orde gesteld.

Acceptatieproblemen bij toetsgebruikers en kandidaten

In handboeken voor toetsconstructie wordt altijd gewezen op de noodzaak dat de items in een toets een goede afspiegeling moeten zijn van de doelstellingen van het onderwijs dat de toets wil evalueren. Echter, de hierboven genoemde selectiemethodes kunnen geenszins de constructie van representatieve toetsen garanderen. Op elk punt in de toetsafname wordt dát item gekozen dat de grootste bijdrage levert aan de toetsinformatie of dat de a posteriori standaarddeviatie rondom de vaardigheidsschatting minimaliseert. Het gevolg hiervan kan zijn dat de totale toets te weinig representatief is voor de doelstellingen. Vanuit psychome-

trisch oogpunt is dit nauwelijks een probleem. Adaptieve toetsen opereren op een unidimensionele itembank waarvan elk item dezelfde latente vaardigheid meet. Dit impliceert dat de vaardigheid van een kandidaat geschat kan worden op basis van een willekeurige set van items. Maar toetsgebruikers (docenten of andere verantwoordelijke personen) en kandidaten kunnen grote moeite hebben met eenzijdig samengestelde toetsen.

De maatregelen die genomen worden om de selectiestrategieën qua inhoud 'aan banden te leggen' worden in de literatuur aangeduid met de term 'content balancing'. Deze maatregelen moeten voorkomen dat bepaalde doelstellingen of onderwerpen te vaak getoetst worden. Enige bekendheid geniet de methode voor content-balancing van Kingsbury en Zara (1989). Deze methode vereist dat voorafgaand aan een toetsafname het percentage items per subdomein gespecificeerd wordt dat in een toets moet worden opgenomen. Tijdens de toetsafname worden de volgende stappen uitgevoerd na elk beantwoord item (uitgegaan wordt van de MaI selectiestrategie):

- per subdomein wordt het percentage (t.o.v. alle tot dan toe afgenomen items) beantwoorde items berekend;
- deze empirische percentages worden vergeleken met de vooraf gespecificeerde percentages en het subdomein met de grootste discrepantie wordt geïdentificeerd;
- binnen dit subdomein wordt het item geselecteerd dat de meeste informatie kan bijdragen.

Context effect

Een tweede probleem heeft betrekking op het effect dat eerder afgenomen items kunnen hebben op de beantwoording van volgende items. Bij conventionele toetsen treden deze effecten ook op maar gemiddeld genomen in gelijke mate voor elke kandidaat. Bij een CAT echter krijgt elke kandidaat in principe een andere toets te maken waardoor context effecten van kandidaat tot kandidaat kunnen verschillen.

Eenzijdige uitputting

Het laatste hier te noemen probleem betreft de eenzijdige uitputting van de itembank. Vooral in relatief kleine itembanken dreigt dit gevaar doordat de gehanteerde selectiestrategie steeds dezelfde items kiest voor kandidaten van wie de ware vaardigheid ongeveer gelijk is. Dit levert een geheimhoudingsprobleem op, vooral omdat adaptieve toetsen vaak weinig items bevatten. Daarnaast vormen de zelden of nooit gekozen items een bron van verspilling van schaarse middelen.

Maatregelen die de gebruikte selectiestrategie dwingen om ook andere dan de meest optimale items te kiezen, worden in de literatuur aangeduid met de term **exposure control** methodes. De meest bekende zijn die welke zich baseren op algoritmen van Sympton en Hetter (1985). De essentie is dat een item pas gekozen wordt als een random gegenereerd getal kleiner of gelijk is aan een vooraf in te stellen 'exposure control parameter' (ECP). De ECP kan gelden voor de hele toets of moet voor elk afzonderlijk item in de bank ingesteld worden. De ECP's moeten door middel van een iteratief proces van gesimuleerde toetsing en bijstelling bepaald worden.

De meeste toetsalgoritmen in een CAT selecteren het volgende item op basis van de laatste, voorlopige schatting van de vaardigheid van de kandidaat. Het adaptieve karakter van deze toetsen schuilt primair in de aanpassing van de moeilijkheidsgraad van het volgende te presenteren item aan de voorlopige geschatte vaardigheid van de kandidaat. Volgens Reckase (1983) en Kingsbury en Weiss (1983) echter is het voor toetsen die tot doel hebben te bepalen of een kandidaat onder of boven een bepaalde aftestgrens (dat is de toetsscore die minimaal behaald moet worden voor het nemen van een beslissing over de kandidaat) zit efficiënter items te presenteren waarvan de moeilijkheidsgraad zo dicht mogelijk in de buurt ligt van deze aftestgrens. Het adaptieve karakter van deze CATs wordt dan vooral gevormd door de stopregel. Na elk item wordt bekeken of er met de vooraf gespecificeerde nauwkeurigheid een beslissing genomen kan worden over de kandidaat. Als dit het geval is, wordt de toets beëindigd.

3.1.2 Toetsen als rekenprocedure

Deze procedure komt in aanmerking als het toetsresultaat gebruikt wordt om een beslissing te nemen over de kandidaat, bijvoorbeeld een plaatsingsbeslissing of een beslissing over zakken of slagen. De methode is gebaseerd op de door Wald (1947) ontwikkelde sequential

probability ratio test (SPRT). De essentie is dat er een indifferentiegebied gelegd wordt rondom de aftestgrens (een punt op de thetaschaal dat de grens vormt tussen twee verschillende vaardigheidsniveaus). Alleen als de statistische toetsing aangeeft dat de vaardigheid onder of boven het indifferentiegebied ligt, wordt een classificatiebeslissing genomen. In alle andere gevallen wordt er een nieuw item gepresenteerd. Voor een plaatsingstoets houdt deze procedure het volgende in:

- Na elk beantwoord item wordt de aannemelijkheid berekend van de waargenomen respons (goed of fout) gegeven een theta die a) gelijk is aan de ondergrens en b) gelijk is aan de bovengrens van het indifferentiegebied.
- Vervolgens wordt de ratio berekend van deze twee aannemelijkheden (aannemelijkheid gegeven de theta van de bovengrens gedeeld door aannemelijkheid gegeven de theta van de ondergrens). Als deze ratio voldoende groot is, wordt de beslissing van toewijzing aan het hogere niveau genomen. Als de ratio voldoende klein is, volgt toewijzing aan het lagere niveau. In alle andere gevallen kan er nog niet met voldoende zekerheid een beslissing genomen worden en wordt de toets vervolgd.
- Wat onder 'voldoende groot of klein' verstaan moet worden, is afhankelijk van de risico's die men bereid is te lopen op fouten van de eerste soort (accepteren van H_1 terwijl H_0 waar is) en de tweede soort (accepteren van H_0 terwijl H_1 waar is). De nulhypothese houdt bijvoorbeeld in dat theta kleiner of gelijk is aan de theta van de ondergrens; de alternatieve hypothese dat theta groter of gelijk is aan de theta van de bovengrens. De toetsconstructeur dient in samenspraak met zijn opdrachtgever de kans op deze verkeerde beslissingen in een percentage uit te drukken.

3.2 Enkele Methodes voor de Beëindiging van een Toetsafname

In het algemeen kunnen drie verschillende stopregels voor een CAT onderscheiden worden:

- 1 Stop de toetsafname als het van tevoren bepaalde aantal items is beantwoord. Deze stopregel wordt in de praktijk niet veel gebruikt omdat hij afbreuk doet aan de efficiëntie van adaptief toetsen. Wel wordt deze regel veel gebruikt bij onderzoek dat tot doel heeft de werking van verschillende toetsalgoritmen met elkaar te vergelijken.
- 2 Stop de toetsafname als de vaardigheid van de kandidaat met een vooraf bepaalde mate van nauwkeurigheid geschat is.
- 3 Stop de toetsafname als de ware vaardigheid met een vooraf bepaalde mate van zekerheid onder of boven de aftestgrens ligt.

4 Gecomputeriseerd Adaptief Toetsen voor het Classificeren van Kandidaten in drie Categorieën: Verslag van een Simulatiestudie¹

4.1 Inleiding

Gecomputeriseerd adaptief toetsen (CAT) op basis van itemresponstheorie (IRT) wordt tegenwoordig veel toegepast, terwijl ook het psychometrisch onderzoek ten behoeve van adaptief toetsen weer volop in de belangstelling staat (Wainer, 1990). Veel nieuwe onderzoeksvragen komen voort uit nieuwe toepassingen en door de testpraktijk geuite wensen.

Het hart van een CAT wordt gevormd door het toetsalgoritme. Het toetsalgoritme, bestaande uit een itemselectieprocedure en een statistische rekenprocedure, regelt de start, de voortgang en de beëindiging van een adaptieve testafname. Voor die afname dient een IRT gecalibreerde itembank ter beschikking te staan. Met behulp van methoden uit de IRT wordt per individuele kandidaat de samenstelling van de toets tijdens het testen aangepast aan de tot dan toe bekende vaardigheid, totdat deze vaardigheid of een beslissing met gespecificeerde nauwkeurigheid kan worden gerapporteerd. Het hoofddoel van CAT, efficiënt meten, wordt daarmee bereikt. De in het toetsalgoritme gehanteerde criteria zijn psychometrisch van aard, hetgeen met name voor de itemselectie mogelijk nadelige praktische consequenties heeft. Als itemselectie plaatsvindt met het gebruikelijke maximale informatie criterium is zowel de inhoud van een toets als de frequentie van het gebruik van items uit de itembank niet onder controle. In recent psychometrisch onderzoek zijn hiervoor oplossingen gesuggereerd. Kingsbury en Zara (1989, 1991) hebben een procedure voorgesteld die er voor zorgt dat elke toets aan bepaalde inhoudsspecificaties voldoet; dat wil zeggen dat een door de praktijk gewenste verdeling van het aantal items over inhoudelijke subdomeinen van de itembank wordt gerealiseerd. Zowel deze inhoudscontrole als ook de afnamecontrole wordt in het toetsalgoritme gerealiseerd door restricties op te leggen aan het oorspronkelijke maximale informatie criterium voor itemselectie. Afnamecontrole, vooral onderzocht door Simpson en Hetter (1985) en Stocking en Swanson (1993), kan betrekking hebben op twee problemen met maximale informatie selectiemethoden: items uit de bank kunnen

¹ Dit hoofdstuk is geschreven door Ir. T.J.H.M. Eggen naar aanleiding van de door hem uitgevoerde simulatiestudies.

te vaak (overexposure) of te weinig (underutilization) gebruikt worden bij adaptief toetsen. Overexposure kan een probleem met de geheimhouding van de items tot gevolg hebben. Underutilization is zonde gelet op de hoge kosten die gemoeid zijn met de constructie van een itembank.

Alhoewel adaptief toetsen traditioneel gericht is op het efficiënt schatten van de vaardigheid van kandidaten, is het ook voor classificatieproblemen, maar dan met name voor twee categorieën (bijvoorbeeld slagen/zakken), mogelijk en zinvol gebleken adaptief te toetsen (Spray & Reckase, 1994).

In deze studie wordt gerapporteerd over onderzoek ten behoeve van de ontwikkeling van een CAT, waarvan de functie is kandidaten te classificeren in één van drie categorieën. Twee mogelijke statistische rekenprocedures, gebaseerd op toetsen en op schatten, in combinatie met itemselectiemethoden, die ook voorzien in een inhouds- en of afnamecontrole, worden beschreven en geëvalueerd.

4.2 Context

In Nederland heeft de Basiseducatie voor volwassenen als doel volwassenen met grote educatieve achterstanden kennis en vaardigheden te laten verwerven die tenminste nodig zijn om persoonlijk en maatschappelijk te kunnen functioneren. Eén van de cursussen die met dit doel wordt gegeven is rekenen/wiskunde. Deze cursus wordt aangeboden op drie niveaus. Potentiële cursisten worden door middel van een plaatsingstoets aan een bepaald niveau toegewezen. De huidige schriftelijke plaatsingstoets is vanwege de grote verschillen in vaardigheid van de kandidaten een tweefasen toets (Lord, 1971). In de eerste fase maken alle kandidaten een sorteertoets van 15 items met een moeilijkheidsgraad die afgestemd is op de gemiddelde vaardigheid van de potentiële cursisten. Afhankelijk van de score op de sorteertoets krijgen de kandidaten een van drie verschillende meettoetsen, elk bestaande uit 10 items, maar qua moeilijkheid verschillend en passend bij het resultaat op de sorteertoets. De huidige schriftelijke plaatsingstoets kent een aantal bezwaren, die samengevat betrekking hebben op de bewerkelijke afnameprocedure, de geheimhouding van de opgaven en de beperkte meetnauwkeurigheid voor grote groepen kandidaten. Het invoeren van een CAT ter vervanging van de schriftelijke toets wordt overwogen om de bezwaren weg te nemen.

4.2.1 De itembank Rekenen/Wiskunde

Voor de adaptieve toets staat een itembank ter beschikking bestaande uit 250 dichotoom scorebare kort antwoord vragen. De items zijn gecalibreerd met het OPLM model (Verhelst, Glas, Verstralen, 1995). De basisvergelijking van dit model is

$$p_i(\theta) = P(X_i = 1 | \theta) = \frac{\exp a_i(\theta - \beta_i)}{1 + \exp a_i(\theta - \beta_i)} \quad (1)$$

De respons op een item x_i is goed (1) of fout (0). De kans op het goed maken van een item is een stijgende functie van de latente vaardigheid θ en is afhankelijk van twee itemkenmerken: de itemmoeilijkheid, β_i en het discriminerend vermogen van een item a_i . Naar inhoud is de bank opgedeeld in drie subdomeinen: hoofdrekenen/ schatten (A), meten/meetskunde (B) en de rest van de leerstofonderdelen (C). Het resultaat van de calibratie is dat de verdeling van de vaardigheid θ in de populatie normaal is met gemiddelde van .294 en een standaarddeviatie van .522. De verdeling van de items over de geschatte itemmoeilijkheden en de drie subdomeinen staat in Tabel 1.

Tabel 1
Verdeling van de items over moeilijkheid en inhoud

moeilijkheid	inhoud			totaal
	A	B	C	
$\hat{\beta}_i \leq -1$	2	1	7	10
$-1 < \hat{\beta}_i \leq -.5$	13	5	19	37
$-.5 < \hat{\beta}_i \leq 0$	18	11	37	66
$0 < \hat{\beta}_i \leq .5$	13	23	66	102
$\hat{\beta}_i > .5$	2	9	24	35
totaal	48	49	153	250

Op de latente vaardigheidsschaal zijn door inhoudelijke beoordeling van de items de grenspunten voor de classificatie van de kandidaten bepaald. Het grenspunt tussen

niveautoewijzing 1 en 2 is $\theta_1 = -.13$ en tussen 2 en 3 is dat $\theta_2 = .33$. Globale inspectie van de itembank leert dat er een goede verdeling is van de moeilijkheid over de latente vaardigheidsschaal, met name in de buurt van de grenspunten is een ruime hoeveelheid items aanwezig.

Voor meer details over de context, de constructie van de itembank en de bepaling van de grenspunten wordt verwezen naar andere hoofdstukken in dit rapport en naar Straetmans (1994a, 1994b).

4.3 Onderzoeksvragen

De globale onderzoeksvraag die in dit hoofdstuk wordt behandeld is: welk toetsalgoritme is de beste keus voor de gecomputeriseerde adaptieve plaatsingstoets voor rekenen/wiskunde in de Basiseducatie voor volwassenen, waarbij rekening gehouden wordt met een aantal praktische randvoorwaarden? Evidentie, vanuit het meetperspectief, wordt gezocht voor de vervanging van de huidige schriftelijke toets door een CAT. Een harde randvoorwaarde daarbij is dat een individuele afname bij een CAT maximaal 25 items lang mag zijn. De mogelijkheden voor inhouds- en afnamecontrole dienden onderzocht te worden. Meer specifiek waren de vragen: Welke statistische rekenprocedures komen in aanmerking voor het classificatieprobleem in drie niveaus? En in relatie daarmee: Welke itemselectiemethoden komen in aanmerking? Hoe functioneren de toetsalgoritmen in termen van meetnauwkeurigheid, het aantal misclassificaties, meetefficiëntie, het voldoen aan inhoudelijke specificaties en de verdeling van het gebruik van de items uit de itembank?

4.4 Statistische Rekenprocedures

De functie van een statistische rekenprocedure in het toetsalgoritme is dat deze op basis van de itemresponsen de beslissing over de geteste kandidaat geeft. Deze inferentie wordt gepleegd aan de hand van de aannemelijkheidsfunctie van de vaardigheid θ van de kandidaat. Gegeven de scores $X_1=x_1, X_2=x_2, \dots, X_k=x_k$ op k items en de parameters van de items, $a_1, \beta_1, a_2, \beta_2, \dots, a_k, \beta_k$, is deze functie:

$$L_k(\theta; x_1, \dots, x_k) = L_k(\theta; \mathbf{x}) = \prod_{i=1}^k p_i(\theta)^{x_i} (1 - p_i(\theta))^{1-x_i}. \quad (2)$$

Voor $p_i(\theta)$ wordt het OPLM model, formule (1), gesubstitueerd.

Er wordt na elk item besloten of gestopt wordt met testen en welke uitslag in dat geval aan de kandidaat wordt gegeven, of dat er nog een item moet worden afgenomen. Voor bovengenoemd classificatieprobleem zijn daarvoor globaal twee benaderingen: statistisch schatten en statistisch toetsen. Beide procedures worden kort beschreven.

4.4.1 Statistisch schatten in het toetsalgoritme

Voor adaptief toetsen met statistisch schatten is een traditionele benadering gekozen (Weiss en Kingsbury, 1984). Gegeven de scores op k items, $X_1=x_1, X_2=x_2, \dots, X_k=x_k$, en de parameters van de items, $a_1, \beta_1, a_2, \beta_2, \dots, a_k, \beta_k$, wordt een schatting van de vaardigheid $\hat{\theta}_k$ en van haar standaardfout $se(\hat{\theta}_k)$ bepaald. Vervolgens wordt een betrouwbaarheidsinterval geconstrueerd voor de ware vaardigheid θ van de kandidaat: $(\hat{\theta}_k - \gamma.se(\hat{\theta}_k), \hat{\theta}_k + \gamma.se(\hat{\theta}_k))$.

γ is een te specificeren constante afhankelijk van de gewenste nauwkeurigheid. Het toetsalgoritme besluit tot het afnemen van een nieuw item zolang er een grenspunt θ_1 of θ_2 in het interval ligt, zo niet dan wordt er een beslissing genomen volgens de beslissingsregels in Tabel 2.

