

**De ontwikkeling van een instrument voor  
'Toets Curriculum Overlap'**

**A.A.H. van Abswoude**



OPD Memorandum

99-1

## **De ontwikkeling van een instrument voor 'Toets Curriculum Overlap'**

A.A.H. van Abswoude

Cito  
Arnhem, 1999

**Cito** Instituut voor Toetsontwikkeling  
Postbus 1034 6801 MG Arnhem  
**Bibliotheek**

8501 002 5782



Prijs f 15, --

© Cito, Arnhem 1999

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

## **Voorwoord**

Dit memorandum is geschreven in het kader van een werkstuk voor de vakgroep Psychologische Methodenleer van de Faculteit der Psychologie van de Universiteit van Amsterdam (UvA). Het uitgevoerde onderzoek heeft plaatsgevonden onder begeleiding van Drs. H.A. Moelands en Prof. Dr. P.F. Sanders.

Graag zou ik de medewerkers van de afdeling Onderzoek en Psychometrische Dienstverlening (OPD) die mij geholpen hebben bij dit werkstuk willen bedanken. In het bijzonder wil ik echter Ron Engelen bedanken voor zijn adviezen en hulp bij de analyses met itemresponsmodellen.

Sandra van Abswoude  
Arnhem, maart 1999



## Samenvatting

Dit memorandum doet verslag van de stappen die ondernomen zijn in de ontwikkeling van een instrument Toets Curriculum Overlap (TCO). TCO is gedefinieerd als de mate van overlap tussen de inhoud van een toets en de inhoud van een geïmplementeerd curriculum. Dit meetinstrument is ontwikkeld in het kader van het project zelfevaluatie basisonderwijs (het ZEBO-project), waarin basisscholen in staat worden gesteld zelf de kwaliteit van hun onderwijs te meten. Het is van belang TCO te kennen, omdat TCO een onbedoeld effect kan hebben op de te schatten vaardigheid van leerlingen en daardoor ook op de beoordeling van de kwaliteit van het onderwijs.

De volgende vragen staan in het onderzoek centraal: (1) Is er sprake van TCO? (2) Is het mogelijk een valide en betrouwbaar TCO-instrument te ontwikkelen? (3) Hoe moet een valide en betrouwbaar TCO-instrument vormgegeven worden? Om de laatste onderzoeksvraag te kunnen beantwoorden zijn drie methoden om TCO te meten gebruikt. Van algemeen naar specifiek zijn dit: het stellen van vragen over de lesmethoden die leerkrachten gebruiken, het stellen van vragen over omschrijvingen van de lesstof en het stellen van vragen over de items van de toets. Deze drie methoden om TCO te meten worden in het memorandum de lesmethoden-, de subcategorie- en de itemmethode genoemd. Welke meetmethode het meest in aanmerking komt om TCO te meten, kan op basis van onder andere betrouwbaarheid en validiteit besloten worden. Als toets is toets E3 van het pakket Rekenen-Wiskunde 1 van het Cito-leerlingvolgsysteem gebruikt.

Na twee kleinschalige proefafnamen is de TCO-vragenlijst bij een aselechte steekproef van 170 leerkrachten van groep drie afgenomen. Om de criteriumvaliditeit van de meetmethoden te kunnen vaststellen, zijn ook de antwoorden van alle 3305 leerlingen van deze leerkrachten op toets E3 verzameld.

Verschillende eigenschappen van de vragen van de meetmethoden zijn onderzocht. Eerst zijn enkele praktische of toetstechnische eigenschappen van de vragen onderzocht, zoals de eenduidige formulering en de belasting voor de leerkrachten. Vervolgens is een aantal psychometrische eigenschappen van de drie meetmethoden onderzocht. Uit de resultaten blijkt een combinatie van de itemmethode en een vraag naar de gebruikte lesmethode als TCO-instrument het meeste perspectief te bieden.



# Inhoudsopgave

Voorwoord

Samenvatting

Inhoudsopgave

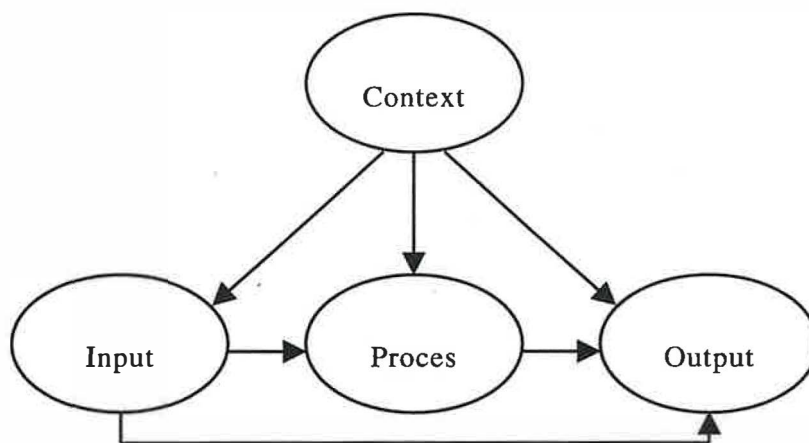
|     |                                     |    |
|-----|-------------------------------------|----|
| 1   | Inleiding                           | 1  |
| 2   | Doel en functie van het instrument  | 3  |
| 2.1 | Doel van het instrument             | 3  |
| 2.2 | Functie van het instrument          | 3  |
| 3   | Ontwikkeling van de vragenlijst TCO | 5  |
| 3.1 | De toets                            | 5  |
| 3.2 | De doelgroep                        | 5  |
| 3.3 | Methoden voor het meten van TCO     | 5  |
| 3.4 | De kwaliteitseisen                  | 7  |
| 4   | Vraagconstructie                    | 9  |
| 5   | Afname van het instrument           | 11 |
| 5.1 | Werving van leerkrachten            | 11 |
| 5.2 | Verwerking van de gegevens          | 11 |