Tabel 2
Beslissingstabel adaptieve toets met statistisch schatten

als	besluit
$\hat{\theta}_k + \gamma.se(\hat{\theta}_k) < -.13$	niveau 1
$\hat{\theta}_k - \gamma.se(\hat{\theta}_k) > -.13$ en $\hat{\theta}_k + \gamma.se(\hat{\theta}_k) < .33$	niveau 2
$\hat{\theta}_k - \gamma.se(\hat{\theta}_k) > .33$	niveau 3
anders	doorgaan met toetsen

Vanwege zijn goede statistische eigenschappen is voor de schatting van de vaardigheid na k items de gewogen grootste aannemelijkheid (WML) methode van Warm (1989) toegepast. Gegeven de itemscores en de itemparameters volgt deze schatting en haar standaardfout uit een iteratieve maximalisatie procedure :

$$\hat{\theta}_k = \max_{\theta} \left(\sum_{i=1}^k I_i(\theta) \right)^{1/2} \cdot \prod_{i=1}^k p_i(\theta)^{x_i} (1-p_i(\theta))^{1-x_i}. \quad (3)$$

Het tweede deel van deze uitdrukking is de aannemelijkheidsfunctie van de vaardigheid (2), gegeven de scores op de items en de itemparameters. Het eerste deel is het gewicht dat aan deze aannemelijkheidsfunctie wordt gegeven. In deze uitdrukking is de zogenaamde iteminformatiefunctie, $I_i(\theta)$, opgenomen. Deze functie geeft de informatie die bevat is in een item als functie van de vaardigheid. De bijdrage van een item aan de nauwkeurigheid van de schatting van de vaardigheid van een kandidaat hangt positief samen met deze iteminformatie. In het OPLM model (1) is de iteminformatiefunctie gegeven door

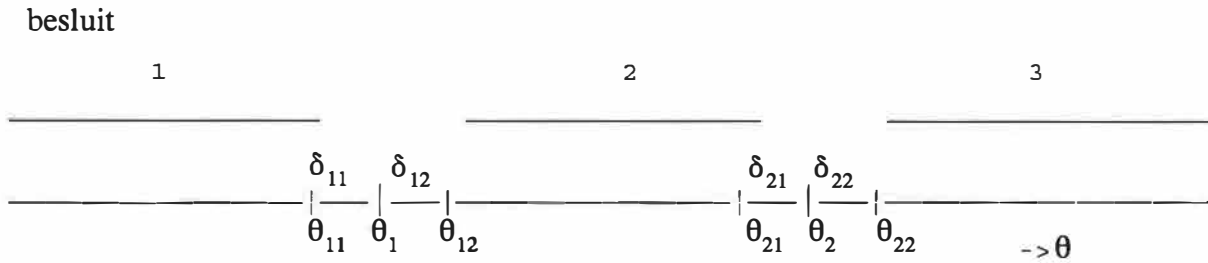
$$I_i(\theta) = a_i^2 p_i(\theta)(1-p_i(\theta)). \quad (4)$$

Voor details over de achtergronden van deze schatter en de wijze waarop deze uitgerekend wordt, wordt verwezen naar Warm (1989) en Verhelst en Kamphuis (1989). De nauwkeurigheid van deze schattingsprocedure in het toetsalgoritme wordt bepaald door het in te stellen niveau van het betrouwbaarheidsinterval.

4.4.2 Statistisch toetsen in het toetsalgoritme

Als alternatief voor de traditionele schattingsaanpak kan in het classificatieprobleem ook een statistische toetsingsprocedure gebruikt worden. Voorgesteld wordt een veralgemening van een eerder door Reckase (1983) voor adaptief toetsen toegepaste procedure, die gebaseerd is op de Sequential Probability Ratio Test (SPRT) van Wald (1947).

Eerst worden zogenaamde indifferentiezones δ gedefinieerd. Dit zijn smalle gebieden rondom de grenspunten θ_1 en θ_2 , waarin, vanwege de feilbaarheid van de metingen, nooit met zekerheid een juiste beslissing genomen kan worden. Daarna worden de statistische hypothesen geformuleerd en de acceptabele kansen op foute beslissingen vastgesteld. In Figuur 1 is het probleem schematisch weergegeven.



Figuur 1
Schematische weergave toetsingsprobleem

De hypothesen luiden:

$H0_1: \theta \leq \theta_{11}$ (niveau 1) $H0_2: \theta \leq \theta_{21}$ (lager dan 3)

$H1_1: \theta \geq \theta_{12}$ (hoger dan 1); $H1_2: \theta \geq \theta_{22}$ (niveau 3).

De acceptabele kansen op foute beslissingen worden, met α_1 , α_2 , β_1 en β_2 te kiezen kleine constanten (niet te verwarren met de eerder met hetzelfde symbool gedefinieerde itemparameters), gespecificeerd door:

$$P(\text{ac. } H0_1 | \theta \leq \theta_{11}) \geq 1 - \alpha_1 \quad P(\text{ac. } H0_2 | \theta \leq \theta_{21}) \geq 1 - \alpha_2$$

$$P(\text{ac. } H0_1 | \theta \geq \theta_{12}) \leq \beta_1 \quad P(\text{ac. } H0_2 | \theta \geq \theta_{22}) \leq \beta_2$$

Voor elk paar hypothesen ($H0_1$ tegen $H1_1$; $H0_2$ tegen $H1_2$) kan deze toetsing, terwijl voldaan wordt aan de nauwkeurigheidseisen, worden uitgevoerd met de SPRT (Wald, 1947). Als toetsingsgrootte wordt de verhouding van de waarden van de aannemelijkheidsfunctie (2) in de nulhypothese en het alternatief gebruikt. Het toetsingsvoorschrift voor $H0_1$ tegen $H1_1$ is:

Ga door met waarnemen als (ook wel de kritieke ongelijkheid statistische toets genoemd):

$$\frac{\beta_1}{1 - \alpha_1} < \prod_{i=1}^k \frac{L_k(\theta_{12}; x)}{L_k(\theta_{11}; x)} < \frac{1 - \beta_1}{\alpha_1}; \quad (5)$$

accepteer HO_1 (niveau 1) als:

$$\prod_{i=1}^k \frac{L_k(\theta_{12}; \underline{x})}{L_k(\theta_{11}; \underline{x})} \leq \frac{\beta_1}{1-\alpha_1}$$

verwerp HO_1 (niveau 2 of 3) als:

$$\prod_{i=1}^k \frac{L_k(\theta_{12}; \underline{x})}{L_k(\theta_{11}; \underline{x})} \geq \frac{1-\beta_1}{\alpha_1}.$$

Door combinatie van de twee SPRT's, zoals weergegeven in Figuur 2, kunnen voor het onderhavige classificatieprobleem eenduidige beslissingen worden geformuleerd.

besluit test 2	besluit test 1	
	1	2 of 3
1 of 2	1	2
3	onmogelijk	3

Figuur 2
Besluiten bij combinatie van twee SPRT's

Deze veralgemening van de SPRT naar een dubbele SPRT is in de literatuur bekend als de combinatieprocedure van Sobel en Wald (1949). Eenvoudig is aan te tonen dat bij gebruik van het OPLM model de onmogelijke oplossing inderdaad nooit optreedt en dat de kritieke ongelijkheid (5) geschreven kan worden als:

$$\frac{\ln \frac{\beta_1}{1-\alpha_1} - C_{k\theta_{11}\theta_{12}}}{\theta_{12} - \theta_{11}} < \sum_{i=1}^k a_i x_i < \frac{\ln \frac{1-\beta_1}{\alpha_1} - C_{k\theta_{11}\theta_{12}}}{\theta_{12} - \theta_{11}},$$

waarbij $C_{k\theta_{11}\theta_{12}}$ een uitdrukking is die alleen afhangt van de itemparameters en de van te voren te kiezen constanten in de toetsingsprocedure (θ_{11} en θ_{12}). Na te gaan is dat de evaluatie van deze kritieke ongelijkheid zeer eenvoudig uit te voeren is, omdat ze

niet meer behelst dan die van een vergelijking van de geobserveerde gewogen score met constanten.

Het beslissingsschema voor de dubbele sequentiële procedure bij de plaatsingstoets rekenen/wiskunde, als we voor de eenvoud aannemen dat $\alpha_1=\alpha_2=\beta_1=\beta_2=\alpha$, $\delta_{11}=\delta_{12}=\delta_{21}=\delta_{22}=\delta$ en $\ln(1-\alpha)/\alpha = A$ staat in Tabel 3.

Tabel 3
Beslissingstabel adaptieve toets met statistisch toetsen

als	besluit
$\sum_{i=1}^k a_i x_i \leq \frac{-A - \sum_{i=1}^k \ln \left(\frac{p_i(-.13-\delta)}{p_i(-.13+\delta)} \right)}{2\delta}$	niveau 1
$\frac{A - \sum_{i=1}^k \ln \left(\frac{p_i(-.13-\delta)}{p_i(-.13+\delta)} \right)}{2\delta} \leq \sum_{i=1}^k a_i x_i \leq \frac{-A - \sum_{i=1}^k \ln \left(\frac{p_i(.33-\delta)}{p_i(.33+\delta)} \right)}{2\delta}$	niveau 2
$\sum_{i=1}^k a_i x_i \geq \frac{A - \sum_{i=1}^k \ln \left(\frac{p_i(.33-\delta)}{p_i(.33+\delta)} \right)}{2\delta}$	niveau 3
anders	doorgaan met toetsen

4.5 Itemselectieprocedures

In het toetsalgoritme zorgt de itemselectieprocedure ervoor dat items worden gekozen bij de tot dan toe van de getoetste persoon bekende vaardigheid, die volgt uit de rekenprocedure. Een aparte positie wordt daarbij ingenomen door de startprocedure omdat er vanuit gegaan wordt dat voor de eerste afname nog niets over de vaardigheid van de persoon bekend is. Hierna wordt eerst de startprocedure en vervolgens de bij

de plaatsingstoets rekenen/wiskunde geïmplementeerde reeks van selectieprocedures kort beschreven.

4.5.1 Startprocedure

De startprocedure die bij de plaatsingstoets rekenen/wiskunde ontwikkeld is, werkt als volgt. Uit de itembank van 250 items is een selectie gemaakt van 54 relatief gemakkelijke items. Uit elk van de drie inhoudelijke subdomeinen krijgt een kandidaat aselect één relatief gemakkelijk item aangeboden. Er waren drie redenen om te kiezen voor deze startprocedure. De belangrijkste is dat de doelgroep van kandidaten, die ten dele bestaat uit laag-opgeleide mensen die niet vertrouwd zijn met de computer, door aanbieding van gemakkelijke startitems wellicht minder last zullen hebben van eventuele toetsangst en computerangst. De tweede reden is dat het in enige zin optimaal kiezen van items bij de vaardigheid van de getoetste persoon na één of twee items nauwelijks mogelijk is, omdat de eerste schattingen van de vaardigheid noodzakelijkerwijs erg onnauwkeurig zullen zijn. Tenslotte draagt de gekozen startprocedure door selectie uit de drie verschillende subdomeinen bij aan de inhoudelijke representativiteit van de toets voor het getoetste domein rekenen/wiskunde.

4.5.2 Vijf selectieprocedures

In samenhang met de rekenprocedures zijn bij de plaatsingstoets vijf itemselectieprocedures onderzocht. Deze worden aangeduid met:

- 1 aselect (R);
- 2 maximale informatie (MI);
- 3 maximale informatie met inhoudscontrole (MI+C);
- 4 maximale informatie met afnamecontrole (MI+E);
- 5 maximale informatie met inhoudscontrole en afnamecontrole (MI+C+E).

De eerste procedure trekt aselect, met uitsluiting van eerder afgenomen items, uit de beschikbare bank het volgende item. De overige procedures kiezen een item voor een kandidaat zodanig dat de informatie van een item, zie (4), bij die kandidaat maximaal is. Maximale informatie kan daarbij de volgende drie mogelijke betekenissen hebben.

Bij statistisch schatten als rekenprocedure:

- a Als volgend item wordt het item geselecteerd waarvan de informatie bij de lopende vaardigheidsschatting van de kandidaat maximaal is. Dit wordt hierna aangegeven met de CE (current estimate). Kies het item i waarvoor: $\max_i I_i(\hat{\theta}_k)$.

- b Spray en Reckase (1994) tonen aan dat het bij classificatieproblemen met één grenspunt bij adaptief toetsen efficiënter (resulterend in gemiddeld kortere toetslengte) is om items te kiezen die de maximale informatie hebben bij dat grenspunt, in plaats van bij de lopende vaardigheidsschatting. Naar analogie hiervan is de volgende selectieprocedure gebruikt. Kies het item dat maximale informatie heeft bij het grenspunt dat het dichtst ligt bij de lopende vaardigheidsschatting: het minimum wordt bepaald van $|-.13 - \hat{\theta}_k|$ en $|.33 - \hat{\theta}_k|$. Deze optie duiden we aan met NC (nearest cutting point).

Bij statistisch toetsen als rekenprocedure worden er geen vaardigheidsschattingen bepaald, zodat daar een variant van b wordt gebruikt.

- c Bepaald wordt bij welke van de twee grenzen van de statistische toets (zie Tabel 3) de huidige score van een kandidaat het dichtst ligt. Gekozen wordt het item dat maximale informatie heeft bij het daarbijbehorende grenspunt.

Maximale informatie is een psychometrisch criterium voor de itemselectie. Door de praktijk worden er doorgaans ook andere wensen aan de toetssamenstelling geuit, die door restricties op dit psychometrisch criterium kunnen worden gerealiseerd. Naar twee van deze restricties is onderzoek gedaan.

De eerste betreft inhoudscontrole: een algoritme zorgt er voor dat de adaptieve toets aan bepaalde inhoudsspecificaties voldoet. Bij de adaptieve plaatsingstoets rekenen/wiskunde is de inhoudscontrole gerealiseerd door vooraf te specificeren dat het wenselijk is dat een toets voor respectievelijk 16% en 20% uit items over de subdomeinen hoofdrekenen (A) en meten/meetkunde (B) zou bestaan en voor 64% uit items voor de overige onderdelen (C). Voor de realisatie hiervan is de benadering van Kingsbury en Zara (1989, 1991) gevolgd. Het geïmplementeerde algoritme bepaalt na elke itemafname het verschil tussen het gewenste en het bereikte percentage items per subdomein. Uit het subdomein waarvoor dit verschil het grootst is, wordt vervolgens het item met maximale informatie gekozen.

De tweede onderzochte restrictie is afnamecontrole. De achtergrond van de afnamecontrole is dat in de praktijk van adaptief toetsen, alhoewel elke kandidaat doorgaans een andere toets krijgt, het vaak voorkomt, dat sommige items vaak en veel items uit de beschikbare itembank niet gebruikt worden. De eenvoudigste variant van deze controle, die bij de plaatsingstoets rekenen/wiskunde is gebruikt, zorgt ervoor dat de beschikbare itembank beter benut wordt, doordat een op grond van maximale informatie gekozen item, slechts in de helft van de gevallen wordt afgenomen. In het algoritme wordt, nadat een item gekozen is, een aselekt getal g uit het interval $(0,1)$ getrokken. Als $g > .5$ wordt het item gekozen, zo niet dan wordt de procedure voortgezet met het kiezen van het volgende meest informatieve item. Items die eenmaal afgewezen zijn door deze controle worden niet meer gekozen voor de kandidaat. Voor

een uitvoerige beschrijving van andere vormen van afnamecontrole wordt verwezen naar Stocking en Swanson (1993).

Ook het simultaan opleggen van beide restricties, afname- en inhoudscontrole, is bij de plaatsingstoets onderzocht.

4.6 Opzet Simulatiestudies

Het functioneren van de besproken reken- en itemselectieprocedures zijn bij de plaatsingstoets onderzocht door middel van simulatiestudies. De algemene opzet was als volgt.

Uit de populatie verdeling, bekend uit het calibratie-onderzoek, $N(.294, .522)$, werd een kandidaat v , dat wil zeggen zijn vaardigheid θ_v , aselekt getrokken. De drie start-items werden gekozen volgens de hiervoor beschreven startprocedure en de volgende items op grond van één van de itemselectieprocedures. Het antwoord van de kandidaat op een item wordt gegenereerd volgens het IRT model. Meer specifiek: bij elke afname werd aselekt een getal g uit het interval $(0,1)$ getrokken. Voor kandidaat v en item i werd formule (1) geëvalueerd en als $p_i(\theta_v) \geq g$ werd het item als "goed" gescoord, $x_i=1$, zo niet dan "fout", $x_i=0$. Deze procedure werd in de simulaties voor $N=5000$ kandidaten herhaald.

In de simulatiestudies zijn de beschreven procedures beoordeeld en vergeleken op enkele hierna te bespreken aspecten. Het toetsalgoritme met schatting als rekenprocedure werd onderzocht op de met de plaatsingstoets te bereiken meetnauwkeurigheid. Door niet op een vooraf ingestelde nauwkeurigheid de gesimuleerde afname te stoppen, maar na een absoluut aantal van 50 items, werd nagegaan wat de verschillen waren in de te bereiken nauwkeurigheid tussen de toetsalgoritmen als functie van het aantal afgenomen items. Gerapporteerd wordt de onnauwkeurigheid van de meting na k items, dat wil zeggen het gemiddelde absolute verschil tussen de ware en de geschatte vaardigheid:

$$LA_k(\theta) = \frac{1}{N} \sum_{v=1}^N |\theta - \hat{\theta}_{k,v}|. \quad (6)$$

Op grond van de resultaten hiervan kon mede bepaald worden of de absolute bovengrens van maximaal 25 items zinvol was.

Om het functioneren van de adaptieve toets met de verschillende toetsalgoritmen in de toepassingssituatie van de plaatsingstoets te beoordelen werden simulaties uitgevoerd, die als stopregel een nauwkeurigheidseis hanteerden met een maximaal aantal af te nemen items van $k_{\max}=25$. Indien dit maximale aantal items nodig is, wordt

afgeweken van de beslissingstabellen 2 en 3 in de zin dat dan de meest voor de hand liggende beslissing wordt genomen. Bij schatten en toetsen als rekenprocedure wordt de beslissing dan genomen zoals vermeld in Tabel 4.

Tabel 4
Beslissingstabel adaptieve plaatsingstoets rekenen/wiskunde
met schatten en toetsen bij $k_{\max}$

schatten als	toetsen als	besluit
$\hat{\theta}_{k_{\max}} < -.13$	$\sum_{i=1}^{k_{\max}} a_i x_i \leq \frac{-\sum_{i=1}^{k_{\max}} \ln \left(\frac{p_i(-.13-\delta)}{p_i(-.13+\delta)} \right)}{2\delta}$	1
$-.13 \leq \hat{\theta}_{k_{\max}} < .33$	$\frac{-\sum_{i=1}^{k_{\max}} \ln \left(\frac{p_i(-.13-\delta)}{p_i(-.13+\delta)} \right)}{2\delta} < \sum_{i=1}^{k_{\max}} a_i x_i < \frac{-\sum_{i=1}^{k_{\max}} \ln \left(\frac{p_i(.33-\delta)}{p_i(.33+\delta)} \right)}{2\delta}$	2
$\hat{\theta}_{k_{\max}} \geq .33$	$\sum_{i=1}^{k_{\max}} a_i x_i \geq \frac{-\sum_{i=1}^{k_{\max}} \ln \left(\frac{p_i(.33-\delta)}{p_i(.33+\delta)} \right)}{2\delta}$	3

De volgende resultaten worden gerapporteerd: de classificatienauwkeurigheid, het gemiddeld aantal benodigde items, de frequentieverdeling van de afnames van items uit de itembank en de verdeling van de afgenomen items naar subdomein.

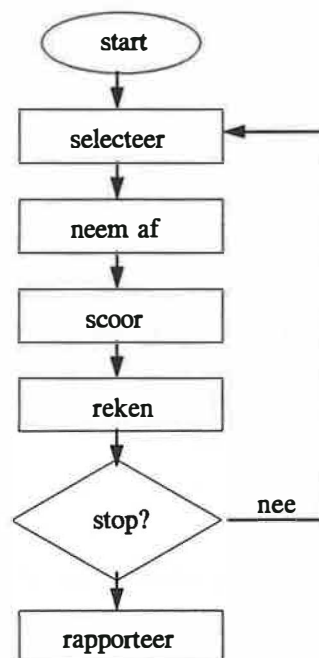
Bij de rekenprocedure "schatten" worden twee verschillende nauwkeurigheden onderzocht: γ is 1.034 en 1.644 (zie Tabel 2), hetgeen respectievelijk 70 en 90% betrouwbaarheidsintervallen geeft. Bij de rekenprocedure "toetsen" varieerden de acceptabele kansen op foute beslissingen (zie Tabel 3): α is .05, .075 en .1. Daarnaast werd bij $\alpha = .075$ de indifferentiezone gevarieerd: δ is .1 en .1333.

Om na te gaan voor welke vaardigheden eventuele verschillen tussen de toetsings- en schattingsprocedure optreden, werd naast de algemene simulatie opzet ook de volgende variant gebruikt. In plaats van aselekt trekken uit de populatieverdeling werden op

67 equidistante punten van de vaardigheid, $-1.1 \leq \theta \leq 1.1$, 500 toetsafnames gesimuleerd.

4.7 De Implementatie van de Toetsalgoritmen

Deze paragraaf geeft een beschrijving van de manier waarop de onderdelen van het toetsalgoritme, zoals die paragraaf 4.4 en 4.5 beschreven zijn, geïmplementeerd zijn in de computerprogramma's die de adaptieve toetsafnames simuleren. De computerprogramma's doorlopen herhaald de in Figuur 2a weergegeven stappen die de elementen vormen van het toetsalgoritme in een adaptieve toetsafname.



Figuur 2a
Schematische weergave van een adaptieve toetsafname

Alvorens de globale inhoud van de programma's te beschrijven, wordt eerst ingegaan op de geïmplementeerde itemselectieprocedure.

4.7.1 Implementatie itemselectieprocedures

In de toetsalgoritmen is er voor gekozen de itemselectie op basis van informatietabellen (Wainer, 1990) te laten plaatsvinden. In elke itemselectiestap worden maximaal informatieve items, eventueel met nadere restricties, bij bepaalde vaardigheids punten gekozen. In principe is het mogelijk om na elke afname een maximalisatiealgoritme toe te passen op de uit te rekenen iteminformaties voor alle nog beschikbare items in de bank. Dit is echter vrij rekenintensief en bovendien inefficiënt, omdat dit voor elke nieuwe persoon zou moeten gebeuren waardoor veel exact dezelfde berekeningen gemaakt moeten worden. Bij het werken met informatietabellen wordt dit vermeden. De tabel wordt ten behoeve van en voorafgaand aan de simulaties of adaptieve toetsafnames in één keer op basis van de parameters uit de itembank en met de toets te bereiken meetnauwkeurigheid gemaakt. De tabel bevat voor een aantal vaardigheids punten een reeks itemnummers uit de itembank in volgorde van grootte van de iteminformatie op dat punt. Bij de rekentoets bevat de informatietabel 73 rijen: voor θ is -1.50 tot 2.00 in stappen van .05 en voor de grenspunten -.13 en .33.; in elke rij staan de 250 itemnummers uit de bank in volgorde van grootte van de iteminformatie. De keuze van minimale en maximale θ is gebaseerd op het calibratieonderzoek, waaruit is gebleken dat de minimale en maximale vaardigheidsschatting van personen uit de steekproef respectievelijk -1.3 en 1.9 waren. De stapgrootte is zodanig gekozen dat er geen verlies van nauwkeurigheid kan optreden: vastgesteld is dat bij afname van 100 items uit de bank de minimale standaardfout van de vaardigheidsschatting .057 bedraagt. Gegeven het toetsdoel en de beschreven itemselectieprocedures zal het opnemen van rijen in de tabel voor de grenspunten evident zijn.

4.7.2 Globale inhoud programma's

Twee basisprogramma's zijn geschreven. De eerste voor adaptief toetsen met schatten, de tweede voor adaptief toetsen met toetsen als rekenprocedure. De programma's hebben de volgende gemeenschappelijke elementen:

- 1 Er zijn twee invoer-itembankbestanden: de startbank en de gehele itembank. Deze bevatten: uniek itemnummer, code voor inhoudscategorie, de itemparameters (a_i en $\hat{\beta}_i$). De startbank bevat de startselectie van relatief gemakkelijke items.
- 2 De hiervoor beschreven informatietabel is ook een invoerbestand .

- 3 De programma's vragen in een dialoog naar de te gebruiken itemselectiemethoden.
- 4 In de programma's is de afnamecontrole parameter een te specificeren constante.
- 5 Voor het trekken van aselechte getallen, het transformeren van deze getallen naar normale afwijkingen en voor diverse sorteringen van getallen is gebruikt gemaakt van routines uit Press, Teukolsky, Vetterling en Flannery (1992).
- 6 De uitvoer van de simulatieprogramma's bevatten de resultaten en de afgenomen items.

De programma's verschillen uiteraard in het rekenalgoritme. Voor het programma met schatten als rekenprocedure is voor het verkrijgen van de vaardigheidsschattingen en de standaardfouten een routine uit het programma OPLM (Verhelst e.a., 1995) enigszins aangepast en gebruikt. De nauwkeurigheidsinstellingen γ (stopregel) zijn als constanten in het programma opgenomen. Het programma met toetsen als rekenprocedure heeft een extra invoerbestand waarin de te toetsen hypothesen, de aanvaardbare beslissingsfouten (α en β) en de indifferentiezone (δ) wordt opgegeven.

4.8 Resultaten Simulatiestudies

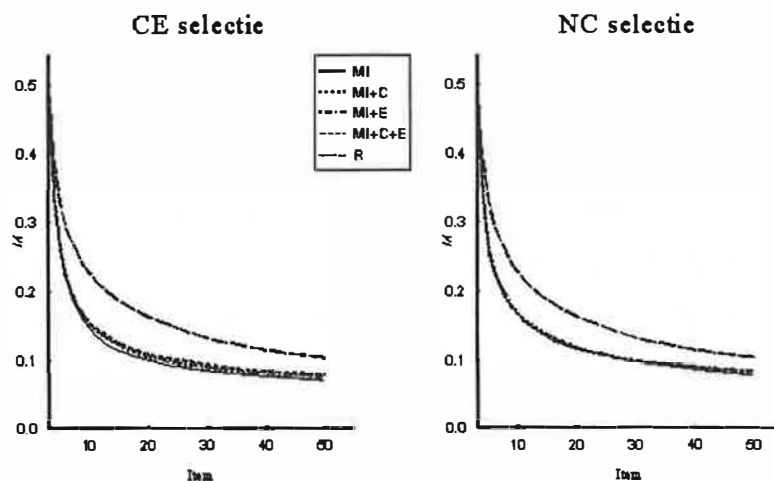
4.8.1 De meetnauwkeurigheid bij statistisch schatten

In Tabel 5 staat de onnauwkeurigheid $IA_k(\theta)$, vermenigvuldigd met 10000, (zie formule (6)) van de meting met de adaptieve toets voor de verschillende itemselectieprocedures na achtereenvolgens 10, 20, 30, 40 en 50 items. Zoals verwacht, neemt bij een toenemend aantal items de onnauwkeurigheid af.

Tabel 5

$IA_k(\theta) \times 10.000$ voor itemselecties met MI bij lopende vaardigheidsschatting (CE) en MI bij dichtstbijzijnde grenspunt (NC)

k	10		20		30		40		50	
selectie	CE	NC	CE	NC	CE	NC	CE	NC	CE	NC
MI	1466	1639	994	1146	834	963	758	847	700	776
MI+C	1519	1674	1037	1166	865	990	771	869	702	804
MI+E	1522	1657	1062	1171	893	983	811	890	754	827
MI+C+E	1541	1652	1082	1162	932	991	835	890	785	822
R	2266		1619		1314		1141		1034	

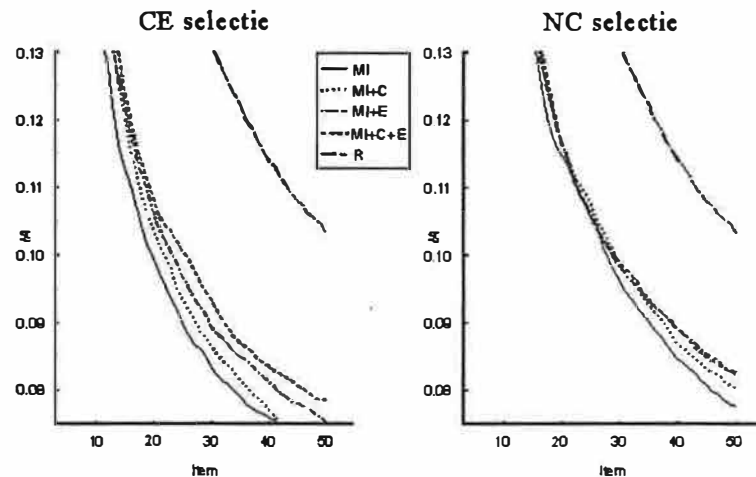


Figuur 3

Onnauwkeurigheid itemselectieprocedures als functie van het aantal items

Het selecteren van items bij lopende vaardigheidsschatting (CE) heeft een kleinere onnauwkeurigheid tot gevolg dan het selecteren bij het dichtstbijzijnde grenspunt (NC). Uit Tabel 5 en Figuur 3 blijkt verder dat de aselechte itemselectie methode tot de grootste onnauwkeurigheid leidt. De afname van de onnauwkeurigheid wordt bij alle overige selectiemethodes zeer klein na ruim 20 items, op grond waarvan geconcludeerd kan worden dat de praktische randvoorwaarde van maximaal 25 af te nemen

items reëel is. Het verschil tussen de itemselectiemethoden is niet groot. In Figuur 4, waarin Figuur 3 op een kleinere schaal is weergegeven, valt echter op dat bij de CE selectie het selecteren zonder restricties de nauwkeurigste schattingen oplevert.



Figuur 4

Onnauwkeurigheid itemselectieprocedures als functie van het aantal items

Toepassing van de afnamecontrole benadeelt de nauwkeurigheid iets meer dan toepassing van de inhoudscontrole. Wanneer beide restricties tegelijk toegepast worden, levert dat het grootste verlies op. Bij de NC selectie zijn de verschillen tussen de selectiemethodes nog geringer. Met name tussen de 20 en 25 items is er qua nauwkeurigheid geen verschil waar te nemen tussen de selectiemethodes.

4.8.2 Het functioneren van de toetsalgoritmen onder afname-omstandigheden die kenmerkend zijn voor een plaatsingstoets

In Tabel 6 staan de resultaten van de simulaties voor de vier itemselectieprocedures, die maximaal informatieve items kiezen, samengevat. Vermeld is het gemiddeld aantal benodigde items (k) om tot een beslissing te komen en het percentage goede beslissingen (%). Of een beslissing voor een kandidaat goed is, kan in de simulatie eenvoudig worden bepaald. Van een kandidaat is de vaardigheid θ bekend en dus ook tot welke subpopulatie (niveau 1, 2 of 3) hij of zij behoort. De beslissing die op grond

van de adaptieve toetsafname wordt genomen, wordt hiermee vergeleken. Ten behoeve van de interpretatie van deze resultaten werd ook de schriftelijke plaatsingstoets gesimuleerd. Het resultaat hiervan was dat het gemiddeld aantal benodigde items, uiteraard, 25 was en het percentage goede beslissingen 87.0%. Verder zij opgemerkt dat, bij de gebruikte steekproefgrootte, verschillen tussen het gemiddelde aantal benodigde items van .6 significant zijn op 99% niveau, terwijl de verschillen tussen percentages goede beslissingen pas significant zijn op dit niveau als dit minstens 3.3 is (2.7 op 95% niveau).

Statistisch schatten

Het variëren van de ingestelde nauwkeurigheid bij schatten als rekenprocedure heeft het volgende effect. Vergroting van de ingestelde nauwkeurigheid, van $\gamma=1.644$ naar $\gamma=1.034$ in Tabel 2, heeft zowel bij de CE selectie als bij NC selectie een significante toename van het gemiddeld aantal benodigde items tot gevolg. Bij de eerstgenoemde selectiemethode varieert deze toename tussen gemiddeld 1.9 en 2.8 items, terwijl dit effect dubbel zo groot is (tussen 3.9 en 4.6) bij selectie bij de NC selectie. Het percentage goede beslissingen neemt doorgaans ook significant toe als de ingestelde nauwkeurigheid verhoogd wordt.

Tabel 6
Gemiddeld aantal benodigde items en percentage goede beslissingen

selectie	MI		MI+C		MI+E		MI+C+E	
	k	%	k	%	k	%	k	%
schatten								
70%-CE	13.8	85.4	13.8	85.7	14.5	85.5	14.2	83.5
90%-CE	16.3	89.1	16.6	88.8	16.4	87.8	16.4	87.7
70%-NC	14.4	88.4	14.8	88.2	14.3	87.4	14.6	85.2
90%-NC	18.7	89.9	19.4	89.8	18.7	87.2	18.5	89.2
toetsen								
5%- $\delta = .1$	17.9	90.9	18.1	88.7	18.6	87.9	19.2	88.7
10%- $\delta = .1$	15.5	89.0	15.4	89.4	15.8	87.1	16.5	86.5
7.5%- $\delta = .1$	16.2	88.5	16.7	89.0	17.1	87.8	17.5	90.4
7.5%- $\delta = .13$	14.2	88.3	13.9	88.5	14.4	89.1	14.9	87.4

Ten opzichte van de schriftelijke versie van de plaatsingstoets neemt het percentage goede beslissingen alleen toe als we met 90% betrouwbaarheidsintervallen werken in het stopcriterium.

Bij vergelijking van CE selectie met NC selectie valt op dat bij 70% betrouwbaarheidsintervallen er een klein verschil is in het gemiddelde aantal benodigde items in het nadeel van de NC selectie, terwijl het percentage goede beslissingen hierbij, soms significant, groter is. Bij 90% evenwel is er een significant voordeel voor CE selectie (tussen 2.1 en 2.8) in het gemiddeld aantal benodigde items, terwijl er geen verschillen zijn in percentages goede beslissingen. Dit is een afwijking van het resultaat dat Spray en Reckase (1994) rapporteren bij een classificatie in twee categorieën.

Onderlinge vergelijking van de itemselectiemethoden geeft bij alle schattingsprocedures nauwelijks verschillen te zien. Significante verschillen zijn er alleen ten opzichte van de aselechte itemselectiemethode, waarbij respectievelijk met 70% en 90% betrouwbaarheidsintervallen de volgende simulatieresultaten werden bereikt: $k = 16.2$ en $\% = 81.7$, $k = 20.7$ en $\% = 83.8$.

Statistisch toetsen

Bij toetsen als rekenprocedure, bij vaste indifferentiezone van $\delta = .1$, stijgt het gemiddelde benodigde aantal items significant met het verlagen van de ingestelde kansen op foute beslissingen. De verschillen bij de foutenkansen van 5% en 10% variëren tussen de 2.4 en 2.8. Tussen de percentages goede beslissingen zijn er, tegen de verwachting in, nauwelijks verschillen. De verklaring hiervoor is dat in relatief veel gesimuleerde afnames een beslissing na 25 items nog niet op grond van de ingestelde foutenkansen kon worden genomen en dat de procedure dan werd afgesloten met het nemen van de meest redelijke beslissing (zie Tabel 4).

Als we het indifferentiegebied vergroten bij acceptabele kansen op foute beslissingen van .075 is een significante afname in het gemiddeld aantal benodigde items te constateren (variërend tussen 2.0 en 2.8), terwijl er geen invloed is op het percentage goede beslissingen. Wordt echter het indifferentiegebied nog verder vergroot dan leidt dit wel tot een significante afname in het percentage goede beslissingen.

Ten opzichte van de schriftelijke plaatsingstoets is bij de gerapporteerde simulaties, waarbij toetsen als rekenprocedure wordt gebruikt, een verbetering (soms significant) in het percentage goede beslissingen te zien.

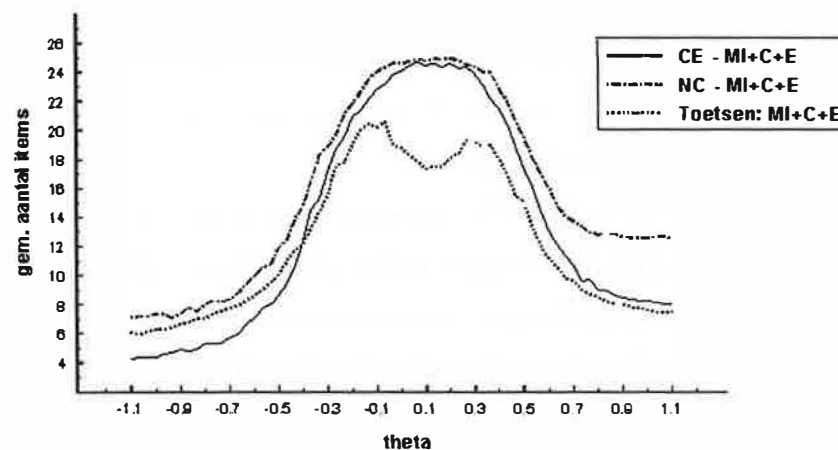
Evenals bij schatten is het verschil bij toetsen tussen de itemselectiemethoden onderling gering: slechts een kleine toename in het gemiddeld aantal benodigde items bij toenemende restricties op de itemselectie is waarneembaar. Indien bij toetsen items aselekt geselecteerd worden, heeft dat wel invloed op de kwaliteit van de procedures: er zijn dan gemiddeld ongeveer 5 items meer nodig en het percentage goede beslissingen daalt ook.

Statistisch schatten en toetsen vergeleken

Bij ongeveer gelijke percentages goede beslissingen, die overigens alle de nauwkeurigheid van de schriftelijke plaatsingstoets halen (87.0%), vergelijken we schatten en toetsen (de drie procedures uit de tweede, vierde en achtste rij in Tabel 6). De globale conclusie die hieruit kan worden getrokken is dat het gemiddeld aantal benodigde items het kleinst is bij het gebruik van statistisch toetsen als rekenprocedure met de volgende kenmerken: de acceptabele kansen op foute beslissingen zijn $\alpha = .075$ en het indifferentiegebied is $\delta = .13$. Deze procedure heeft gemiddeld ruim twee items minder nodig dan de combinatie van schatten als rekenprocedure en selectie van maximaal informatieve items bij de lopende vaardigheidsschatting (CE); nog eens twee

items zijn extra nodig als we schatten en selecteren bij het dichtstbijzijnde grenspunt (NC). Als inhoudscontrole en of afnamecontrole als extra restricties worden opgelegd aan de itemselectieprocedure, worden dezelfde verschillen in het aantal benodigde items verkregen.

Voor de drie genoemde procedures is onderzocht waar in de populatie de winst in het gemiddeld aantal benodigde items behaald wordt. Daarvoor zijn op 67 equidistante punten van de vaardigheid, $-1.1 \leq \theta \leq 1.1$, 500 toetsafnames gesimuleerd. Omdat het resultaat voor alle itemselecties ongeveer hetzelfde is, is het gemiddeld aantal benodigde items als functie van de vaardigheid alleen voor de selectieprocedure met toepassing van zowel afname- als inhoudscontrole weergegeven in Figuur 5.



Figuur 5

Het gemiddeld aantal benodigde items voor selectieprocedure MI+C+E

Voor alle drie procedures geldt dat voor vaardigheden rondom het gemiddelde van de populatieverdeling gemiddeld de meeste items nodig zijn. Bij vaardigheden hoger dan het gemiddelde zijn gemiddeld meer items nodig dan bij lagere vaardigheden. Dit kan te maken hebben met de eerder geconstateerde vulling van de itembank rekenen/wiskunde waarin, zoals eerder geconstateerd, relatief meer relatief gemakkelijke items aanwezig zijn. Meer voor de hand liggend echter is de oorzaak te zoeken in de gehanteerde startprocedure van de adaptieve toets: er wordt gestart met drie gemakkelijke items.

Vergelijking van de procedures laat zien dat voor alle vaardigheden het kiezen van maximaal informatieve items bij de lopende vaardigheidsschatting (CE) een winst oplevert ten opzichte van het kiezen bij het dichtstbijzijnde grenspunt (NC). Bij vergelijking van de efficiëntere schattingsprocedure met de toetsingsprocedure, geeft dit voor de toetsingsprocedure een winst in het gemiddeld aantal benodigde items vooral voor de vaardigheden tussen de grenspunten (-.13 en .33). Alleen bij lage vaardigheden werkt de schattingsprocedure gemiddeld iets sneller.

4.8.3 Gegevens over het gebruik van de itembank

In het voorgaande is de globale conclusie getrokken dat het opleggen van restricties aan het itemselectieproces geen grote gevolgen heeft voor de kwaliteit van de procedures als ze gebruikt worden in de afnamesituatie van de plaatsingstoets. De vraag die nu wordt besproken is, is of de opgelegde restricties ook inderdaad het beoogde effect hebben.