|       |                                                                  |    |
|-------|------------------------------------------------------------------|----|
| 6     | Vaststellen van de eigenschappen van de vragen                   | 13 |
| 6.1   | Toetstechnische aspecten van de TCO-vragen                       | 13 |
| 6.1.1 | Eenduidige formulering                                           | 13 |
| 6.1.2 | Belasting voor de leerkracht                                     | 14 |
| 6.2   | Mate van Toets Curriculum Overlap                                | 15 |
| 6.2.1 | TCO bij de subcategorieën                                        | 15 |
| 6.2.2 | TCO bij de items                                                 | 17 |
| 6.2.3 | TCO bij de lesmethoden                                           | 19 |
| 6.2.4 | Conclusies over de mate van TCO                                  | 24 |
| 6.3   | Discriminerend vermogen                                          | 26 |
| 6.4   | Betrouwbaarheid                                                  | 26 |
| 6.5   | Dimensionaliteit                                                 | 27 |
| 6.5.1 | Twee-factormodel                                                 | 27 |
| 6.5.2 | Homogeniteitsanalyse                                             | 29 |
| 6.5.3 | Principale Componenten Analyse                                   | 31 |
| 6.6   | Schaalbaarheid van de drie meetmethoden                          | 34 |
| 6.6.1 | Het Mokkenmodel                                                  | 35 |
| 6.6.2 | Het Raschmodel                                                   | 35 |
| 6.6.3 | Het OPLM                                                         | 36 |
| 6.7   | Validering van TCO met toetsscores van leerlingen als criterium  | 38 |
| 6.7.1 | Validering van de items                                          | 39 |
| 6.7.2 | Validering van de subcategorieën                                 | 43 |
| 6.7.3 | Validiteit van de lesmethoden                                    | 44 |
| 6.8   | Detectie van vraagonzuiverheid                                   | 46 |
| 6.8.1 | Algemene calibratie van de items van toets E3                    | 46 |
| 6.8.2 | Vraagonzuiverheid samenhangend met de gebruikte lesmethode       | 47 |
| 6.8.3 | Vraagonzuiverheid samenhangend met TCO                           | 48 |
| 6.8.4 | Vraagonzuiverheid samenhangend met TCO,<br>itemgewijs onderzocht | 49 |
| 6.8.5 | Conclusies over vraagonzuiverheid                                | 50 |
| 7     | Conclusie en discussie                                           | 51 |
| 8     | Literatuur                                                       | 55 |
|       | Bijlagen                                                         | 57 |

## 1. Inleiding

In 1995 is het project 'Schoolzelfevaluatie in het basisonderwijs' (het ZEBO-project) gestart. Dit project is een gezamenlijke activiteit van het Instituut voor Toetsontwikkeling (Cito), het Instituut voor Leerplanontwikkeling (SLO) en het Onderzoekscentrum voor Toegepaste Onderwijskunde (OCTO). Het doel van het ZEBO-project is het ontwikkelen van een instrument dat scholen in staat stelt de kwaliteit van hun onderwijs te evalueren. De term 'schoolzelfevaluatie' duidt aan dat het de school zelf is die het initiatief voor de evaluatie neemt, en dat de school deze evaluatie vervolgens ook zelf uitvoert.

Uit de onderwijskundige literatuur blijkt dat vele variabelen aan de kwaliteit van het onderwijs gerelateerd zijn. Deze variabelen zijn te herleiden tot de input, het proces en de output van het onderwijs. Een onderwijskundig model dat de relatie tussen deze variabelen weergeeft, is het CIPO-model (Scheerens, 1990), een acroniem voor context, input, proces en output (zie figuur 1).



Figuur 1.1  
Het CIPO-model

Het onderzoek waar dit memorandum betrekking op heeft richt zich alleen op de proces-component van het CIPO-model. Een belangrijke procesvariabele is *Opportunity to learn* (OTL). OTL wordt gedefinieerd als de mate waarin leerlingen in staat zijn gesteld zich de lesstof eigen te maken.

Dit onderzoek is gericht op een deelgebied van OTL dat *Toets Curriculum Overlap* (TCO) wordt genoemd. TCO is de mate waarin de inhoud van een toets aansluit bij het door leerkrachten geïmplementeerd curriculum. Het slecht aansluiten van een toets bij het door een leerkracht verzorgde onderwijs kan een reden zijn dat de geschatte vaardigheid van leerlingen onjuist is, waardoor mogelijk verkeerde conclusies over de kwaliteit van het onderwijs getrokken worden. Valide en betrouwbare instrumenten om TCO te meten ontbreken. In dit onderzoek is getracht een valide en betrouwbaar TCO-meetinstrument te ontwikkelen.

De volgende vraagstellingen staan in dit onderzoek centraal:

1. Is er sprake van TCO?
2. Is het mogelijk een valide en betrouwbaar TCO-instrument te ontwikkelen?
3. Hoe moet een TCO-instrument vormgegeven worden?

In dit memorandum wordt de constructie van het instrument 'Toets Curriculum Overlap', beschreven. Hierbij zijn grotendeels de stappen die Eggen en Sanders in het boek 'Psychometrie in de Praktijk' voor het ontwikkelen van toetsen beschrijven, aangehouden (Eggen & Sanders, 1993).

De zes stappen zijn:

- specificatie van het doel en de functie van het instrument (Hoofdstuk 2);
- specificatie van de kenmerken van het instrument (Hoofdstuk 3);
- vraagconstructie (Hoofdstuk 4);
- afname van het instrument (Hoofdstuk 5);
- vaststellen van de eigenschappen van de vragen (Hoofdstuk 6);
- samenstellen van het instrument (Hoofdstuk 7).

In het memorandum worden begrippen gebruikt die in het dagelijks taalgebruik nagenoeg dezelfde betekenis hebben, maar die in dit onderzoek een specifieke betekenis hebben. Deze begrippen en hun betekenis staan in tabel 1.1. In de eerste kolom staat het begrip en in de tweede kolom staat de betekenis die het begrip in het memorandum heeft.

Tabel 1.1  
Belangrijke begrippen en hun betekenis in het verslag

| Begrip      | Betekenis                                                                                   |
|-------------|---------------------------------------------------------------------------------------------|
| Vragenlijst | Duidt op het conceptinstrument waarin drie methoden om TCO te meten zijn opgenomen.         |
| Instrument  | Duidt op het instrument dat uiteindelijk in het ZEBO-project gebruikt kan worden.           |
| Toets       | Duidt op toets E3 van het pakket Rekenen-Wiskunde 1 van het Cito-leerlingvolgsysteem (LVS). |
| Items       | Duiden op de opgaven van toets E3.                                                          |
| Vragen      | Duiden op de vragen van de TCO-vragenlijst.                                                 |

## 2. Doel en functie van het instrument

De eerste stap in de constructie van het instrument bestaat uit het vaststellen van het doel en de functie die het instrument moet vervullen. Eerst zal het doel en daarna de functie van het instrument beschreven worden.

### 2.1 Doel van het instrument

Het doel van het TCO-instrument is het meten van het construct 'Toets Curriculum Overlap' (TCO). Toets Curriculum Overlap is gedefinieerd als de overeenstemming tussen de inhoud van de toets en de inhoud van het door de leerkracht onderwezen curriculum. TCO betreft de maakbaarheid van een toets: hoe hoger de TCO-score, hoe maakbaarder de toets. Merk op dat lesstof die wel is onderwezen, maar geen onderdeel uitmaakt van de inhoud van de toets, buiten het bereik van de definitie van het begrip valt. Dit geldt ook voor lesstof die een leerling zichzelf eigen heeft gemaakt, of die anderen (de ouders, grootouders, etcetera) het kind hebben geleerd.

### 2.2 Functie van het instrument

De functie van het TCO-instrument is het verkrijgen van inzicht in de bijdrage van TCO bij de kwaliteit van het onderwijs van een school (zie Inleiding). Eén mogelijkheid is de kwaliteit van het onderwijs te voorspellen met behulp van een regressiemodel met een aantal voorspellende variabelen, die ieder een gewicht hebben en die één intercept delen. Eén van die voorspellende variabelen zou TCO kunnen zijn. Bij deze mogelijkheid zou bij het vaststellen van de kwaliteit van het onderwijs *achteraf*, dus ná afname van de toets voor TCO gecorrigeerd kunnen worden.

Een tweede mogelijkheid is *vooraf* met TCO rekening te houden, door te stellen dat eerst alle lesstof behandeld moet zijn voordat een toets afgenomen mag worden. Ook kan vooraf met TCO rekening gehouden worden door leerlingen alleen de maakbare opgaven van een toets te laten maken, en de vaardigheid van de leerlingen alleen op deze opgaven te baseren.

Op welke manier het instrument het beste gebruikt kan worden, wordt nader onderzocht.



### 3. Ontwikkeling van de vragenlijst TCO

Op basis van het in hoofdstuk twee beschreven doel en de functie van het instrument, is de TCO-vragenlijst ontwikkeld. Allereerst wordt in deze paragraaf ingegaan op *de toets* die gebruikt is om het instrument Toets Curriculum Overlap te ontwikkelen. Daarna wordt de doelgroep van de toets beschreven. Vervolgens worden drie methoden om TCO te meten behandeld. Ten slotte zullen de eisen waaraan het instrument moet voldoen besproken worden. De uiteindelijke vragenlijst zoals deze is afgenomen, staat in bijlage 1.

#### 3.1 De toets

Om te achterhalen of het mogelijk is een TCO-instrument te ontwikkelen (zie onderzoeksvraag 2) is toets E3 van het pakket Rekenen-Wiskunde 1 uit het Cito-Leerlingvolgsysteem (LVS) gebruikt. De toets wordt door scholen in groep 3 van het basisonderwijs gebruikt om de voortgang in rekenvaardigheid van leerlingen te meten. Voor toets E3 is gekozen omdat deze toets een hoge penetratiegraad van de LVS-toetsen kent en omdat van deze toets bekend is dat deze ééndimensioneel is (zie Handleiding bij E3, 1992).

#### 3.2 De doelgroep

In de vragenlijst zijn leerkrachten van groep drie gevraagd de items op TCO te beoordelen, omdat zij van alle betrokkenen het meeste zicht hebben op de overlap tussen een toets en het door hen onderwezen curriculum. Het instrument is bij deze doelgroep individueel en schriftelijk afgenomen. De instructie en de vragen moesten daarom voldoende uitleg bieden, niet dubbelzinnig geformuleerd zijn, en eenvoudig te hanteren zijn.