Allereerst wordt in Tabel 7 de verdeling van het itemgebruik uit de itembank gegeven voor het toetsalgoritme met schatten als rekenprocedure, waarbij 90% betrouwbaarheidsintervallen zijn gehanteerd. Per selectiemethode wordt het aantal items vermeld (van de in totaal 250) met de frequentie van gebruik (f) in 1000 simulaties.

Tabel 7
Frequentieverdeling itemgebruik bij schatten (90%) als rekenprocedure

selectie	R	MI		MI+C		MI+E		MI+C+E	
frequentie %		CE	NC	CE	NC	CE	NC	CE	NC
$f=0$	0	132	156	126	156	75	103	53	75
$0 < f \leq 2.5$	0	5	1	11	2	47	21	56	40
$2.5 < f \leq 5$	0	28	28	31	29	26	27	42	36
$5 < f \leq 7.5$	130	26	14	16	14	20	21	17	20
$7.5 < f \leq 10$	68	8	4	18	4	13	11	18	12
$10 < f \leq 15$	43	17	11	11	6	26	11	19	18
$15 < f \leq 20$	9	13	3	5	3	24	26	27	22
$20 < f \leq 25$	0	2	8	13	7	8	11	9	10
$25 < f \leq 30$	0	2	5	4	6	7	4	5	4
$30 < f \leq 40$	0	7	7	7	10	4	10	4	7
$40 < f \leq 50$	0	4	1	4	2	0	5	0	5
$50 < f \leq 75$	0	6	9	3	5	0	0	0	1
$75 < f \leq 100$	0	0	3	1	6	0	0	0	0

Het beste gebruik van de itembank wordt gemaakt bij de aselechte itemselectieprocedure: alle items worden in tussen de 5% en 20% van de afnames gebruikt. De kwalitatief betere procedures maken eenzijdiger gebruik van de itembank. Bijvoorbeeld bij selectie van maximaal informatieve items bij de lopende vaardigheidsschattingen, worden 132 items nooit gebruikt, terwijl er 21 items in meer dan 20% van de afnames worden gebruikt. Zowel het feit dat er weinig items uit de bank worden gebruikt, als dat sommige items vaak worden gebruikt zou problemen met geheimhouding van de items kunnen veroorzaken. Vergelijking van de NC selectieprocedures met de CE selectieprocedures laat zien dat bij alle varianten de eerste minder efficiënt gebruik maakt van de bank: zowel het aantal items dat nooit en het aantal items dat vaak (20% of meer) gebruikt wordt, is groter.

Beschouw verder alleen de gebruiksfrequenties bij CE itemselectie. Er valt op dat door de toepassing van inhoudscontrole het aantal items dat niet gebruikt wordt, licht vermindert (van 132 naar 126), maar ook dat het aantal items dat vaak gebruik wordt, toeneemt (van 21 naar 32). Toepassing van de afnamecontrole heeft het verwachte gunstige effect op het aantal niet gebruikte items (75), terwijl ook het aantal vaak gebruikte items daalt naar 19 en er bovendien geen items meer zijn die in meer dan 40% van de afnames worden gebruikt. Het gecombineerd gebruiken van beide controles heeft voor de frequentieverdeling van het itemgebruik duidelijk het beste effect. In Tabel 8 worden dezelfde gegevens gerapporteerd als in Tabel 7 maar in deze tabel hebben ze betrekking op de rekenprocedure toetsen met $\alpha = .075$ en het vergrote indifferentiegebied.

Tabel 8
Frequentieverdeling itemgebruik bij toetsen ($\alpha = .075$; $\delta = .13$) als rekenprocedure

selectie frequentie %	R	MI	MI+C	MI+E	MI+C+E
$f=0$	0	156	156	104	77
$0 < f \leq 2.5$	0	1	2	27	43
$2.5 < f \leq 5$	1	28	28	32	38
$5 < f \leq 7.5$	143	17	16	18	25
$7.5 < f \leq 10$	62	10	13	13	11
$10 < f \leq 15$	36	12	12	27	23
$15 < f \leq 20$	8	7	6	11	16
$20 < f \leq 25$	0	4	1	6	7
$25 < f \leq 30$	0	2	4	2	2
$30 < f \leq 40$	0	2	4	7	5
$40 < f \leq 50$	0	4	2	3	3
$50 < f \leq 75$	0	5	4	0	0
$75 < f \leq 100$	0	2	2	0	0

Zoals kan worden nagegaan levert de vergelijking tussen de itemselectieprocedures hierbij een soortgelijk resultaat op als bij schatten als rekenprocedure.

Vergelijking van de resultaten van Tabel 7 en 8 maakt duidelijk dat de verdelingen van het gebruik van de items uit de itembank bij toetsen als rekenprocedure gunstiger is dan bij schatten met de NC selectieprocedure. Het aantal items dat vaak (20% of meer) wordt afgenomen is bij de NC procedure ongeveer 1.5 keer zo groot. (bijvoorbeeld 17 tegen 27 items bij de selectieoptie MI+C+E). Vergelijken met schatten met de CE selectieprocedure is de verdeling van het itembankgebruik evenwel ongunstiger.

Tenslotte wordt voor de in Tabel 7 en 8 gerapporteerde selectieprocedures de verdeling van de items over de drie subdomeinen in Tabel 9 gegeven.

Tabel 9
Verdeling itemgebruik over subdomeinen voor verschillende itemselecties

subdomein (gewenst %)	A 16	B 20	C 64
selectie-schatten			
R	21.2	21.7	57.2
CE-MI	21.3	20.0	58.7
NC-MI	20.9	15.6	63.6
CE-MI+C	16.2	20.8	62.9
NC-MI+C	16.2	20.4	63.4
CE-MI+E	22.6	21.0	56.4
NC-MI+E	22.7	18.7	58.7
CE-MI+C+E	16.3	20.9	62.8
NC-MI+C+E	16.5	20.4	63.1
selectie-toetsen			
R	21.2	21.2	57.1
MI	23.5	14.5	62.0
MI+C	16.7	21.5	61.8
MI+E	23.2	18.6	58.1
MI+C+E	16.6	21.0	62.3

Uit Tabel 9 wordt duidelijk dat de inhoudelijk gewenste verhoudingen tussen de aantallen van de subdomeinen A, B en C (16:20:64), alleen dicht benaderd worden als expliciet de inhoudscontrole wordt toegepast bij de itemselectie. Bij alle overige selectiemethodes is in de gemiddeld afgenomen toets onderwerp A oververtegenwoordigd en onderwerp C ondervertegenwoordigd met ongeveer 4%.

4.9 Conclusie

Op grond van de uitgevoerde studies worden ten aanzien van de ontwikkeling van een adaptieve plaatsingstoets rekenen/wiskunde de volgende conclusies getrokken.

- 1 De itembank lijkt van voldoende kwaliteit om adaptief te kunnen toetsen.
- 2 Het absolute maximum van 25 items per afname is reëel.
- 3 De winst in het aantal waarnemingen ten opzichte van de schriftelijk toets bedraagt naar verwachting tussen de 25% en 45% van het aantal items.
- 4 Als rekenprocedure in het toetsalgoritme komt toepassing van de dubbele SPRT het meest in aanmerking.
- 5 Additionele restricties op itemselectieprocedures in de vorm van inhoudscontrole en een milde variant van afnamecontrole, kunnen, zonder dat de kwaliteit van de procedures in negatieve zin beïnvloed wordt, worden toegepast.
- 6 Alvorens tot definitieve uitgave van een CAT voor de plaatsingstoets rekenen/wiskunde wordt besloten, zal experimenteel moeten worden vastgesteld dat de werking van de toetsalgoritmen in reële toetssituaties niet in strijd is met de simulatieresultaten.

Ten aanzien van de gebruikte toetsalgoritmen voor de classificatie van kandidaten in drie categorieën kan in het algemeen geconcludeerd worden, dat de toepassing van toetsen als rekenprocedure een beloftevol alternatief is voor de traditionele rekenprocedure met schatten. Naast de in de simulaties gevonden winst in het gemiddeld benodigde aantal waarnemingen bij gelijkblijvende nauwkeurigheid, heeft deze procedure het voordeel dat het rekenwerk tijdens de afname van de toets zeer gering is. Bij schatten moet een iteratieve maximalisatie procedure worden gevolgd; bij toetsen kan worden volstaan, althans bij het gehanteerde IRT model, met de eenvoudige vergelijking van de geobserveerde score met constanten. De invloed van de te kiezen acceptabele kansen op foute beslissingen in samenhang met de breedte van de indifferentiezones op de kwaliteit van de toetsalgoritmen verdient nader onderzoek. Bij de onderzochte itemselectieprocedures is de constatering, dat bij schatten als rekenprocedure het in het algemeen beter is om items te kiezen die maximaal informatief zijn bij de lopende vaardigheidsschatting dan items die maximaal informatief zijn bij het dichtstbijzijnde grenspunt, interessant. In hoeverre dit mede veroorzaakt is door de specifieke karakteristieken van de gebruikte itembank en de gekozen grenspunten is een nog niet beantwoorde vraag. Ten aanzien van de gehanteerde itemselectiemethode bij toetsen als rekenprocedure wordt opgemerkt dat deze

naar alle waarschijnlijkheid nog verbeterd kan worden. In deze studie is er voor gekozen om bij toetsen als rekenprocedure geen gebruik te maken van schattingen van de vaardigheid van de leerlingen. Voor de selectie van items wordt het grenspunt gezocht dat het dichtst ligt bij de geobserveerde score en persoon. Verwacht wordt dat uit een vervolgstudie, waarin voor de itemselectie wel vaardigheidsschattingen zullen worden gemaakt, zal blijken dat de toetsprocedure, analoog aan de resultaten van Spray en Reckase (1994) voor een classificatieprobleem in twee categorieën, nog efficiënter kan worden gemaakt.

Tenslotte zij opgemerkt dat momenteel verder gewerkt wordt aan de verfijningen van de afnamecontrole in de itemselectieprocedure en aan de uitbreiding van de toepassing van de sequentiële toetsingsprocedure voor meer dan drie categorieën.

5 Machinale Antwoordverwerking bij 'On-Line Testing'

Naast het toetsalgoritme bevat de software voor een CAT ook een algoritme voor de verwerking van de responsen die kandidaten geven. Dit algoritme moet elke respons van een kandidaat kunnen beoordelen als goed of fout (gedeeltelijk goed/fout is in het kader van deze adaptieve toets niet mogelijk). De wijze waarop dit gebeurt varieert met de wijze waarop gerepsondeerd moet worden. Vooralsnog zijn er vier typen responsen. Hieronder worden deze beknopt beschreven.

Het enkelvoudige geselecteerde antwoord

Het meest eenvoudige type antwoord dat verwerkt moet kunnen worden is het enkelvoudige geselecteerde antwoord. De kandidaat respondeert door het aanklikken van een antwoord met de muis. De betreffende codering (letter of cijfer) wordt vergeleken met het goede antwoord en geëvalueerd als juist of onjuist.

Voorbeeldopgave:

100 gram is evenveel als ...

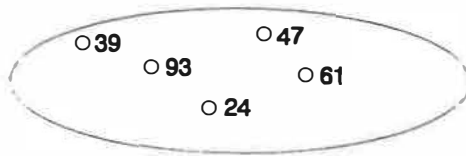
- A 1 kg
 - B 0,1 kg
 - C 0,01 kg
 - D 0,001 kg
-

Het meervoudige geselecteerde antwoord

Ook eenvoudig te verwerken is het meervoudige geselecteerde antwoord. Het verschil met het eerste antwoordtype is dat voor de beantwoording meer dan één antwoordalternatief aangeklikt moet worden. In principe betekent dit dat een antwoord behalve juist of onjuist ook gedeeltelijk juist kan zijn. Voorlopig is er echter voor gekozen om antwoorden alleen als juist of onjuist te evalueren.

Voorbeeldopgave:

Welke getallen zijn groter dan 50?



Het geformuleerde numerieke antwoord

De antwoorden in de vorm van één of meer zelf in te voeren getallen, de numerieke antwoorden, zijn over het algemeen ook gemakkelijk te verwerken.

Voorbeeldopgave:

Tel verder met sprongen van vijf.

50 - 55 - ____ - ____

Te verwachten problemen bij numerieke antwoorden worden veroorzaakt door cultuurverschillen met betrekking tot de notatie van getallen en de creativiteit die kandidaten aan de dag leggen bij het formuleren van een numeriek antwoord (bijvoorbeeld door een getal in woorden te schrijven). De cultuurverschillen spitsen zich toe op het gebruik van punten en komma's in getallen. In Angelsaksische landen is dit vaak precies andersom dan in Nederland: punten worden als decimaal teken gebruikt en komma's als scheidingstekens.

De problemen bij het beoordelen van door kandidaten in te voeren numerieke antwoorden kunnen grotendeels vermeden worden door in de itembank verschillende typen opgaven te onderscheiden die elk een specifieke antwoordverwerking vragen.

Opgaven waarvan de antwoordstring een getal is

Bij opgaven waar als antwoord alleen een **getal** verwacht wordt, moeten alle karakters geblokkeerd (afgevangen) worden behalve de cijfers 0 t/m 9 alsmede de punt, de komma en het minteken (dit laatste voor negatieve getallen). De punt en de komma

worden beide geïnterpreteerd als decimaal teken en dus nooit als scheidingstekens. In de instructie moeten kandidaten leren dat scheidingstekens niet gebruikt mogen worden.

Wordt door een kandidaat een geblokkeerd karakter gebruikt dan wordt hij geattendeerd op deze fout door een akoestisch en visueel (alleen cijfers!) signaal.

Bij gebroken getallen moeten decimale nullen op de laatste plaats genegeerd worden (als het antwoordmodel bijv. 240,03 is, dan wordt een antwoordstring '240,030' goed gerekend).

Opgaven waarvan de antwoordstring een geldbedrag is

Als het antwoord een **geldbedrag** is (d.w.z. wanneer in het invoerveld een valuta-codering staat), wordt ook het liggende streepje (-) vrijgegeven (om aan te geven dat het een bedrag zonder centen betreft). Het liggende streepje kan niet gebruikt worden als aanduiding voor 'schuldbedragen'. Opgaven waarin schuldbedragen uitgerekend moeten worden, worden zódanig geformuleerd dat toch met een positief getal geantwoord kan worden. Bijvoorbeeld: Hoeveel geld komt Dhr.Smits tekort?

Het correcte antwoord waarmee de ingevoerde antwoorden vergeleken worden is een getal met één of meer cijfers vóór de komma en twee cijfers erachter.

Verwerking invoer:

- als decimaal teken worden de punt en de komma geaccepteerd;
- als het goede antwoord een bedrag in hele guldens is, worden de volgende antwoorden goed gerekend:
 - x (bijvoorbeeld: f 10)
 - x,- (bijvoorbeeld: f 10,-)
 - x,00 (bijvoorbeeld: f 10,60);
- de volgende invoer wordt geaccepteerd maar fout gerekend. De reden hiervoor is dat de foutmelding die moet volgen bij het niet accepteren van de invoer als ongewenst effect kan hebben dat de kandidaat op het spoor van het goede antwoord gezet wordt.
 - x, (bijvoorbeeld: f 10,)
 - x,y (bijvoorbeeld: f 10,6)
 - ,yy (bijvoorbeeld: f ,60)
 - x,yyy (bijvoorbeeld: f 10,600).

Al deze notatiewijzen zijn maatschappelijk ongebruikelijk en worden daarom fout gerekend.

- Niet geaccepteerd wordt (er volgt een foutmelding):

invoer van niet vrijgegeven karakters (bijvoorbeeld als een kandidaat een geldbedrag voluit wil schrijven) meer dan 1 decimaal teken in een bedrag (bijvoorbeeld: *f 10,60,*).

Opgaven waarin met geld gerekend moet worden, kunnen ook zó vormgegeven zijn dat er voor het antwoord twee invoervelden zijn. Bijvoorbeeld: _____ gulden en _____ centen. Dit komt voor bij opgaven die tot doel hebben om vast te stellen of kandidaten kunnen tellen. In dergelijke gevallen worden de ingevoerde getallen verwerkt als 'normale getallen' (zie hierboven).

Opgaven waarvan de antwoordstring een tijdstip is

In veel opgaven waarin 'gerekend moet worden met tijd' wordt als antwoord een **tijdstip** of een **tijdsduur** verlangd. Dat is een antwoord met de volgende vorm: uu_mm (Het algoritme voorziet nog niet in de verwerking van opgaven die antwoorden verlangen van de vorm uu_mm_ss of mm_ss). De vorm waarin het antwoord gegeven moet worden, wordt aangegeven door een uur-aanduiding (uur) in het invoerveld. Behalve de cijfers 0 t/m 9 worden de punt, de dubbele punt en de spatie vrijgegeven. De laatste twee karakters kunnen gebruikt worden om uren, minuten en seconden van elkaar te scheiden.

Het correcte antwoord waarmee de ingevoerde antwoorden vergeleken moeten worden, bestaat uit twee cijfers vóór en twee cijfers achter het scheidingsteken.

Verwerking invoer:

- als scheidingsteken worden de punt, de dubbele punt en de spatie geaccepteerd;
- als het goede antwoord in hele uren uitgedrukt kan worden, wordt een antwoord van de vorm 'uu' (bijvoorbeeld: 12 uur) ook goed gerekend;
- antwoorden waarvan $uu > 24$ en $mm > 60$ worden geaccepteerd maar fout gerekend;
- antwoorden waarin meer dan één scheidingsteken staat of waarin niet vrijgegeven karakters gebruikt worden, worden niet geaccepteerd (er komt een foutmelding).

Opgaven waarin met tijd gerekend moet worden, kunnen - net als bij opgaven waarin met geld gerekend wordt - ook zó vormgegeven zijn dat er voor het antwoord twee invoervelden zijn. Bijvoorbeeld: _____ uren en _____ minuten. In dergelijke

gevallen worden de ingevoerde getallen verwerkt als 'normale getallen' (zie hierboven).

Het geformuleerde alfanumerieke antwoord

De grootste problemen doen zich voor bij opgaven waarop met één of meer woorden, al dan niet gecombineerd met getallen, geantwoord moet worden. Wat moet er gedaan worden met spelfouten, verkeerde interpunctie, verkeerd gebruik van hoofdletters, synoniemen? Binnen de voor dit project gestelde termijn was het niet haalbaar om hiervoor een algoritme te schrijven dat op basis van elke mogelijke respons tot een correcte beoordeling komt. In plaats daarvan zijn lijsten aangelegd van goed te keuren antwoorden. Bij het samenstellen van die lijsten is gebruik gemaakt van de antwoorden die cursisten bij de proefafname gegeven hebben.

Een aparte probleemcategorie vormen de antwoorden waarin een breuk staat. Zonder speciale voorzieningen kunnen computerprogramma's breuken niet netjes op het beeldscherm afbeelden.

6 De Ontwikkeling van een Instructiemodule voor Kandidaten²

Computergestuurde adaptieve toetsen hebben als meest opvallende kenmerk dat de vragen op een beeldscherm worden gepresenteerd en dat de kandidaat het antwoord met de muis aanwijst op het beeldscherm of invoert met behulp van het toetsenbord. Het verschil tussen een computergestuurde adaptieve toets en een 'gewone pen-en-papier toets' lijkt op het eerste gezicht triviaal, maar een nadere beschouwing leert dat het maken van een adaptieve toets 'aan de computer' op bepaalde punten essentieel anders is dan het maken van een pen-en-papier toets. Zo is het bijvoorbeeld bij een computergestuurde adaptieve toets niet mogelijk om, voorafgaand aan de beantwoording, eerst een globale indruk te krijgen van de toets door alle items even 'langs te lopen'. De toets ligt namelijk niet klaar maar wordt 'on-line' samengesteld. Eveneens onmogelijk is het om de beantwoording van een gepresenteerd item uit te stellen. Voor kandidaten die toch al last hebben van toetsangst en/of weinig ervaring hebben met het bedienen van computers, kunnen deze verschillen ten opzichte van pen-en-papier toetsen aanleiding geven tot een versterking van de toetsangst.