#### 3.3 Methoden voor het meten van TCO

Om te achterhalen hoe het TCO-instrument het best vormgegeven kan worden (onderzoeksvraag 3), zijn drie methoden om TCO te meten onderzocht.

De drie meetmethoden betreffen het stellen van vragen over:

- de lesmethoden die de leerkrachten gebruiken;
- de maakbaarheid van omschrijvingen van de lesstof (zogenaamde subcategorieën) in toets E3;
- de maakbaarheid van de specifieke items van toets E3.

Deze drie meetmethoden worden in dit verslag, respectievelijk **lesmethode**, **subcategorie** en **item** genoemd. Als over de drie gezamenlijk wordt gesproken, worden zij **meetmethoden** of **methoden** genoemd. De drie meetmethoden zijn in de vragenlijst TCO opgenomen. De redenen voor de drie meetmethoden, wat ze moeten opleveren en de problemen die bij de constructie van het instrument ondervonden zijn, worden eerst voor de subcategorie- en de itemmethode besproken en daarna voor de lesmethode.

## I Subcategorie

De lesstof van toets E3 is in de toetshandleiding onderverdeeld in vijf hoofdcategorieën en 14 subcategorieën (zie Handleiding bij E3). In de meetmethode die gebaseerd is op de subcategorieën van toets E3, beoordelen leerkrachten de TCO van deze subcategorieën. De *Toets Curriculum Overlap* van een subcategorie is in de vragenlijst TCO als volgt geoperationaliseerd. TCO wordt uitgedrukt in het begrip maakbaarheid. Een subcategorie van een toets is maakbaar als de omschrijving van de lesstof van een subcategorie, voor het kunnen maken van opgaven over het onderwerp, onderwezen is en de leerlingen er oefeningen mee gemaakt hebben. Het gaat er in deze definitie niet om of leerlingen de opgave van de toets *goed kunnen maken*, want dat impliceert een bepaalde vaardigheid van hen. In deze definitie is wel essentieel dat leerlingen *in staat zijn gesteld* — de mogelijkheid hebben gehad — de lesstof met betrekking tot de opgave te beheersen. Met de subcategoriemethode kan duidelijk worden welke delen van de lesstof van toets E3 door de leerkrachten maakbaar worden bevonden, en welke niet.

De subcategorieën worden in deze meetmethode door de leerkracht dichotoom gescoord: de omschrijving van een subcategorie is wel of niet maakbaar, òf er is wel of geen TCO. Omdat een derde optie moeilijk te verantwoorden is — een opgave is wel of niet maakbaar, niet ‘een beetje’ maakbaar — is deze achterwege gelaten. Bovendien is een aantrekkelijk kenmerk van dichotome antwoordmogelijkheden dat men twijfelaars tot een keuze dwingt.

Er kleven wel een aantal bezwaren aan deze meetmethode. Een eerste bezwaar is dat het ingewikkeld is de omschrijvingen zo te formuleren dat alle lezers het begrijpen en er hetzelfde onder verstaan. Een tweede bezwaar is dat leerkrachten zich door de moeilijkheid en niet door de maakbaarheid van de subcategorieën laten leiden. Een derde bezwaar is dat leerkrachten zich bij het beoordelen van de subcategorieën onbedoeld laten leiden door de concrete items van toets E3. Door het opnemen van een instructie is getracht met deze bezwaren rekening te houden. Een vierde bezwaar waar in de vragenlijstinstructie geen rekening mee gehouden kon worden is dat als een subcategorie niet maakbaar is beoordeeld, het onbekend is of de gehele subcategorie niet maakbaar is, of dat het slechts om een deel van de subcategorie gaat.

## II Item

In de tweede meetmethode beoordelen leerkrachten de maakbaarheid van de items van toets E3. De definitie van de maakbaarheid van de items is vergelijkbaar met die van de subcategorieën. Anders bij deze itemmethode is dat de concrete items van de toets voorgelegd worden in plaats van de omschrijvingen van de subcategorieën. De eerste drie bezwaren van de subcategoriemethode zijn ook van toepassing op de itemmethode. Het vierde bezwaar echter niet, omdat in deze meetmethode de maakbaarheid van alle items onderzocht wordt, waardoor de gehele lesstof die met toets E3 gemeten wordt, bestreken is.

Bij deze meetmethode kan men ervoor kiezen het construct TCO met alle items of een deel ervan (bijvoorbeeld met alleen kwalitatief sterke items) vast te stellen.

De eerste drie bezwaren die op de subcategoriemethode van toepassing waren, gelden ook voor deze meetmethode. Leerkrachten kunnen zich wederom door de moeilijkheid en niet door de maakbaarheid van de items laten leiden. Leerkrachten kunnen zich bij deze methode nog sterker dan bij de subcategorieën door onbedoelde eigenschappen van het item laten leiden in plaats van door de maakbaarheid ervan. Een extreem voorbeeld hiervan is dat leer-

krachten een item niet als maakbaar beoordelen omdat zijn of haar leerlingen het optellen van knikkers wel hebben gehad, maar het optellen van bonen niet.

### **III Lesmethode**

Met het stellen van vragen over de door leerkrachten gebruikte lesmethode wordt TCO op een geheel andere manier gemeten dan met de voorafgaande twee methoden. Leerkrachten worden niet gevraagd de maakbaarheid van de toets te beoordelen, maar zij worden alleen gevraagd aan te geven welke lesmethode (lesboeken) zij gebruiken. De gedachtegang is dat als lesmethoden verschillen in de mate waarin zij de inhoud van toets E3 dekken, dit terug te zien moet zijn in de toetsresultaten van leerlingen. Met het vragen naar de lesmethode zou men dus voldoende informatie over de TCO moeten hebben.

In de praktijk kleven er enkele bezwaren aan deze gedachtegang. Ten eerste sluit het feit dat lesstof niet in een methode staat niet uit dat leerkrachten deze lesstof toch behandelen. Ten tweede hoeft wat het ene jaar geldt niet altijd te gelden. TCO kan dus niet alleen tussen maar ook binnen lesmethoden verschillen. In de meetmethode zijn twee extra vragen opgenomen om deze bezwaren te ondervangen.

#### **3.4 De kwaliteitseisen**

De keuze voor één van de drie methoden en daarmee de keuze voor de vorm van het uiteindelijke instrument zal zowel op basis van psychometrische als toetstechnische overwegingen gebeuren.

Er zijn twee toetstechnische eisen aan het TCO-instrument gesteld. In de eerste plaats moeten de vragen zo geformuleerd zijn dat iedere leerkracht onder elke vraag hetzelfde verstaat. In de tweede plaats mag het instrument geen grote tijdsinvestering van de leerkrachten vergen. Om een beeld te krijgen hoeveel tijd het invullen van de meetmethoden de leerkrachten kost, is hen na iedere meetmethode gevraagd een tijdsindicatie te geven.

Daarnaast dienen ook psychometrische eisen aan het TCO-instrument gesteld te worden. De belangrijkste zijn dat het instrument TCO valide en betrouwbaar meet. De betrouwbaarheid van het instrument kan na afname van de TCO-vragenlijst berekend worden. Om de begripsvaliditeit van het instrument te bepalen, is de dimensionaliteit van de meetmethoden onderzocht en is getracht de vragen van de methoden op één schaal te plaatsen. Daarnaast zijn de resultaten van de leerlingen op toets E3 verzameld om de criteriumvaliditeit van de meetmethoden te kunnen vaststellen.



## 4. Vraagconstructie

Vraagconstructie bestond uit het schrijven van de vragen, het laten beoordelen van de vragen door medewerkers van de afdeling Onderzoek en Psychometrische Dienstverlening (OPD) van het Cito, en het nagaan of de bedoeling van de vragen helder was door deze in proefafnamen te laten beoordelen.

Er hebben twee proefafnamen van het instrument plaatsgevonden. De eerste vond plaats in maart 1998. Een prototype van het instrument, ontwikkeld door Moelands, werd afgenomen bij een leerkracht van een basisschool in Duiven. Uit de eerste proefafname bleek onder meer het belang van een heldere en eenduidige formulering en het steeds goed voor ogen houden van het doel van een vraag.

Nadat de bevindingen van de eerste proefafname verwerkt waren tot een nieuwe versie van de vragenlijst en het instrument intern door vijf medewerkers van de afdeling OPD gescreend was, is een tweede proefafname gehouden. Twee leerkrachten van een basisschool in Oss hebben in juni aan de proefafname van de tweede versie deelgenomen. De tekstinhoudelijke eigenschappen van de vragen zijn met de twee leerkrachten doorgenomen, waardoor het mogelijk was tekst die voor de leerkrachten onduidelijk was aan te passen.

De vragenlijst die na deze twee proefafnamen ontstond is in juli 1998 op grote schaal afgenomen. Dit testmateriaal bestond uit de volgende onderdelen (zie bijlage 1):

- twee instructiepagina's
- vragenlijst TCO
- twee antwoordformulieren voor toets E3



## **5. Afname van het instrument**

Bij de afname van de TCO-vragenlijst is een bepaalde werkwijze gevolgd. Deze werkwijze wordt in de onderstaande paragrafen beschreven. Eerst wordt de gevolgde procedure bij het werven van leerkrachten beschreven en daarna de manier waarop de gegevens verwerkt zijn.

### **5.1 Werving van leerkrachten**

Om voldoende waarnemingen per vraag te hebben voor de beoogde analyses, is getracht op een steekproef van ongeveer 200 leerkrachten uit te komen. Daartoe is een naar postcode-gebied gestratificeerde steekproef van 450 scholen getrokken uit een bestand van 1490 gebruikers van toets E3 van het pakket Rekenen-Wiskunde 1. De schoolleiding of de leerkrachten van deze 450 scholen zijn middels een brief uitgenodigd deel te nemen aan het onderzoek. Het antwoord kon men op een antwoordstrookje kosteloos terugsturen.

Er is een aantal voorwaarden voor deelname gesteld. Een eerste voorwaarde is dat de deelnemende scholen de toets — zoals in de toetshandleiding is geadviseerd — aan het einde van groep drie (mei, juni, of juli) afnemen. Hiervoor is gekozen om de tijd van afname als variabele constant te houden. Een tweede voorwaarde is dat slechts één leerkracht per school aan het onderzoek deelneemt. De kans is namelijk groot dat twee leerkrachten van een school een zelfde TCO-profiel opleveren, terwijl het doel van het onderzoek is verschillende TCO-profielen te meten. Daarom is er voor gekozen één leerkracht per school te laten deelnemen. Een derde voorwaarde is dat de leerkracht zelf het hele jaar voor de klas heeft gestaan. Dit is om er zo zeker mogelijk van te zijn dat de leerkracht weet wat wel en wat niet behandeld is.