Mede daarom is het belangrijk kandidaten van tevoren zorgvuldig te instrueren over de specifieke wijze waarop een computergestuurde adaptieve toets gemaakt moet worden. Daarnaast moet voor wat betreft de vaardigheden van kandidaten om een computer(programma) te bedienen, worden uitgegaan van het 'worst case scenario' dat zich onder de potentiële kandidaten mensen bevinden die nog nooit een computer bediend hebben. Daarom moet in heldere en beknopte bewoordingen de werking van het toetsprogramma worden uitgelegd. Essentieel daarbij is de uitleg die gegeven wordt over de bediening met de muis, omdat de computerprogramma's gebruik maken van het besturingssysteem Windows95^R.

De ontwikkelde instructie bestaat uit twee onderdelen:

- ▶ programmabesturing met behulp van een muis;
- ▶ de bediening van het toetsprogramma.

De muisinstructie is afkomstig van het besturingssysteem Windows 3.1^R. Van de instructie voor de bediening van het toetsprogramma zijn twee versies vervaardigd in verband met het geplande onderzoek naar de relatieve effectiviteit van toetsprocedures bij plaatsingsbeslissingen. In dat onderzoek wordt een adaptieve toetsprocedure

² Dit hoofdstuk is gebaseerd op de werkzaamheden die in het kader van een afstudeerstage werden uitgevoerd door Mw. Katja Clemens (Universiteit Twente).

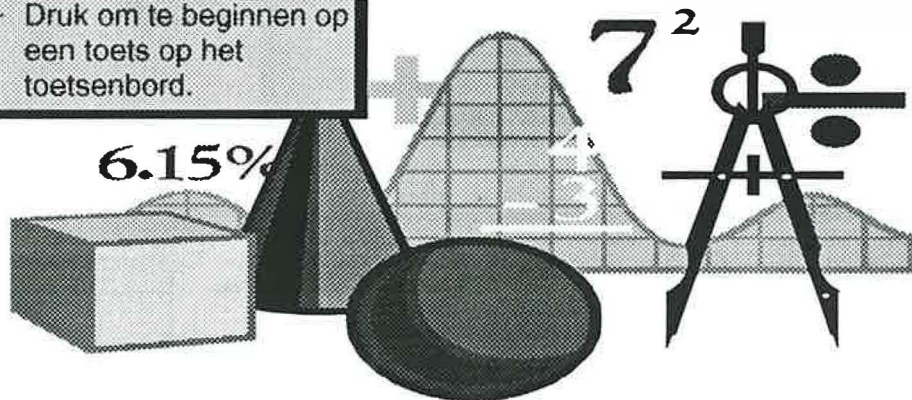
vergeleken met respectievelijk een computergestuurde lineaire toetsprocedure en een conventionele pen-en-papiertoets. Aangezien de bediening van de computergestuurde lineaire toetsprocedure op diverse punten afwijkt van die van een adaptieve toetsprocedure is hiervoor een aparte instructie ontwikkeld.

Hieronder volgen de beeldschermen die kandidaten te zien krijgen bij de instructie over:

- ▶ de computergestuurde adaptieve toets (CAT-versie);
- ▶ de computergestuurde lineaire toets (CIT-versie);
- ▶ het werken met een muis.

Welkom bij de
Intaketoets rekenen / wiskunde
voor de basiseducatie
Cito (1995)

➤ Druk om te beginnen op een toets op het toetsenbord.

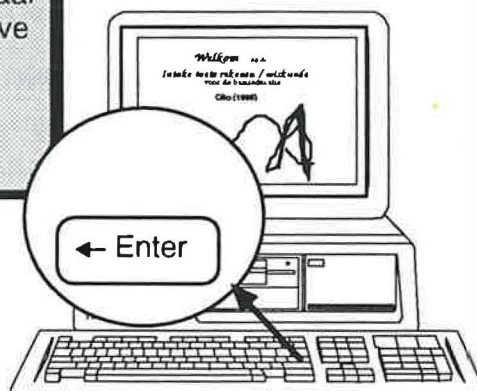


cat 1

Deze **Intaketoets rekenen / wiskunde** is bedoeld om te kijken welk onderwijsniveau voor u geschikt is.

U gaat deze toets maken op de computer. De computer zet een rekenopgave op het scherm. U leest deze opgave goed door en geeft het antwoord. Vervolgens geeft de computer een nieuwe opgave. In totaal krijgt u 25 opgaven. Na de laatste opgave krijgt u de uitslag.

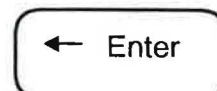
➤ Druk op de **Enter-toets** op het toetsenbord om verder te gaan.



cat 2

Voordat de toets begint, krijgt u een korte uitleg.
Voor het beantwoorden van sommige opgaven moet u de muis gebruiken.

- Als u met de muis kunt omgaan, klik dan met de muis op het hokje met het pijltje.
- Als u niet weet wat een muis is, of er niet (goed) mee kunt omgaan, druk dan op de **Enter-toets** op het toetsenbord.



cat 3

Hieronder ziet u een voorbeeldopgave.
Aan de onderkant van het scherm ziet u ook een aantal knoppen. In de tekstballonnen leest u waarvoor die knoppen zijn. Straks kunt u met deze knoppen oefenen.

- Als u klaar bent met lezen klik dan op .

ve 1 van 25

$200 - 50 = \boxed{}$


Verbeteren


Volgende opgave


Help

Als u zeker bent van uw antwoord, klikt u op deze knop om door te gaan met de volgende opgave.

Als u niet meer weet hoe u een opgave moet beantwoorden, klikt u op deze knop.

Als u een fout antwoord hebt gegeven en u wilt het veranderen, klikt u op deze knop.

cat 4

Dit is een voorbeeldopgave waarbij u de muis moet gebruiken om het antwoord te geven. U kunt het antwoord kiezen door met de muis op het rondje vóór het goede antwoord te klikken. Het rondje wordt zwart. Probeer maar.

➤ Klik daarna met de muis op .

Opgave 2 van 25

$$3 + 2 =$$

☐ 5

☐ 6

☐ 1



Verbeteren



Volgende
opgave



Help

cat 5

Dit is een voorbeeldopgave waarbij een verkeerd antwoord is gekozen.

➤ Verbeter het antwoord. Klik hiervoor met de muis op . Het foute antwoord verdwijnt. U kunt nu het goede antwoord kiezen.

➤ Als u zeker bent van uw antwoord, klikt u op .

Opgave 2 van 25

$$3 + 2 =$$

☐ 5

☒ 6

☐ 1



Verbeteren



Volgende
opgave



Help

cat 6

Dit is een voorbeeldopgave waarbij u meerdere antwoorden moet geven.

U kunt de antwoorden kiezen door één voor één op de rondjes vóór de goede antwoorden te klikken.

Als u een fout antwoord hebt gekozen, klikt u op .

De antwoorden verdwijnen. U kunt opnieuw antwoord geven. Probeer maar.

➤ Als u zeker bent van uw antwoorden, klikt u op .

Opgave 3 van 25

Welke getallen zijn groter dan 55?

- | | | |
|--------------------------|--------------------------|--------------------------|
| ☐ 13 | ☐ 23 | ☐ 55 |
| ☐ 69 | ☐ 35 | ☐ 39 |
| ☐ 51 | ☐ 77 | ☐ 97 |



Verbeteren



Volgende
opgave



Help

cat 7

Dit is een opgave waarbij u het toetsenbord moet gebruiken.

Meestal is het antwoord een getal, maar soms moet er een woord ingevuld worden.

Let op: Er zijn mensen die punten of komma's gebruiken om grote getallen beter leesbaar te maken (bijvoorbeeld 1.000.000). Dat mag hier **niet**.

Gebruik een punt of komma **alleen** in gebroken getallen (bijvoorbeeld: 1% van 150 = 1,5).

➤ Typ het antwoord, gebruik hierbij geen punten of komma's.

➤ Als u zich vergist heeft, klik dan op . Het antwoord verdwijnt. U kunt dan opnieuw een antwoord geven.

➤ Als u zeker bent van uw antwoord, klikt u op .

Opgave 4 van 25

$$999 + 1 = \boxed{}$$



Verbeteren



Volgende
opgave



Help

cat 8

Als u tijdens de toets niet meer weet hoe u een bepaalde opgave moet beantwoorden, klikt u met de muis op .

- Probeer maar.
- Type daarna het antwoord en klik op .

De betreffende helpboodschap wordt hier afgebeeld...

Opgave 5 van 25

Fatima koopt een bos bloemen van f7,95. Ze geeft tien gulden. Hoeveel krijgt ze terug?

f |



Verbeteren



Volgende
opgave



Help

cat 9

Bij de meeste opgaven mag u pen en papier gebruiken om het antwoord te berekenen. Maar bij sommige opgaven mag dit niet.

Er staat dan boven: **Uit het hoofd !** of **Schatten!**

- Bereken de opgave uit het hoofd en typ het antwoord.
- Als u zeker bent van uw antwoord, klikt u op .

Opgave 6 van 25

Uit het hoofd!

15 - 13 = |



Verbeteren



Volgende
opgave



Help

cat 10

Als u een bepaalde opgave niet kunt beantwoorden, mag u deze opgave overslaan door op  te klikken. Een overgeslagen opgave wordt wél fout gerekend.

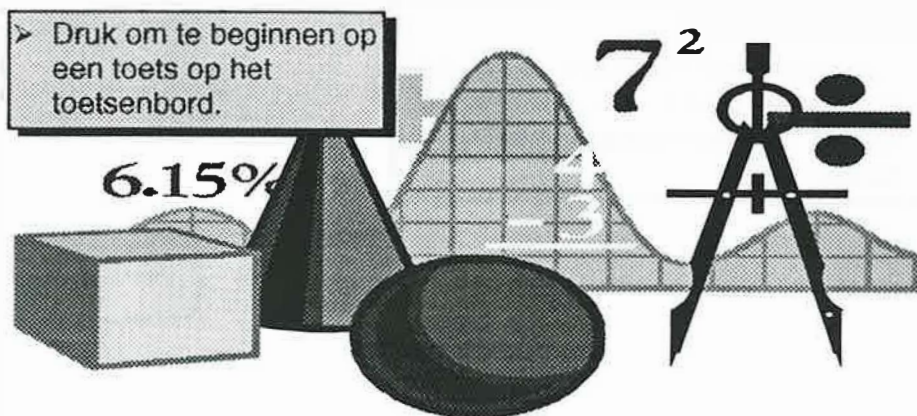
De uitleg is afgelopen. Als u op **starten van toets** klikt, begint de toets. Wilt u voordat u start de uitleg herhalen? Klik dan op **Herhalen**.

Herhalen

Starten van toets

cat 11

Welkom bij de
Intaketoets rekenen / wiskunde
voor de basiseducatie
Cito (1995)

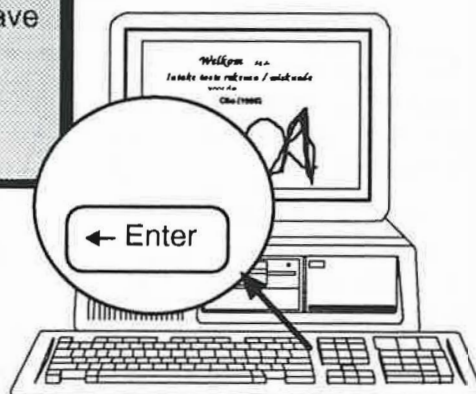


cat 1

Deze **Intaketoets rekenen / wiskunde** is bedoeld om te kijken welk onderwijsniveau voor u geschikt is.

U gaat deze toets maken op de computer. De computer zet een rekenopgave op het scherm. U leest deze opgave goed door en geeft het antwoord. Vervolgens geeft de computer een nieuwe opgave. In totaal krijgt u 25 opgaven. Na de laatste opgave krijgt u de uitslag.

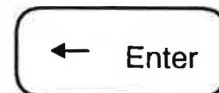
➤ Druk op de **Enter-toets** op het toetsenbord om verder te gaan.



cat 2

Voordat de toets begint, krijgt u een korte uitleg.
Voor het beantwoorden van sommige opgaven moet u de muis gebruiken.

- Als u met de muis kunt omgaan, klik dan met de muis op het hokje met het pijltje.
- Als u niet weet wat een muis is, of er niet (goed) mee kunt omgaan, druk dan op de **Enter-toets** op het toetsenbord.



cat 3

Hieronder ziet u een voorbeeldopgave.
Aan de onderkant van het scherm ziet u ook een aantal knoppen. In de tekstballonnen leest u waarvoor die knoppen zijn. Straks kunt u met deze knoppen oefenen.

- Als u klaar bent met lezen klikt u op

1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu opgaven van deel 1

opgave 2

3 + 2 =

☐ 5

☐ 6

☐ 1

Als u naar een eerdere opgave wilt terugkeren, klikt u één of meerdere keren op deze knop.

Als u een fout antwoord hebt gegeven en u wilt het veranderen, klikt u op deze knop.

Klik op deze knop voor de volgende opgave.

Als u niet meer weet hoe u een opgave moet beantwoorden, klikt u op deze knop.

Verbeteren

Vorige opgave

Volgende opgave

Help

Naar deel 2

cit 1

Dit is een voorbeeldopgave waarbij u de muis moet gebruiken om het antwoord te geven. U kunt het antwoord kiezen door met de muis op het rondje vóór het goede antwoord te klikken. Het rondje wordt zwart. Probeer maar.

➤ Klik daarna met de muis op  .

☐ 1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu opgaven van deel 1

opgave 2

$$3 + 2 =$$

- ☐ 5
☐ 6
☐ 1



Verbeteren



Vorige opgave



Volgende opgave



Help

Naar deel 2

cat 5+

Dit is een voorbeeldopgave waarbij een verkeerd antwoord is gekozen.

➤ Verbeter het antwoord. Klik hiervoor met de muis op  . Het foute antwoord verdwijnt. U kunt nu het goede antwoord kiezen.

➤ Als u uw antwoord heeft ingevuld, klikt u op  .

☐ 1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu opgaven van deel 1

opgave 2

$$3 + 2 =$$

- ☐ 5
☒ 6
☐ 1



Verbeteren



Vorige opgave



Volgende opgave



Help

Naar deel 2

cat 6+

Dit is een voorbeeldopgave waarbij u meerdere antwoorden moet geven.
 U kunt de antwoorden kiezen door één voor één op de rondjes vóór de goede antwoorden te klikken.
 Als u een fout antwoord hebt gekozen, klikt u op **X**.
 De antwoorden verdwijnen. U kunt opnieuw antwoord geven.
 Probeer maar.
 ➤ Klik op **▶** als u klaar bent.

1 **2** 3 4 5 6 7 8 9 10
 11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 3

Welke getallen zijn
groter dan 55?

☐ 13 ☐ 23 ☐ 55
☐ 69 ☐ 35 ☐ 39
☐ 51 ☐ 77 ☐ 97



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

cat 7+

Dit is een opgave waarbij u het toetsenbord moet gebruiken.
 Meestal is het antwoord een getal, maar soms moet er een woord ingevuld worden.
Let op: Er zijn mensen die punten of komma's gebruiken om grote getallen beter leesbaar te maken (bijvoorbeeld 1.000.000). Dat mag hier niet.
 Gebruik een punt of komma **alleen** in gebroken getallen (bijvoorbeeld: 1% van 150 = 1,5).
 ➤ Typ het antwoord, gebruik hierbij geen punten of komma's.
 ➤ Als u zich vergist, klik dan op **X**. Het antwoord verdwijnt. U kunt dan opnieuw een antwoord geven.
 ➤ Als u zeker bent van uw antwoord, klikt u op **▶**.

1 **2** **3** 4 5 6 7 8 9 10
 11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 4

$999 + 1 =$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

Als u tijdens de toets niet meer weet hoe u een bepaalde opgave moet beantwoorden, klikt u met de muis op .

- Probeer maar.
- Typ daarna het antwoord en klik op .

Betreffende helpboodschap wordt hier afgebeeld.

    5 6 7 8 9 10
11 12 13 14 15

U maakt nu opgaven van deel 1

opgave 5

Fatima koopt een bos bloemen van f7,95. Ze geeft tien gulden. Hoeveel krijgt ze terug?

f |



Verbeteren



Vorige opgave



Volgende opgave



Help

Naar deel 2

cat 9+

Bij de meeste opgaven mag u pen en papier gebruiken om het antwoord te berekenen. Maar bij sommige opgaven mag dit niet.

Er staat dan boven: **Uit het hoofd !** of **Schatten!**

- Bereken de opgave uit het hoofd en typ het antwoord.
- Als u zeker bent van uw antwoord, klikt u op .

     6 7 8 9 10
11 12 13 14 15

U maakt nu opgaven van deel 1

opgave 6

Uit het hoofd!

15 - 13 = |



Verbeteren



Vorige opgave



Volgende opgave



Help

Naar deel 2

cat 10+

De toets die u gaat maken bestaat uit twee delen. In deel 1 krijgt u 15 opgaven en in deel 2 krijgt u 10 opgaven. Tijdens de toetsafname kunt u bovenaan het scherm lezen in welk deel u bezig bent. Binnen een deel kunt u heen en terug 'bladeren' door de toets. Bovenaan het scherm wordt ook bijgehouden welke opgaven al door u beantwoord zijn. Als er een hokje om het opgavenummer staat, heeft u de opgave al beantwoord.

➤ Klik op  om verder te gaan.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 6 ☐ 7 8 9 10
11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 8

$$200 - 50 = \boxed{1}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

cit 2

U mag zelf weten in welke volgorde u de opgaven van deel 1 of van deel 2 maakt. U kunt door de toets 'bladeren' met  en . De computer houdt voor u bij welke opgaven u al beantwoord heeft. Bovenaan het scherm kunt u bijvoorbeeld zien dat opgave 6 nog niet beantwoord is.

- Blader terug naar opgave 6. Klik hiervoor twee keer op . Let op: u ziet dat bij opgave 7 al een antwoord is gegeven, maar bij opgave 6 niet.
- Klik op twee keer op  om bij opgave 8 verder te gaan.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 6 ☐ 7 8 9 10
11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 8

$$200 - 50 = \boxed{1}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

cit 3

U mag zelf weten in welke volgorde u de opgaven van deel 1 of van deel 2 maakt. U kunt door de toets 'bladeren' met  en . De computer houdt voor u bij welke opgaven u al beantwoord heeft. Bovenaan het scherm kunt u bijvoorbeeld zien dat opgave 6 nog niet beantwoord is.

- Blader terug naar opgave 6. Klik hiervoor twee keer op . Let op: u ziet dat bij opgave 7 al een antwoord is gegeven, maar bij opgave 6 niet.
- Klik op twee keer op  om bij opgave 8 verder te gaan.

1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 7

$$5 \times 15 = 75$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

hulpcit 1

U mag zelf weten in welke volgorde u de opgaven van deel 1 of van deel 2 maakt. U kunt door de toets 'bladeren' met  en . De computer houdt voor u bij welke opgaven u al beantwoord heeft. Bovenaan het scherm kunt u bijvoorbeeld zien dat opgave 6 nog niet beantwoord is.

- Blader terug naar opgave 6. Klik hiervoor twee keer op . Let op: u ziet dat bij opgave 7 al een antwoord is gegeven, maar bij opgave 6 niet.
- Klik op twee keer op  om bij opgave 8 verder te gaan.

1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 6

Uit het hoofd!

$$15 - 13 = \boxed{1}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

hulpcit 2

Als u het antwoord op een opgave niet meteen weet kunt u deze overslaan. U kunt deze opgave later nog maken.

- Klik op , om de opgave over te slaan.
- Klik nogmaals op  om verder te gaan.