Achtendertig procent van de aangeschreven leerkrachten hebben deelgenomen aan het onderzoek (170 van de aangeschreven 450). De relatief geringe respons is verklaarbaar. De vragenlijst is tegen het einde van het schooljaar afgenomen wanneer leerkrachten het erg druk hebben. Hierdoor hadden sommige leerkrachten geen tijd mee te werken. Acht leerkrachten bevestigden dit vermoeden door te informeren of zij later aan het onderzoek deel mochten nemen. Met deze leerkrachten is afgesproken dat zij alleen de ingevulde TCO-vragenlijst zouden opsturen. Eén vragenlijst kwam ver na de sluitingsdatum binnen en kon daarom niet verwerkt worden.

Van de 170 deelnemende leerkrachten zijn 3305 antwoorden van leerlingen op toets E3 verzameld. Niet van alle leerlingen waren de verzamelde antwoorden volledig. Als meer dan 25% van de antwoorden van een leerling ontbrak of onduidelijk was, zijn deze niet gebruikt in het onderzoek. Om deze redenen zijn de resultaten van 40 leerlingen niet gebruikt.

### **5.2 Verwerking van de gegevens**

Het verwerken van de antwoorden van de leerkrachten tot een digitaal databestand (ponsen) is door een extern bedrijf gebeurd. Om de foutenmarge te minimaliseren is dit twee keer gedaan. De vragenlijst en de twee scoreformulieren zijn tot drie asci-bestanden verponst. Het opschonen van de gegevens bestond grotendeels uit het koppelen van de drie bestanden en het nagaan of de ontbrekende waarnemingen werkelijk ontbraken of dat er sprake was van een ponsfout.



## 6. Vaststellen van de eigenschappen van de vragen

Voordat het samenstellen van het uiteindelijke instrument kan plaatsvinden, moeten eerst de eigenschappen van de geconstrueerde vragen van de drie meetmethoden — subcategorie, item, lesmethode — onderzocht worden. De resultaten worden in dit hoofdstuk beschreven.

In paragraaf 3.4 is beschreven dat de keuze voor één van de drie methoden op basis van toetstechnische en psychometrische kenmerken van de vragen gebeurt.

De toetstechnische kenmerken van de vragen van de meetmethoden zijn op twee manieren onderzocht. Eerst is bepaald of de vragen van de meetmethoden eenduidig geformuleerd zijn. Daarna is nagegaan of een methode geen onnodig grote belasting voor de leerkracht oplevert. Deze analyses staan beschreven in paragraaf 6.1.

De empirische karakteristieken van de vragen van de drie meetmethoden zijn onderzocht door eerst een overzicht te maken van de mate waarin er bij de drie meetmethoden sprake is van TCO. Dit is met name gedaan om te kunnen achterhalen of bij het schatten van de vaardigheid van leerlingen met TCO rekening gehouden moet worden. Met andere woorden: is een instrument TCO noodzakelijk? Dit gebeurt in paragraaf 6.2. Daarna zijn het discriminerend vermogen en de betrouwbaarheid van de vragen van de methoden bepaald. Dit wordt in de paragrafen 6.3 en 6.4 beschreven. Het onderzoek naar de dimensionaliteit en de schaling van de vragen van de drie meetmethoden, worden in de paragrafen 6.5 en 6.6 beschreven. De criteriumvaliditeit van de meetmethoden is nagegaan met behulp van de resultaten van leerlingen op toets E3. Dit onderzoek wordt in paragraaf 6.7 beschreven. Ten slotte wordt in de paragrafen 6.8 tot en met 6.12 *Differential Item Functioning* (DIF) bij de vragen van de subcategorie- en de itemmethode besproken.

### 6.1 Toetstechnische aspecten van de TCO-vragen

In deze paragraaf wordt besproken of de ene methode beter is dan de andere wanneer men twee toetstechnische eisen aan de vragen stelt. Eerst is onderzocht of de vragen van de drie meetmethoden verschillen in de mate waarin zij eenduidig te formuleren zijn. Daarna is onderzocht of de methoden verschillen in de mate van belasting die het beantwoorden van de vragen voor de leerkrachten vormt.

#### 6.1.1 Eenduidige formulering

Het eenduidig formuleren van een vraag is een noodzakelijke maar niet een voldoende voorwaarde voor een goed begrip van een vraag door proefpersonen. Het bepalen of de vragen van de drie methoden eenduidig te formuleren zijn, kan op verschillende momenten gebeuren. In de constructiefase van het instrument, dat wil zeggen, gedurende de proefafnamen van het instrument, zijn de vragen door wetenschappelijke medewerkers en leerkrachten inhoudelijk beoordeeld. Na afname van de vragenlijst kan achterhaald worden of de leerkrachten de vragen al dan niet op een zelfde manier beoordeeld hebben. De mate waarin de methoden tijdens de constructiefase van het instrument eenduidig te formuleren waren, wordt in deze paragraaf besproken. De mate waarin achteraf is gebleken dat de vragen van de drie meetmethoden eenduidigheid geformuleerd zijn, komt in de paragrafen 6.5 en 6.6 aan de orde.

Of de vragen van de drie meetmethoden even goed eenduidig te formuleren waren tijdens de constructie van het instrument is op basis van twee punten beoordeeld:

- de mogelijkheid dat leerkrachten het begrip van maakbaarheid onjuist interpreteren. Dit is bijvoorbeeld het geval als een leerkracht maakbaarheid als moeilijkheid interpreteert.
- de mogelijkheid dat leerkrachten zich laten leiden door specifieke itemkarakteristieken. Dit is het geval als leerkrachten zich door bijvoorbeeld de vormgeving van een item laten in plaats van door de inhoud.