1 **2** **3** **4** **5** 6 **7** 8 9 10
11 12 13 14 15

**U maakt nu
opgaven van deel 1**

opgave 8

$$200 - 50 = \boxed{}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

cit 4

Als u het antwoord op een opgave niet meteen weet kunt u deze overslaan. U kunt deze opgave later nog maken.

- Klik op , om de opgave over te slaan.
- Klik nogmaals op  om verder te gaan.

1 **2** **3** **4** **5** 6 **7** 8 9 10
11 12 13 14 15

**U maakt nu
opgaven van deel 1**

opgave 9

$$5 \% \text{ van } 60 \text{ is } \boxed{}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

hulpcit 3

Als u van deel 1 alle opgaven die u kunt beantwoorden gemaakt heeft, gaat u verder met deel 2. U mag dus opgaven overslaan. Overgeslagen opgaven worden wél fout gerekend!
LET OP: Als u in deel 2 van de toets bezig bent, kunt u niet terug naar de opgaven van deel 1.

- Alle opgaven van deel 1 zijn beantwoord. Ga nu naar deel 2 door op **naar deel 2** te klikken.
- Klik op **▶** om verder te gaan.

1 2 3 4 5 6 7 8 9 10
11 12 13 14 15

U maakt nu
opgaven van deel 1

opgave 8

$$200 - 50 = \boxed{150}$$



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Naar deel 2

cit 5

Als u van deel 1 alle opgaven die u kunt beantwoorden gemaakt heeft, gaat u verder met deel 2. U mag dus opgaven overslaan. Overgeslagen opgaven worden wél fout gerekend!
LET OP: Als u in deel 2 van de toets bezig bent, kunt u niet terug naar de opgaven van deel 1.

- Alle opgaven van deel 1 zijn beantwoord. Ga nu naar deel 2 door op **naar deel 2** te klikken.
- Klik op **▶** om verder te gaan.

1 2 3 4 5 6 7 8 9 10

U maakt nu
opgaven van deel 2

opgave 1

Josje koopt een appel van 50 ct en
een sinaasappel van 1 gulden.
Hoeveel moet Josje betalen?

F I



Verbeteren



Vorige
opgave



Volgende
opgave



Help

Stoppen

hulpcit 4

Dit is de laatste opgave van deel 2. Onderaan het scherm ziet u welke opgaven beantwoord zijn. U kon opgave 4 niet beantwoorden. U heeft deze opgave overgeslagen. U wilt stoppen met de toets

➤ U kunt stoppen door op te klikken.

12345678910

U maakt nu opgaven van deel 2

opgave 10

$10 \times 4 =$

Verbeteren

Vorige opgave

Volgende opgave

Help

cit 6

De computer ziet dat u opgave 4 nog niet gemaakt heeft. Hij vraagt nu of u wilt stoppen of dat u terug wilt naar de toets om de overgeslagen opgave te maken.

➤ Klik op .

12345678910

U maakt nu opgaven van deel 2

opgave 10

$10 \times 4 =$

Verbeteren

Vorige opgave

Volgende opgave

Help

U heeft één of meer opgaven niet beantwoord. Wilt u echt stoppen?

cit 7

De uitleg is afgelopen. Als u op **starten van toets** klikt, begint de toets. Wilt u voordat u start de uitleg herhalen? Klik dan op **Herhalen**.

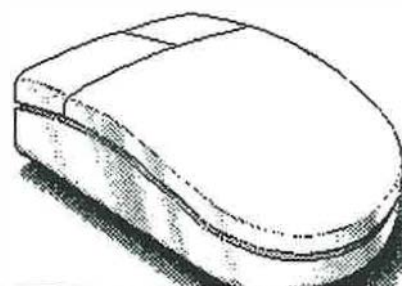
Herhalen

Starten van toets

cit 8

Hieronder ziet u een plaatje van een muis. Aan uw computer zit ook zo'n apparaatje. De muis zit met een snoertje aan de computer vast.

- Druk op de **Enter-toets** op het toetsenbord om verder te gaan.



m1

Op het plaatje ziet u hoe u de muis kunt vasthouden.

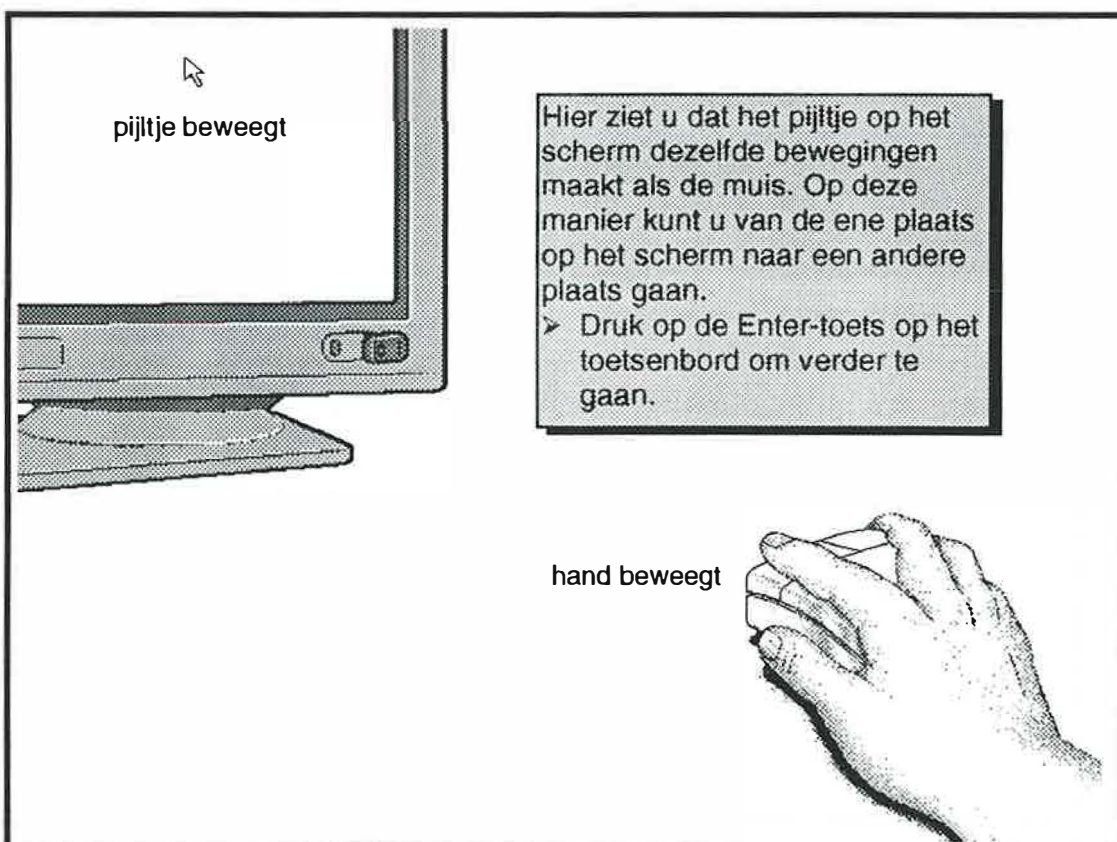
- Probeer maar.
- Druk daarna op de enter-toets op het toetsenbord om verder te gaan.



m2



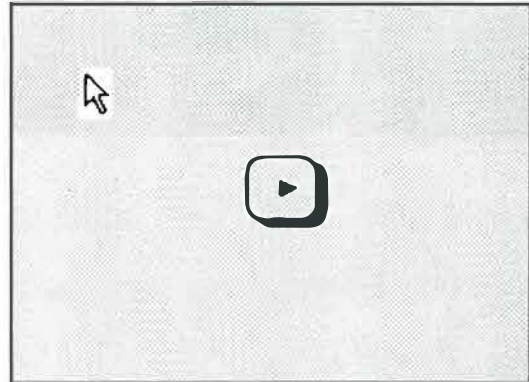
m3



m4

Hieronder ziet u een knop met een pijltje erin.

- Beweeg de muis zo dat u met het pijltje het midden van de knop aanwijst.



m5

Hieronder ziet u een knop met een pijltje erin.

- Beweeg de muis zo dat u met het pijltje het midden van de knop aanwijst.

U weet nu hoe u met de muis het pijltje kunt verplaatsen.

- Druk op de **Enter-toets** op het toetsenbord om door te gaan.



m6

Op de muis zitten twee of drie knoppen. In deze toets gebruikt u alleen de **linker knop**.

- Druk op de linker knop van de muis om door te gaan.



m7

Hieronder ziet u weer de knop met een pijltje erin. Als u met behulp van de muis op de knop 'kijkt', geeft de computer een nieuw scherm.

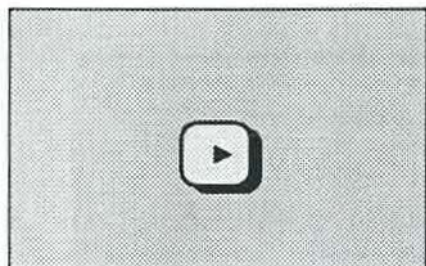
- Beweeg de muis zo dat u met het pijltje het midden van de knop aanwijst.
- Klik op de linker knop van de muis om verder te gaan



m8

U weet nu hoe u met de muis moet omgaan.

➤ Klik op  om door te gaan.



7 De Behoefte aan Adaptief Toetsen in de Volwasseneneducatie³

7.1 Samenvatting

In het kader van het project adaptief toetsen is een onderzoek uitgevoerd om informatie te verkrijgen over mogelijkheden van en eventuele interesse voor adaptief toetsen bij instellingen van de volwasseneneducatie. Aan de hand van een vragenlijst is informatie verzameld over de huidige toetspraktijk binnen instellingen van de VE. Hoe vaak wordt er gebruik gemaakt van toetsen? Met welk doel worden toetsen over het algemeen afgenomen? Wat is de herkomst van het merendeel van de gebruikte toetsen? Bij de vragenlijst was een informatiebrochure over adaptief toetsen gevoegd. Daarin werd uitgelegd wat adaptief toetsen is, hoe het werkt en wat de voordelen ervan zijn. Aan de hand van bovengenoemde vragenlijst is ook gevraagd naar de mening van de respondenten over de kenmerken van adaptief toetsen, zodat geïnventariseerd kon worden of er binnen de VE interesse voor adaptieve toetsen rekenen/wiskunde is als deze op de markt gebracht zouden worden. Indien instellingen gebruik willen maken van adaptieve toetsen, moeten zij beschikken over bepaalde computerfaciliteiten. Daarom is ook informatie verzameld over de aanwezige computerinfrastructuur van de VE-instellingen.

Uit de resultaten (percentages, gemiddelden en standaarddeviaties) van het onderzoek blijkt dat in de praktijk veelvuldig gebruik wordt gemaakt van toetsen. Het merendeel van de instellingen voor de volwasseneneducatie heeft interesse voor adaptieve toetsen. De instellingen beschikken voor het invoeren van adaptieve toetsen over voldoende computerfaciliteiten.

7.2 Inleiding

In instellingen voor de volwasseneneducatie, waar cursisten met zeer verschillende achtergronden naar toe komen om zeer uiteenlopende doelen te bereiken, is het van belang objectieve informatie te verzamelen over hun kennis en vaardigheid. Om het onderwijsproces zo efficiënt mogelijk te laten verlopen, wordt in de volwassenen-

³ Dit hoofdstuk is geschreven door Mw. J. Bakermans in het kader van haar stage vanuit het Instituut voor Onderwijskunde (KU Nijmegen).

educatie op verschillende niveaus les gegeven. Bij binnenkomst is het daarom van belang te weten op welk niveau cursisten les zouden moeten krijgen. De prestatie op een plaatsingstoets geeft objectieve informatie voor de beslissing over het niveau waarop een cursist een onderdeel of vak het beste kan volgen. Doordat er in de volwasseneneducatie grote verschillen kunnen zijn tussen de cursisten met betrekking tot hun vooropleiding, werkervaring en capaciteiten, zit er voor de toetsconstructeur niets anders op dan de opgaven in de toets flink te spreiden qua moeilijkheidsgraad. Dit betekent automatisch dat veel van de opgaven in een toets te moeilijk of te makkelijk zullen zijn voor een kandidaat. Deze opgaven kunnen maar weinig bijdragen aan de meetnauwkeurigheid van een toets. Een ander probleem binnen de volwasseneneducatie is de geheimhouding van de opgaven. Toetsen worden meerdere keren per jaar gebruikt. De praktijk leert dat al na korte tijd opgaven met vraag en antwoord bekend zijn bij cursisten. Uiteraard tast dit de kwaliteit van de beslissing aan die op grond van de toets genomen moet worden. Veel van deze problemen zijn in principe op te lossen door adaptief te toetsen.

Computergestuurd adaptief toetsen kent veel voordelen, maar wordt nog nauwelijks in de onderwijspraktijk toegepast (Moerkerke en Munsters, 1994). Er is een inventarisatie-onderzoek uitgevoerd waarvan de vraagstellingen als volgt luiden:

- 1 Wat is de huidige toetspraktijk binnen de instellingen van de volwasseneneducatie?
- 2 Wat is de mening van docenten rekenen/wiskunde binnen de volwasseneneducatie over adaptief toetsen?
- 3 Over welke computerfaciliteiten beschikken de verschillende instellingen van de volwasseneneducatie?

7.3 Methode

7.3.1 Respondenten

Aan 280 instellingen van de volwasseneneducatie is een vragenlijst gestuurd ter attentie van de contactpersoon rekenen/wiskunde. Deze instellingen zijn onder te verdelen in de volgende vier categorieën: Basiseducatie (BE), Centrum voor Beroepsoriëntatie en Beroepsoefening (CBB), Centrum Vakopleiding (CV) en het Voortgezet

Algemeen Volwassenen Onderwijs (VAVO). Soms hebben meerdere docenten van dezelfde instelling een vragenlijst ingevuld.

7.3.2 Materiaal

Bij het onderzoek is gebruik gemaakt van een informatiebrochure over adaptief toetsen en een vragenlijst.

De informatiebrochure had tot doel docenten binnen de volwasseneneducatie te informeren over de werkwijze en kenmerken van adaptief toetsen.

De vragenlijst bestond uit drie delen. Het eerste deel betrof de huidige toetspraktijk binnen de betreffende instelling. Hier werd onder andere gekeken naar het doel van toetsafname, de frequentie van toetsafname en de herkomst van de gebruikte toetsen. In het tweede deel werd door middel van een vijfpuntsschaal naar de meningen van docenten rekenen/wiskunde over adaptief toetsen gevraagd. Het derde deel van de vragenlijst bestond uit een inventarisatie van de computerinfrastructuur van de betreffende instelling. Voor adaptief toetsen moeten namelijk voldoende DOS-computers aanwezig zijn om cursisten met behulp van de computer te kunnen toetsen.

7.3.3 Resultaten

Hieronder worden de resultaten van het inventarisatie-onderzoek beschreven. Er wordt gerapporteerd per type instelling en over alle instellingen heen. De data worden over het algemeen samengevat met behulp van de volgende maten: percentages, gemiddelden en standaarddeviaties. In Tabel 1 wordt de respons op de vragenlijst weergegeven. Tabel 2 geeft aan hoeveel cursisten gemiddeld genomen de cursus rekenen/wiskunde in de betreffende instelling volgen. In Tabel 3 tot en met Tabel 9.6 worden de resultaten beschreven met betrekking tot de huidige toetspraktijk binnen instellingen van de volwasseneneducatie. In Tabel 10 tot en met Tabel 10.7 en Tabel 11 worden de meningen van de docenten rekenen/wiskunde binnen de volwasseneneducatie over adaptief toetsen weergegeven. In het laatste deel, Tabel 12.1 tot en met Tabel 12.4, komt naar voren welke computerfaciliteiten de verschillende instellingen binnen de volwasseneneducatie tot hun beschikking hebben.

7.3.4 De respons

Tabel 1
Aantal verstuurde en verwerkte vragenlijsten per type instelling

	<i>BE</i>	<i>CBB</i>	<i>CV</i>	<i>VAVO</i>	<i>Totaal</i>
<i>aantal verstuurd</i>	130	28	72	50	280
<i>aantal verwerkt</i>	41	11	14	4	70
<i>% verwerkt</i>	31.54	39.29	19.44	8.00	25.00

Uit Tabel 1 blijkt dat 25% van het totaal aantal verzonden vragenlijsten is verwerkt in de analyse. Een lage maar vrij normale respons gelet op de omstandigheden waaronder het onderzoek plaatsvond. Zo lagen er ten tijde van de enquête alleen al van het Cito nog twee verzoeken tot deelname aan onderzoeken, was het cursusjaar in de opstartfase en waren de docenten en directieleden vooral bezig met de gevolgen van de vele fusies. Vooral de respons uit het VAVO is erg laag.

7.3.5 Aantal cursisten

Tabel 2

Het aantal instellingen per type en het gemiddeld aantal cursisten
dat een cursus rekenen/wiskunde volgt

	<i>BE</i>	<i>CBB</i>	<i>CV</i>	<i>VAVO</i>	<i>Totaal</i>
<i>1-10 cursisten</i>	8	0	0	1	9
<i>11-20 cursisten</i>	7	3	0	1	11
<i>21-30 cursisten</i>	6	1	2	0	9
<i>31-40 cursisten</i>	4	1	1	0	6
<i>41-50 cursisten</i>	5	0	1	0	6
<i>51-60 cursisten</i>	2	2	2	0	6
<i>61-70 cursisten</i>	0	0	0	0	0
<i>71-80 cursisten</i>	1	0	1	0	2
<i>81-90 cursisten</i>	0	0	0	0	0
<i>91-100 cursisten</i>	1	1	2	0	4
<i>> 100 cursisten</i>	4	1	2	2	9

Kijkend naar de frequenties in bovenstaande tabel blijkt dat de grootte van cursusgroepen die rekenen/wiskunde volgt, zeer divers is. Er zijn zowel redelijk wat kleine als enkele grote groepen.

7.3.6 Toetsgebruik

Tabel 3
Percentage respondententen per type instelling en het gebruik van toetsen

	<i>BE</i> (<i>n</i> =40)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =13)	<i>VAVO</i> (<i>n</i> =4)	<i>Totaal</i> (<i>n</i> =68)
<i>wel gebruik</i>	92.50	100.00	92.31	75.00	92.65
<i>geen gebruik</i>	7.50	0.00	7.69	25.00	7.35

Uit Tabel 3 is af te lezen dat in de VE veel gebruik wordt gemaakt van toetsen. Van de ondervraagde respondenten geeft 93 % aan dat ze cursisten toetsen.

Tabel 4
Percentage respondententen per type instelling en het doel van toetsafname.
Bij de beantwoording konden meerdere toetsdoelen worden aangegeven

	<i>BE</i> (<i>n</i> =38)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =13)	<i>VAVO</i> (<i>n</i> =3)	<i>Totaal</i> (<i>n</i> =65)
<i>verwijzing/plaatsing</i>	84.21	90.91	61.54	66.67	80.00
<i>volgen vorderingen</i>	60.53	54.55	61.54	66.67	60.00
<i>diagn. leerprobl.</i>	52.63	45.45	36.36	33.33	46.15
<i>afsluiting/certif.</i>	50.00	27.27	38.46	66.67	44.62
<i>effectmeting</i>	31.58	27.27	7.69	66.67	27.69
<i>anders</i>	2.63	0	7.69	0	3.08

Uit Tabel 4 blijkt dat 80% van alle instellingen toetsen afneemt om cursisten naar een ander onderwijsniveau te verwijzen en/of om te ondersteunen bij de plaatsing van cursisten. Ook het volgen van vorderingen is een vaak genoemd doel van het afnemen van toetsen. Diagnose van leerproblemen en afsluiting/certificering zijn twee toetsfuncties die ook redelijk vaak genoemd worden. Toetsen worden daarentegen redelijk weinig gebruikt om te onderzoeken of cursussen het gewenste effect bereiken.