Het eerste punt is van toepassing op de subcategorie- en de itemmethode, maar niet op de lesmethoden. Hoewel bij zowel de subcategorieën als de items het begrip 'maakbaarheid' gedefinieerd is, is het toch een reële mogelijkheid dat proefpersonen het begrip verkeerd interpreteren. Bij de lesmethoden wordt de maakbaarheid niet beoordeeld en speelt dat probleem dus niet.

Het tweede punt is wederom alleen op de subcategorie- en de itemmethode van toepassing. Bij de subcategorieën kan men zich door onbedoelde aspecten van de omschrijvingen van de subcategorieën laten leiden en bij de items door onbedoelde aspecten van de items.

Gezien het voorgaande lijkt het eenduidig interpreteren van lesmethoden op de beoordeelde punten geen problemen op te leveren. De subcategorie- en de itemmethode daarentegen blijken wel kwetsbaar voor een niet eenduidige interpretatie. Voorlopig moet daarom geconcludeerd worden dat het vragen naar de lesmethode de voorkeur verdient als TCO-instrument als men zich uitsluitend laat leiden door de mogelijkheid tot het eenduidig formuleren van de vragen. Later in het verslag wordt besproken of de pogingen tot het eenduidig formuleren van de vragen van de drie methoden daadwerkelijk geslaagd is (zie paragrafen 6.5 en 6.6).

### **6.1.2 Belasting voor de leerkracht**

Om een indruk te krijgen van de belasting voor de leerkracht van iedere methode is de leerkrachten gevraagd in minuten aan te geven hoeveel tijd het invullen kost. Doel van deze vraag is te achterhalen of een bepaalde methode dermate lang duurt dat het een te grote belasting voor de leerkracht oplevert. Vooral bij itemmethode kan de belasting erg groot zijn, omdat dan 53 keer beoordeeld moet worden of een item maakbaar is, terwijl dat bij de subcategorieën slechts 14 keer is. Vanwege de belasting voor de leerkracht zou men kunnen overwegen de itemmethode als potentieel TCO-instrument te laten afvallen.

Het invullen van de items duurde met een gemiddelde lengte van zeven minuten (gem.= 7.34; sd.= 3.52) nauwelijks langer dan het invullen van de subcategorieën. Voor dit onderdeel had men gemiddeld zes minuten nodig (gem.= 5.56; sd.= 4.05). Het invullen van de vragen over de lesmethoden duurde gemiddeld twee minuten (gem.= 2.22, sd.= 1.02). Dit verschil in tijd is dan ook geen reden om een van de onderdelen als instrument af te laten vallen.

Voorlopig moet geconcludeerd worden dat op basis van de twee toetstechnische eisen geen van de drie meetmethoden als mogelijk TCO-instrument afvalt. Toch moet men rekening houden met de mogelijkheid dat de betekenis van de items en de subcategorieën niet voor iedere leerkracht even duidelijk is geweest.

## 6.2 Mate van Toets Curriculum Overlap

Een tweede overweging voor het kiezen van een TCO-instrument, is de methode met de meest aantrekkelijke psychometrische kenmerken te kiezen. In de komende paragrafen zullen tal van psychometrische kenmerken van de vragen van de drie methoden besproken worden.

Een psychometrisch kenmerk dat wordt besproken, is de mate waarin sprake is van TCO in deze steekproef bij de drie meetmethoden. Als de toets in de steekproef perfect aansluit bij het onderwezen curriculum — alle leerkrachten zeggen dat de toets maakbaar is — dan hoeft er geen instrument voor het construct TCO ontwikkeld te worden. Als dit echter niet zo is, is het voor de constructie van het instrument van belang te weten of de drie methoden een vergelijkbaar beeld van TCO geven.

In de volgende drie paragrafen wordt eerst de mate van TCO bij de drie meetmethoden beschreven. Daarna wordt de mate van TCO van de vragen van de methoden met een kwaliteitscriterium vergeleken.

### 6.2.1 TCO bij de subcategorieën

Om de mate van TCO in de subcategorieën te achterhalen is een analyse over de veertien subcategorieën uitgevoerd. De resultaten hiervan staan in tabel 6.1. De analyse is uitgevoerd op de TCO-oordelen van 169 leerkrachten. Eén leerkracht liet de antwoorden op de subcategorieën open en is daarom niet in de tabel opgenomen.

In de eerste kolom van tabel 6.1 staan de categorieën waarin de lesstof van toets E3 onderverdeeld is. In de tweede kolom staan de bijbehorende subcategorieën. Het aantal ontbrekende waarnemingen per subcategorie staat in de derde kolom van de tabel. In de vierde kolom staat de p-waarde. De p-waarde van een subcategorie is de proportie leerkrachten die een subcategorie maakbaar vindt. In de vijfde kolom staat de standaardafwijking. Deze is maximaal .50, dat wil zeggen als de ene helft van de leerkrachten de subcategorie als maakbaar beoordelen en de andere helft de subcategorie als niet maakbaar beoordelen.

Uit tabel 6.1 valt af te leiden dat de proportie personen die een subcategorie maakbaar vindt erg hoog is: een gemiddelde subcategorie is door 88% van de leerkrachten als maakbaar beoordeeld. Omdat uit tabel 6.1 blijkt dat de overlap tussen de toets (E3) en het geïmplementeerde curriculum (de lessen van de leerkracht) groot is, lijkt TCO niet iets waar men zich ernstig zorgen over hoeft te maken.