Tabel 5

Modaal toetsgebruik per type instelling gedurende een cursus rekenen/wiskunde

	<i>BE</i> (<i>n</i> =34)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =12)	<i>VAVO</i> (<i>n</i> =2)	<i>Totaal</i> (<i>n</i> =59)
<i>modus</i>	2	2	2	niet te bepalen	2

Het blijkt dat over het algemeen per cursus rekenen/wiskunde twee keer getoetst wordt.

7.3.7 Herkomst toetsen

Tabel 6

Percentage respondenten per type instelling en de herkomst van de gebruikte toetsen bij de cursus rekenen/wiskunde.

Bij de beantwoording van deze vraag konden meerdere antwoorden worden gegeven

	<i>BE</i> (<i>n</i> =38)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =13)	<i>VAVO</i> (<i>n</i> =3)	<i>Totaal</i> (<i>n</i> =65)
<i>door respondent ontwikkeld</i>	34.21	54.55	15.38	100.00	36.92
<i>door docententeam ontwikkeld</i>	7.89	18.18	69.23	0	21.54
<i>met andere instellingen ontwikkeld</i>	15.79	18.18	23.08	0	16.92
<i>in de handel verkrijgbare toetsen</i>	65.79	45.45	7.69	0	47.69

Uit Tabel 6 blijkt dat over het algemeen 48% van alle instellingen gebruik maakt van in de handel verkrijgbare toetsen. 37% van de ondervraagden ontwikkelt de toetsen geheel zelfstandig en 39% ontwikkelt toetsen in samenwerking met collega's uit de eigen instelling of uit andere instellingen. Kijkend per type instelling zien we dat in

de BE vaker gebruik wordt gemaakt van in de handel verkrijgbare toetsen terwijl in het CBB en de CV juist vaker toetsen zelf ontwikkeld worden.

Tabel 7
Frequentie van gebruik van in de handel verkrijgbare toetsen

<i>Titel</i>	<i>Uitgever</i>	<i>jaartal</i>	<i>aantal keer</i>
Compensatieprogramma rekenen	Dijkstra		3
Plaatsingstoets beroepsopleidingen	Stichting AROC	1993	3
Rekenschap	Stichting IVIO	1981	2
Rekentraject	Stichting IVIO	1991	9
ROC plaatsingstoets BE	Stichting AROC	1994	11
ROC niveauvorderingentoets BE	Stichting AROC	1994	9
Supermarktstrategie	Stichting IDEE	1992	6

Tabel 8
Percentage respondenten per type instelling en de redenen om in de handel verkrijgbare toetsen niet in te zetten.

Bij de beantwoording konden meerdere redenen worden aangegeven

	<i>BE</i> <i>(n=12)</i>	<i>CBB</i> <i>(n=6)</i>	<i>CV</i> <i>(n=9)</i>	<i>VAVO</i> <i>(n=2)</i>	<i>Totaal</i> <i>(n=29)</i>
<i>sluit niet aan</i>	58.33	16.67	66.67	100.00	55.17
<i>onbekend</i>	16.67	16.67	22.22	0	17.24
<i>te duur</i>	0	16.67	0	0	3.45
<i>anders</i>	25.00	50.00	11.11	0	24.14

Uit Tabel 8 komt naar voren dat het niet aansluiten van de toetsen op de leerstof de meest genoemde reden is voor het niet gebruik maken van in de handel verkrijgbare toetsen.

Enkele andere redenen die genoemd werden om in de handel verkrijgbare toetsen niet in te zetten zijn:

- het taalniveau van de toetsen is vaak te hoog;
- methodes die gebruikt worden, bevatten ook toetsmateriaal;
- er waren nooit zoveel goede toetsen;
- het is traditie om zelf toetsen te ontwikkelen;
- nog geen gelegenheid gehad om zich goed op dit gebied te informeren.

7.3.8 Tijdsbesteding ten behoeve van toetsontwikkeling en toetsgebruik

Tabel 9.1

Percentage respondenten per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan het ontwikkelen van toetsmateriaal

	<i>BE</i> (<i>n</i> =15)	<i>CBB</i> (<i>n</i> =4)	<i>CV</i> (<i>n</i> =6)	<i>VAVO</i> (<i>n</i> =3)	<i>Totaal</i> (<i>n</i> =28)
<i>1 uur</i>	13.33	0	16.67	0	10.71
<i>2 uur</i>	6.67	0	33.33	0	10.71
<i>3 uur</i>	6.67	50.00	0	0	10.71
<i>4 uur</i>	6.67	50.00	16.67	0	14.29
<i>6 uur</i>	6.67	0	0	0	3.57
<i>8 uur</i>	0	0	16.67	33.33	7.14
<i>10 uur</i>	26.67	0	0	0	14.29
<i>20 uur</i>	13.33	0	0	33.33	10.71
<i>40 uur</i>	6.67	0	16.67	33.33	10.71
<i>50 uur</i>	6.67	0	0	0	3.57
<i>160 uur</i>	6.67	0	0	0	3.57

Tabel 9.1 toont een grote diversiteit aan tijdsbesteding met betrekking tot het ontwikkelen van toetsmateriaal. Ongeveer 75 % van de instellingen besteedt echter tussen de 1 en de 10 uur aan het ontwikkelen van toetsmateriaal.

Tabel 9.2

Percentage respondenten per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan het organiseren en toezicht houden bij een toetsafname

	<i>BE</i> (<i>n</i> =29)	<i>CBB</i> (<i>n</i> =7)	<i>CV</i> (<i>n</i> =7)	<i>VAVO</i> (<i>n</i> =2)	<i>Totaal</i> (<i>n</i> =45)
<i>1 uur</i>	17.24	28.57	14.29	0	17.78
<i>2 uur</i>	6.90	28.57	28.57	0	13.33
<i>3 uur</i>	6.90	0	14.29	0	6.67
<i>4 uur</i>	10.34	0	0	0	6.67
<i>5 uur</i>	10.34	0	0	0	6.67
<i>6 uur</i>	10.34	0	0	0	6.67
<i>7 uur</i>	0	14.29	0	0	2.22
<i>8 uur</i>	6.90	14.29	14.29	100.00	13.33
<i>9 uur</i>	6.90	0	0	0	4.44
<i>10 uur</i>	6.90	0	0	0	4.44
<i>15 uur</i>	13.79	14.29	0	0	11.11
<i>40 uur</i>	3.45	0	0	0	2.22
<i>60 uur</i>	0	0	14.29	0	2.22
<i>160 uur</i>	0	0	14.29	0	2.22

Uit Tabel 9.2 blijkt dat de schattingen van de respondenten over de tijd die besteed wordt aan het organiseren en toezicht houden sterk uiteen liggen. Van de ondervraagden zegt 75% maximaal een werkdag te investeren ten behoeve van het organiseren en het toezicht houden bij een toetsafname.

Tabel 9.3

Percentage respondentent per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan het nakijken van gemaakte toetsen

	<i>BE</i> (<i>n</i> =31)	<i>CBB</i> (<i>n</i> =8)	<i>CV</i> (<i>n</i> =8)	<i>VAVO</i> (<i>n</i> =3)	<i>Totaal</i> (<i>n</i> =50)
<i>1 uur</i>	19.35	37.50	50.00	0	26.00
<i>2 uur</i>	12.90	12.50	0	0	10.00
<i>3 uur</i>	6.45	0	0	0	4.00
<i>4 uur</i>	9.68	12.50	0	0	8.00
<i>5 uur</i>	16.13	0	0	0	10.00
<i>7 uur</i>	3.23	0	0	0	2.00
<i>8 uur</i>	0	12.50	12.50	0	4.00
<i>10 uur</i>	16.13	12.50	0	33.33	14.00
<i>15 uur</i>	3.23	0	0	0	2.00
<i>16 uur</i>	0	0	0	33.33	2.00
<i>20 uur</i>	3.23	0	0	0	2.00
<i>30 uur</i>	0	12.50	0	0	2.00
<i>40 uur</i>	3.23	0	0	33.33	4.00
<i>45 uur</i>	3.23	0	0	0	2.00
<i>50 uur</i>	0	0	12.50	0	2.00
<i>60 uur</i>	0	0	12.50	0	2.00
<i>70 uur</i>	0	0	12.50	0	2.00
<i>250 uur</i>	3.23	0	0	0	2.00

Ook met betrekking tot de tijd die besteed wordt aan het nakijken van gemaakte toetsen liggen de schattingen sterk uiteen. Van de instellingen besteedt ongeveer 75% minimaal 1 uur en maximaal 10 uur aan het nakijken van gemaakte toetsen.

Tabel 9.4

Percentage respondentent per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan het rapporteren van toetsprestaties naar cursisten

	<i>BE</i> (<i>n</i> =26)	<i>CBB</i> (<i>n</i> =4)	<i>CV</i> (<i>n</i> =7)	<i>VAVO</i> (<i>n</i> =2)	<i>Totaal</i> (<i>n</i> =39)
<i>1 uur</i>	34.62	75.00	57.14	0	41.03
<i>2 uur</i>	23.08	0	0	0	15.38
<i>3 uur</i>	11.54	0	0	0	7.69
<i>4 uur</i>	7.69	0	0	50.00	7.69
<i>5 uur</i>	3.85	25.00	0	0	5.13
<i>8 uur</i>	0	0	28.57	0	5.13
<i>10 uur</i>	7.69	0	0	50.00	7.69
<i>15 uur</i>	7.69	0	0	0	5.13
<i>30 uur</i>	0	0	14.29	0	2.56
<i>250 uur</i>	3.85	0	0	0	2.56

Het merendeel van de respondenten besteedt maximaal twee uur aan de rapportage van de toetsprestaties.)

Tabel 9.5

Percentage respondententen per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan onderzoek naar de kwaliteit van ontwikkelde toetsen

	<i>BE</i> (<i>n</i> =12)	<i>CBB</i> (<i>n</i> =2)	<i>CV</i> (<i>n</i> =2)	<i>VAVO</i> (<i>n</i> =1)	<i>Totaal</i> <i>n</i> =17)
<i>1 uur</i>	16.67	50.00	100.00	0	29.41
<i>2 uur</i>	25.00	50.00	0	100.00	29.41
<i>3 uur</i>	16.67	0	0	0	11.76
<i>4 uur</i>	25.00	0	0	0	17.65
<i>5 uur</i>	8.33	0	0	0	5.88
<i>10 uur</i>	8.33	0	0	0	5.88

Uit Tabel 9.5 is af te lezen dat 17 instellingen tijd besteden aan onderzoek naar de kwaliteit van ontwikkelde toetsen. Van de ondervraagden die de kwaliteit van de toetsen onderzoeken is ongeveer 60% daar maximaal 2 uur mee bezig.

Tabel 9.6

Percentage respondententen per type instelling en geschatte tijdsbesteding gedurende een cursusjaar aan andere activiteiten dan bovengenoemd

	<i>BE</i> (<i>n</i> =8)	<i>CBB</i> (<i>n</i> =2)	<i>CV</i> (<i>n</i> =0)	<i>VAVO</i> (<i>n</i> =0)	<i>Totaal</i> (<i>n</i> =10)
<i>1 uur</i>	0	50.00	0	0	10.00
<i>2 uur</i>	12.50	50.00	0	0	20.00
<i>3 uur</i>	12.50	0	0	0	10.00
<i>4 uur</i>	12.50	0	0	0	10.00
<i>5 uur</i>	25.00	0	0	0	20.00
<i>6 uur</i>	12.50	0	0	0	10.00
<i>7 uur</i>	12.50	0	0	0	10.00
<i>10 uur</i>	12.50	0	0	0	10.00

Uit Tabel 9.6 komt naar voren dat 10 ondervraagden tijd besteden aan andere activiteiten dan eerder genoemde. Enkele andere activiteiten die uit worden gevoerd zijn:

- het actief leiden van de toetsafname bij de supermarktstrategie;
- het kopiëren van toetsen;
- het voeren van plaatsings- en voortgangsgesprekken;
- het rapporteren naar de opdrachtgever;
- het bijhouden van de administratie en het overleggen met collega's en cursisten.

7.3.9 *Opinie over het belang van de kenmerken van adaptief toetsen*

Tabel 10

Gemiddelden en standaarddeviaties per type instelling en het belang dat gehecht wordt aan kenmerken van adaptief toetsen in de eigen toetssituatie

	<i>BE</i>		<i>CBB</i>		<i>CV</i>		<i>VAVO</i>		<i>Totaal</i>	
	<i>gem.</i>	<i>sd.</i>	<i>gem.</i>	<i>sd.</i>	<i>gem.</i>	<i>sd.</i>	<i>gem.</i>	<i>sd.</i>	<i>gem.</i>	<i>sd.</i>
<i>pc stelt toets samen</i>	3.66 n=38	1.05	3.36 n=11	1.21	3.33 n=12	1.23	3.75 n=4	1.50	3.55 n=65	1.12
<i>toets op niveau kandidaat</i>	4.78 n=40	0.53	4.18 n=11	0.87	4.58 n=12	0.67	4.00 n=4	0.82	4.60 n=67	0.68
<i>per kandidaat andere toets</i>	3.67 n=40	1.23	3.20 n=10	0.79	4.00 n=12	1.21	3.00 n=4	0.00	3.62 n=66	1.15
<i>pc verzorgt de scoring</i>	3.90 n=39	0.97	3.82 n=11	1.08	3.70 n=10	1.25	4.50 n=4	0.58	3.89 n=64	1.01
<i>uitslag direct bekend</i>	4.08 n=40	1.07	4.36 n=11	0.67	4.25 n=12	0.87	4.00 n=4	0.82	4.15 n=67	0.96
<i>relatief weinig opgaven</i>	3.72 n=39	1.07	3.45 n=11	0.82	3.75 n=12	1.22	3.00 n=4	0.82	3.64 n=66	1.05
<i>afname op diverse tijden</i>	4.12 n=40	0.97	4.18 n=11	1.08	4.33 n=12	0.78	3.25 n=4	0.96	4.12 n=67	0.96

Gescoord is er op een vijfpuntsschaal waarbij 1 = zeer onbelangrijk en 5 = zeer belangrijk. Uit Tabel 10 blijkt dat de respondenten over het algemeen de kenmerkende eigenschappen van adaptief toetsen van belang achten voor de eigen toetssituatie. Het gemiddelde oordeel is voor elk kenmerk groter dan 3 (= neutraal).

Tabel 10.1
Percentage respondenten per type instelling en het belang dat gehecht
wordt aan het feit dat de computer de toets samenstelt

	<i>BE</i> (<i>n</i> =38)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =12)	<i>VAVO</i> (<i>n</i> =4)	<i>Totaal</i> (<i>n</i> =65)
<i>zeer onbelangrijk</i>	5.26	9.09	8.33	0	6.15
<i>onbelangrijk</i>	2.63	9.09	16.67	25.00	7.69
<i>neutraal</i>	36.84	36.36	25.00	25.00	33.85
<i>belangrijk</i>	31.58	27.27	33.33	0	29.23
<i>zeer belangrijk</i>	23.68	18.18	16.67	50.00	23.08

Uit Tabel 10.1 blijkt dat een kleine meerderheid het voor de eigen toetssituatie van (groot) belang vindt dat de computer de toetsen samenstelt. Een relatief grote groep (34 %) is echter onzeker over de meerwaarde van de computer bij het samenstellen van toetsen.

Tabel 10.2
Percentage respondententen per type instelling en het belang dat gehecht
wordt aan het feit dat het niveau van de toets wordt
afgestemd op het niveau van de kandidaat

	<i>BE</i> (<i>n=40</i>)	<i>CBB</i> (<i>n=11</i>)	<i>CV</i> (<i>n=12</i>)	<i>VAVO</i> (<i>n=4</i>)	<i>Totaal</i> (<i>n=67</i>)
<i>zeer onbelangrijk</i>	0	0	0	0	0
<i>onbelangrijk</i>	0	0	0	0	0
<i>neutraal</i>	5.00	27.27	8.33	25.00	10.45
<i>belangrijk</i>	12.50	27.27	25.00	50.00	19.40
<i>zeer belangrijk</i>	82.50	45.45	66.67	25.00	70.15

In Tabel 10.2 is af te lezen dat 90% van alle respondenten van mening is dat het (zeer) belangrijk is voor de eigen toetssituatie dat het niveau van de toets wordt afgestemd op het niveau van de kandidaat. Geen enkele respondent is van mening dat het afstemmen van de toets op het niveau van de kandidaat onbelangrijk is.

Tabel 10.3
Percentage respondententen per type instelling en het belang dat gehecht wordt
aan het feit dat elke kandidaat in principe een andere toets maakt

	<i>BE</i> (<i>n=40</i>)	<i>CBB</i> (<i>n=10</i>)	<i>CV</i> (<i>n=12</i>)	<i>VAVO (n=4)</i>	<i>Totaal</i> (<i>n=66</i>)
<i>zeer onbelangrijk</i>	5.00	0	8.33	0	4.55
<i>onbelangrijk</i>	15.00	10.00	0	0	10.61
<i>neutraal</i>	20.00	70.00	16.67	100.00	31.82
<i>belangrijk</i>	27.50	10.00	33.33	0	24.24
<i>zeer belangrijk</i>	32.50	10.00	41.67	0	28.79

In Tabel 10.3 komt naar voren dat een kleine meerderheid (53%) het voor de eigen toetssituatie van (groot) belang vindt dat elke kandidaat een andere toets maakt. Opvallend is dat respondenten uit CBB's en VAVO's daar anders over denken. Respectievelijk 70 en 100% van de respondenten kon niet aangeven of dit kenmerk van adaptief toetsen al dan niet van belang was voor de eigen toetssituatie.

Tabel 10.4

Percentage respondenten per type instelling en het belang dat gehecht wordt aan het feit dat de computer zorg draagt voor de scoring

	<i>BE</i> (<i>n</i> =39)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =10)	<i>VAVO</i> (<i>n</i> =4)	<i>Totaal</i> (<i>n</i> =64)
<i>zeer onbelangrijk</i>	2.56	0	0	0	1.56
<i>onbelangrijk</i>	5.13	18.18	30.00	0	10.94
<i>neutraal</i>	20.51	9.09	0	0	14.06
<i>belangrijk</i>	43.59	45.45	40.00	50.00	43.75
<i>zeer belangrijk</i>	28.21	27.27	30.00	50.00	29.69

Uit Tabel 10.4 blijkt dat het merendeel van de respondenten (73%) het (zeer) belangrijk vindt voor de eigen toetssituatie dat de computer zorg draagt voor de scoring.

Tabel 10.5

Percentage respondententen per type instelling en het belang dat gehecht wordt aan het feit dat de uitslag van de toets direct na afloop van de toets bekend is

	<i>BE</i> (<i>n=40</i>)	<i>CBB</i> (<i>n=11</i>)	<i>CV</i> (<i>n=12</i>)	<i>VAVO</i> (<i>n=4</i>)	<i>Totaal</i> (<i>n=67</i>)
<i>zeer onbelangrijk</i>	2.50	0	0	0	1.49
<i>onbelangrijk</i>	7.50	0	8.33	0	5.97
<i>neutraal</i>	15.00	9.09	0	25.00	11.94
<i>belangrijk</i>	30.00	45.45	50.00	50.00	37.31
<i>zeer belangrijk</i>	45.00	45.45	41.67	25.00	43.28

In Tabel 10.5 komt naar voren dat het merendeel van alle respondenten (80%) van mening is dat het (zeer) belangrijk is voor de eigen toetssituatie dat de uitslag direct na afloop van de toets bekend is.

Tabel 10.6

Percentage respondententen per type instelling en het belang dat gehecht wordt aan het feit dat een toets uit relatief weinig opgaven bestaat

	<i>BE</i> (<i>n=39</i>)	<i>CBB</i> (<i>n=11</i>)	<i>CV</i> (<i>n=12</i>)	<i>VAVO</i> (<i>n=4</i>)	<i>Totaal</i> (<i>n=66</i>)
<i>zeer onbelangrijk</i>	2.56	0	8.33	0	3.03
<i>onbelangrijk</i>	10.26	9.09	8.33	25.00	10.61
<i>neutraal</i>	28.21	45.45	8.33	50.00	28.79
<i>belangrijk</i>	30.77	36.36	50.00	25.00	34.85
<i>zeer belangrijk</i>	28.21	9.09	25.00	0	22.73

Een kleine meerderheid van de respondenten (58%) vindt het (zeer) belangrijk voor de eigen toetssituatie dat toetsen uit weinig opgaven bestaan. Bijna 30% van de respondenten heeft geen duidelijke mening over het belang van dit kenmerk van adaptief toetsen voor de eigen toetssituatie.

Tabel 10.7

Percentage respondententen per type instelling en het belang dat gehecht wordt aan het feit dat een toets kan worden afgenomen op een voor de individuele kandidaat geschikt tijdstip

	<i>BE</i> (<i>n</i> =40)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =12)	<i>VAVO</i> (<i>n</i> =4)	<i>Totaal</i> (<i>n</i> =67)
<i>zeer onbelangrijk</i>	0	0	0	0	0
<i>onbelangrijk</i>	10.00	9.09	0	25.00	8.96
<i>neutraal</i>	10.00	18.18	16.67	25.00	13.43
<i>belangrijk</i>	37.50	18.18	33.33	50.00	34.33
<i>zeer belangrijk</i>	42.50	54.55	50.00	0	43.28

Uit Tabel 10.7 blijkt dat een ruime meerderheid van de respondenten (78%) van mening is dat het (zeer) belangrijk is voor de eigen toetssituatie dat kandidaten getoetst kunnen worden op een voor hen geschikt tijdstip.

Tabel 11

Percentage respondententen per type instelling en de interesse voor adaptieve toetsen voor rekenen/wiskunde indien deze beschikbaar zouden zijn

	<i>BE</i> (<i>n</i> =39)	<i>CBB</i> (<i>n</i> =11)	<i>CV</i> (<i>n</i> =13)	<i>VAVO</i> (<i>n</i> =4)	<i>Totaal</i> (<i>n</i> =67)
<i>ja</i>	69.23	63.64	61.54	75.00	67.16
<i>ja, mits</i>	28.21	27.27	30.77	25.00	28.36
<i>nee</i>	2.56	9.09	7.69	0	4.48

Uit Tabel 11 blijkt dat het merendeel van alle instellingen (67%) interesse heeft voor adaptieve toetsen indien deze beschikbaar zouden zijn. Een aantal instellingen (28%) is geïnteresseerd als de adaptieve toetsen voldoen aan een of meer voorwaarden.

Enkele van die voorwaarden zijn:

- de toetsen moeten gericht zijn op de doelgroep;
- de gehanteerde aftestgrenzen moeten direct gerelateerd zijn aan de SVE-niveaus;
- de kosten mogen niet te hoog zijn;
- de cursisten mogen geen taalproblemen ondervinden met betrekking tot het programma;
- de bruikbaarheid van deze toetsen moet in de praktijk bewezen zijn;
- het programma moet voor alle cursisten (ook de cursisten met computerangst) eenvoudig te gebruiken zijn.

7.3.10 Computerinfrastructuur

Tabel 12.1

Percentage instellingen, per type, dat beschikt over één of meerdere typen PC en het gemiddelde aantal PC's dat aanwezig is

	<i>BE (n=37)</i>			<i>CBB (n=10)</i>			<i>CV (n=13)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	18.92	4.71 n=11	2.56	30.00	5.67 n=4	4.51	0.00	n=0	
286	43.24	6.87 n=20	4.69	50.00	8.00 n=6	5.96	23.08	7.67 n=4	7.37
386	40.54	8.47 n=19	5.21	40.00	11.75 n=5	9.67	61.54	28.87 n=9	20.75
486	67.57	12.24 n=29	10.81	60.00	5.50 n=7	4.64	61.54	31.12 n=9	29.69
<i>Pentium</i>	8.11	4.67 n=7	4.16	0.00	n=0		7.69	4.00 n=2	

Vervolg Tabel 12.1

	<i>VAVO (n=4)</i>			<i>Totaal (n=64)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	25.00	20.00 n=1		17.19	6.36 n=17	5.35
286	50.00	30.00 n=2	14.14	40.63	8.96 n=32	8.35
386	50.00	27.50 n=2	17.68	45.31	15.86 n=35	15.24
486	75.00	43.33 n=3	5.77	65.63	17.10 n=48	18.60
<i>Pentium</i>	0.00	n=0		6.25	4.50 n=10	3.42

Uit Tabel 12.1 blijkt dat 65% van alle instellingen (n=48) een 486 heeft. Gemiddeld staan er bij 48 instellingen in totaal 17 486-computers. Kijken we naar de afzonderlijke instellingen dan blijkt dat het merendeel van de instellingen over een 486 beschikt. Er zijn verschillende instellingen die beschikken over een 286 en/of een 386.

Tabel 12.2

Percentage instellingen, per type en per type PC, dat PC's in een netwerk heeft en het gemiddelde aantal PC's dat is aangesloten

	<i>BE (n=37)</i>			<i>CBB (n=10)</i>			<i>CV (n=13)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	2.70	6.00 n=5		20.00	5.50 n=3	6.36	0.00	n=0	
286	16.22	4.83 n=10	5.34	0.00	n=0		7.69	5.00 n=2	
386	18.92	8.00 n=11	5.45	30.00	9.67 n=4	7.37	46.15	35.83 n=7	19.25
486	27.03	10.40 n=14	6.80	30.00	8.00 n=4	5.29	53.85	29.43 n=8	16.41
<i>Pentium</i>	5.41	7.00 n=6	1.41	0.00	n=0		7.69	4.00 n=2	

Vervolg Tabel 12.2

	<i>VAVO (n=4)</i>			<i>Totaal (n=64)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	0.00	n=0		4.69	5.67 n=9	4.51
286	25.00	40.00 n=1		12.50	9.25 n=14	13.22
386	50.00	17.50 n=2	3.54	28.13	18.61 n=24	17.09
486	75.00	36.67 n=3	15.28	35.94	19.30 n=29	15.60
<i>Pentium</i>	0.00	n=0		4.69	6.00 n=9	2.00

Uit Tabel 12.2 blijkt dat voornamelijk de 386 en de 486 in een netwerk staan. Kijken we bijvoorbeeld naar de 486 bij de BE dan blijkt dat bij 27% van de respondenten uit de BE de 486-computers in een netwerk te staan. Gemiddeld staan er bij 14 instellingen voor de BE (n=14) 10 486-computers in een netwerk. De interpretatie van Tabel 12.3 en 12.4 kan op bovengenoemde wijze geschieden.

Tabel 12.3

Percentage instellingen, per type en per type PC, dat PC's heeft met een kleurenscherm en het gemiddelde aantal PC's dat hiermee is uitgerust

	<i>BE (n=37)</i>			<i>CBB (n=10)</i>			<i>CV (n=13)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	5.41	3.50 n=6	4.95	0.00	n=0		0.00	n=0	
286	35.14	5.38 n=17	4.27	0.00	n=0		23.08	7.67 n=4	7.37
386	40.54	7.80 n=19	5.93	40.00	2.75 n=5	1.26	46.15	28.67 n=7	21.34
486	62.16	12.83 n=27	11.06	60.00	5.50 n=7	4.64	61.54	28.50 n=9	23.80
<i>Pentium</i>	8.11	4.67 n=7	4.16	0.00	n=0		7.69	4.00 n=2	

Vervolg Tabel 12.3

	<i>VAVO (n=4)</i>			<i>Totaal (n=64)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
8086	0.00	n=0		3.13	3.50 n=8	4.95
286	25.00	20.00 n=1		26.56	6.65 n=23	5.75
386	25.00	20.00 n=1		40.63	12.31 n=32	14.30
486	75.00	43.33 n=3	5.77	62.50	17.15 n=46	16.92
<i>Pentium</i>	0.00	n=0		6.25	4.50 n=10	3.42

Uit Tabel 12.3 komt naar voren dat over het algemeen de instellingen voor zowel de 286, 386 als de 486 een kleurenscherm hebben. Alleen bij de 8086 beschikken de instellingen over het algemeen niet over een kleurenscherm.

Tabel 12.4

Percentage instellingen, per type en per type PC, dat Windows 3.1x geïnstalleerd heeft en het gemiddelde aantal PC's waarop dit besturingssysteem aanwezig is

	<i>BE (n=37)</i>			<i>CBB (n=10)</i>			<i>CV (n=13)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
<i>8086</i>	5.41	3.50 n=6	4.95	0.00	n=0		0.00	n=0	
<i>286</i>	21.6 7	6.87 n=12	6.10	0.00	n=0		7.69	0.00 n=2	
<i>386</i>	35.1 4	8.77 n=17	4.75	30.00	9.00 n=4	11.36	46.15	12.00 n=7	12.57
<i>486</i>	64.8 6	12.67 n=28	10.82	50.00	6.20 n=6	4.82	61.54	24.37 n=9	17.89
<i>Pentium</i>	8.11	4.67 n=7	4.16	0.00	n=0		7.69	4.00 n=2	

Vervolg Tabel 12.4

	<i>VAVO (n=4)</i>			<i>Totaal (n=64)</i>		
	%	<i>gem.</i>	<i>sd.</i>	%	<i>gem.</i>	<i>sd.</i>
<i>8086</i>	0.00	n=0		3.13	3.50 n=8	4.95
<i>286</i>	25.00	20.00 n=1		15.63	7.50 n=16	7.28
<i>386</i>	25.00	40.00 n=1		35.94	11.00 n=29	10.10
<i>486</i>	50.00	40.00 n=2	0.00	60.94	15.64 n=45	14.03
<i>Pentium</i>	0.00	n=0		6.25	4.50 n=10	3.42

Uit Tabel 12.4 blijkt dat 61% van alle instellingen die een 486 heeft ($n=45$), ook over Windows 3.1x beschikt. Dit geldt voor 36% van alle instellingen die een 386 heeft ($n=29$).

7.4 Conclusies en Discussie

7.4.1 Conclusies

Aangezien er sprake is van een lage respons, vooral met betrekking tot de afzonderlijke instellingen, is enige voorzichtigheid geboden bij het interpreteren van onderstaande conclusies.

- Er wordt door de meeste instellingen binnen de VE gebruik gemaakt van toetsen. Het merendeel van de respondenten toetst twee keer per cursus.
- Toetsen worden over het algemeen afgenomen om cursisten te verwijzen naar een ander onderwijsniveau of als hulpmiddel bij de plaatsing. Ook worden toetsen regelmatig gebruikt om vorderingen van de cursisten bij te houden.
- 47% van alle respondenten maakt gebruik van in de handel verkrijgbare toetsen. Het merendeel van deze respondenten komt uit de BE (65%).
- 57% van alle respondenten ontwikkelt de gebruikte toetsen zelfstandig of in samenwerking met collega's. Veel van de respondenten die liever zelf toetsen ontwikkelen dan gebruik maken van toetsen die in de handel verkrijgbaar zijn, doen dit omdat zij van mening zijn dat laatstgenoemde toetsen niet aansluiten op de leerstof.
- Respondenten die zelf of in samenwerking met hun collega's toetsen ontwikkelen, besteden hier over het algemeen tussen de 1 en de 10 uur aan.
- Het merendeel van de respondenten ervaart alle kenmerken van adaptief toetsen als (zeer) belangrijk voor de eigen toetssituatie.
- Over het algemeen hebben de instellingen in de VE interesse voor adaptieve toetsen op het gebied van rekenen/wiskunde.
- De computerfaciliteiten in de VE zijn zowel in kwantitatief als in kwalitatief opzicht over het algemeen voldoende om adaptief te kunnen toetsen.

7.4.2 Discussie

Het blijkt dat instellingen binnen de VE veelvuldig toetsen. Het merendeel van de respondenten geeft aan dat elke cursist over het algemeen twee keer per cursus getoetst wordt.

In de afgelopen jaren zijn diverse toetspakketten voor de VE op de markt gebracht. Ondanks deze ontwikkelingen blijkt dat veel respondenten om uiteenlopende redenen toch zelf toetsen construeren. Een veel voorkomende reden die respondenten aangeven, is dat zij van mening zijn dat de toetsen niet goed aansluiten op de leerstof. Het construeren van toetsen, de afname en het nakijken ervan vergen veel tijd van docenten. De schattingen lopen zeer uiteen, maar liggen voor 75% van de respondenten, voor elke bovengenoemde activiteit afzonderlijk, tussen de 1 en de 10 uur per cursus. Aan het rapporteren van toetsprestaties en het onderzoeken van ontwikkelde toetsen wordt in mindere mate tijd besteed, tussen de 1 en 2 uur. Indien er van uitgegaan wordt dat gemiddeld 5 uur aan iedere activiteit wordt besteed, besteedt het merendeel van de respondenten ongeveer 15 uur aan activiteiten die betrekking hebben op toetsen. Als ze gebruik maken van adaptieve toetsen zullen docenten waarschijnlijk minder tijd kwijt zijn, maar de verschillen zullen niet spectaculair zijn. Dit roept enkele vragen op. Hoe nauwkeurig zijn de schattingen van de docenten over hun tijdsbesteding eigenlijk? Uit de resultaten blijkt dat er grote verschillen bestaan tussen de hoeveelheid tijd die respondenten zeggen te besteden aan de toetsactiviteiten. Het nauwkeurig schatten van de juiste tijdsbesteding zou voor respondenten moeilijk kunnen zijn aangezien het activiteiten betreft die een tijd geleden hebben plaatsgevonden. Een volgende vraag die opkomt, heeft te maken met de kwaliteit van de door de respondenten ontwikkelde toetsen. Is het mogelijk om in vijf uur een toets te ontwikkelen op basis waarvan valide beslissingen genomen kunnen worden over cursisten? De vele beschrijvingen van het toetsconstructieproces in de onderwijskundige literatuur laten eigenlijk geen andere conclusie toe dan dat dit onmogelijk is. Uit het onderzoek blijkt verder dat respondenten over het algemeen veel belang hechten aan de aspecten van adaptief toetsen. Deze positieve opinie over de kenmerken van adaptief toetsen komt ook tot uitdrukking in de belangstelling die respondenten zeggen te hebben voor adaptieve toetsen voor rekenen/wiskunde, indien deze op de markt komen. Een aantal instellingen is alleen geïnteresseerd indien de adaptieve toetsen voldoen aan één of meer voorwaarden.

Het blijkt dat instellingen beschikken over voldoende computerfaciliteiten. De computerinfrastructuur kan dus voor de meeste instellingen binnen de VE nauwelijks een belemmering vormen om het adaptief toetsstelsel in te voeren.

Literatuur

- Baker, F.B. (1985). *The basics of item response theory*. University of Wisconsin. Portsmouth, New Hampshire: Heinemann.
- Buning, G.R., Eggen, T.J.H.M., & Kelderman, H. (1986). Interpretaties van het Raschmodel. In W.J. van der Linden. *Het gebruik van het Raschmodel voor een decentraal toetsservicesysteem*. Vakgroep Onderwijskundige Meetmethoden en Data-analyse. Enschede: Universiteit Twente.
- Dijkman, W.M. (1993). *Didactiek en lerende volwassenen: Leren professionaliseren*. Leiden: Spruyt, Van Mantgem & De Does.
- Heuvelmans, A.P.J.M. (1992). *Constructie en verwerking van vragenlijsten*. OPD Memorandum 92-2. Arnhem: Cito, Instituut voor Toetsontwikkeling.
- Kingsbury, G.G., & Weiss, D.J. (1983). A comparison of IRT-based adaptive mastery testing and a sequential mastery testing procedure. In D.J. Weiss (Ed.), *New horizons in testing* (Pps. 257-286). New York: Academic Press.
- Kingsbury, G.G., & Zara, A.R. (1989). Procedures for selecting items for computerized adaptive testing. *Applied Measurement in Education*, 2, 359-375.
- Kingsbury, G.G., & Zara, A.R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education*, 4, 241-261.
- Lord, F.M. (1971). A theoretical study of two-stage testing. *Psychometrika*, 36, 227-242.
- Mellenbergh, G.H. (1986). Twintig jaar onderwijs-meet-kunde. In W.J. van der Linden (red.) *Moderne methoden van toetsenconstructie en toetsgebruik*. Bijdragen aan de Onderwijs Research Dagen. Lisse: Swets & Zeitlinger.
- Moerkerke, G., & Munsters, J. (1994). Over de inzet van adaptief testen in het onderwijs. *Tijdschrift voor Hoger Onderwijs*, 12(1), 33-43.
- Poelmans, P., & Martens, R. (1995). Toetsen per computer: Een geautomatiseerd toetsservicesysteem. *Tijdschrift Volwassenen Educatie*, 3, 46-49.
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (1992). *Numerical recipes in FORTRAN: The art of scientific computing* (2nd ed.). Cambridge: University Press.
- Reckase, M.D. (1983). A procedure for decision making using tailored testing. In: D.J. Weiss (Ed.) *New horizons in testing* (p.p. 237-255). New York: Academic Press.

- Sobel, M., & Wald, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Annals of Mathematical Statistics*, 20, 502-522.
- Spray, J.A., & Reckase, M.D. (1994). *The selection of test items for decision making with a computer adaptive test*. Paper presented at the national meeting of the National Council on Measurement in Education. New Orleans.
- Stocking, M.L., & Swanson, L. (1993). A method for severely constrained item selection in adaptive testing. *Applied Psychological Measurement*, 17, 277-292.
- Straetmans, G.J.J.M. (1993). Een itembank rekenen/wiskunde voor de Basiseducatie. *Tijdschrift Volwassenen Educatie*, 5, 29-32.
- Straetmans, G.J.J.M. (1994a). De constructie van een rekenvaardigheidsschaal voor de Basiseducatie. *Tijdschrift Volwassenen Educatie*, 3, 45-48.
- Straetmans, G.J.J.M. (1994b). Een plaatsingstoets rekenen/wiskunde voor de Basiseducatie. *Tijdschrift Volwassenen Educatie*, 4, 37-41.
- Straetmans, G.J.J.M. (1995a). Een rekenvaardigheidsschaal voor de basiseducatie: Beschrijving en toepassingen. *Over Rekenen* 7, 2, 61-68.
- Straetmans, G.J.J.M. (1995b) Adaptief toetsen: Beschrijving van een veelbelovend toetsconcept. *Tijdschrift Volwassenen Educatie*, 6, 13-17.
- Straetmans, G.J.J.M. (1995c). Het toetsen van vorderingen bij rekenen/wiskunde. *Tijdschrift Volwassenen Educatie*, 1, 35-40.
- Straetmans, G.J.J.M. (1995d). Adaptief toetsen: De hergeboorte van een oud principe. *Opleiding en Ontwikkeling*, 8(6), 23-29.
- Simpson, J.B., & Hetter, R.D. (1985). *Controlling item-exposure rates in computerized adaptive testing*. Paper presented at the annual conference of the Military Testing Association, San Diego.
- Thissen, D., & Mislevy, R.J. (1990) Testing algorithms. In H. Wainer (Ed.). *Computerized adaptive testing: A primer* (p.p. 103-135). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Valcke, M.M.A. (1992). Toetsservicesystemen: Nu ook al toetsen met de computer? *Onderwijs en Media*, 2, 54-67.
- Verhelst, N.D., & Kamphuis, F.H. (1989). *Statistiek met $\hat{\theta}$* . Bulletinreeks nr 77. Arnhem: Cito.
- Verhelst, N.D., Glas, C.A.W., & Verstralen, H.H.F.M. (1995). *One-parameter logistic model (OPLM)*. Arnhem: Cito.

- Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale (NJ): Lawrence Erlbaum Associates, Publishers.
- Wald, A. (1947). *Sequential analysis*. New York: Wiley.
- Warm, T.A. (1989). Weighted maximum likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.
- Weiss, D.J., & Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.