De hoogte van de p-waarde van de subcategorieën levert niet alleen informatie over de mate van TCO, maar ook over de geschiktheid van de subcategorieën om als vragen in een TCO-instrument gebruikt te worden. Als er weinig itemvariantie is, meten de vragen niks. De verdeling van p-waarden in deze steekproef is scheef en dit levert problemen op bij het vaststellen van een construct TCO. Volgens Veldhuijzen, Goldebeld & Sanders (1993) en Goldebeld (1992, gerefereerd door Veldhuijzen, Goldebeld & Sanders) zijn p-waarden tussen .27 en .79 voldoende informatief, omdat ze dan een bevredigende itemvariantie opleveren. Als dit criterium gebruikt wordt om de kwaliteit van de subcategorieën te beoordelen, zijn slechts twee subcategorieën voldoende informatief (zie tabel 6.1). Dit zijn de subcategorieën *Splitsen* ( $p = .79$ ,  $sd = .41$ ) en *Tijd* ( $p = .59$ ,  $sd = .49$ ). In tabel 6.1 staat bij deze subcategorieën een '\*'.

Tabel 6.1  
Analyse over de subcategorieën

| Categorie         | Subcategorie                        | Missing | P-waarde | Sd. | R <sub>it</sub> | R <sub>ir</sub> |
|-------------------|-------------------------------------|---------|----------|-----|-----------------|-----------------|
| Tellen en ordenen | Structuur van de telrij             | 3       | .98      | .13 | .24             | .17             |
|                   | Resultatief en structurerend tellen | 1       | .97      | .17 | .15             | .06             |
|                   | Vergelijkingen en ordenen           | 2       | .95      | .21 | .35             | .24             |
| Structuren        | Splitsen*                           | 4       | .79      | .41 | .76             | .64             |
|                   | Samenstellen                        | 2       | .89      | .31 | .48             | .33             |
|                   | Aanvullen                           | 3       | .81      | .39 | .62             | .46             |
| Bewerkingen       | Optellen                            | 1       | .96      | .20 | .36             | .26             |
|                   | Aftrekken                           | 2       | .92      | .27 | .49             | .36             |
|                   | Diversen                            | 5       | .81      | .39 | .50             | .31             |
| Rekendictee       | Optellen                            | 2       | .97      | .17 | .33             | .25             |
|                   | Aftrekken                           | 3       | .96      | .19 | .35             | .26             |
|                   | Splitsen                            | 3       | .83      | .37 | .46             | .28             |
| Meten en tijd     | Metten                              | 3       | .88      | .32 | .31             | .14             |
|                   | Tijd*                               | 3       | .59      | .49 | .45             | .20             |

### Decielscores van de subcategorieën

Uit tabel 6.1 is gebleken dat de subcategorieën door een grote proportie leerkrachten als maakbaar zijn beoordeeld. Een overzicht van p-waarden geeft echter geen antwoord op de vraag hoe maakbaar leerkrachten de subcategorieën in totaliteit beoordelen. Waren er enkele leerkrachten die veel subcategorieën niet maakbaar vonden of vonden vrijwel alle leerkrachten één of twee subcategorieën niet maakbaar?

Om de resultaten van leerkrachten op de subcategorieën te kunnen interpreteren zijn de toetsscores van leerkrachten — het aantal maakbare subcategorieën per leerkrachten — in decielklassen gezet en zijn decielscores berekend. De indeling in decielklassen geeft een snelle indruk over de positie die een leerkracht in de totale groep van leerkrachten inneemt.

Tabel 6.2  
Indeling van de subcategorieën in decielklassen

| Decielklasse | Decielscore                                                         | Toetsscore |
|--------------|---------------------------------------------------------------------|------------|
| 1            | 9.1<br>10.4<br>11.4<br>12.2<br>12.8<br>13.2<br>13.5<br>13.9<br>13.9 | 0-9        |
| 2            |                                                                     | 10         |
| 3            |                                                                     | 11         |
| 4            |                                                                     | 12         |
| 5            |                                                                     |            |
| 6            |                                                                     | 13         |
| 7            |                                                                     |            |
| 8            |                                                                     | 14         |
| 9            |                                                                     |            |
| 10           |                                                                     |            |

In tabel 6.2 staan in de eerste kolom de decielklassen, in de tweede kolom de bij de decielklassen behorende decielscores op de subcategorieën en in de derde kolom staan de toetscores. De interpretatie is als volgt: wanneer een leerkracht een maakbaarheidsscore van 11 heeft, is in de eerste kolom af te lezen dat 30% van de leerkrachten hetzelfde aantal of minder subcategorieën maakbaar beoordeelde en 70% meer.

Net als bij de indeling in p-waarden lijkt TCO hoog: 50% van de leerkrachten heeft een toetsscore van 13 of 14. Het mogelijke negatieve effect van een lage TCO-score op de prestaties van leerlingen, zal men voornamelijk in de andere 50% moeten zoeken. Deze laatste groep leerkrachten vinden *tenminste* twee van de 14 subcategorieën niet maakbaar.

Kort samengevat blijkt de mate van TCO niet zo hoog als wat het in eerste instantie leek. De decielscores geven een minder gunstig beeld van de maakbaarheid van de subcategorieën, dan het beeld dat men van de p-waarden krijgt. Daarnaast blijken de subcategorieën slechts twee kwalitatief sterke vragen te bevatten als men voldoende variantie in antwoorden op de vragen eist.

### 6.2.2 TCO bij de items

In tabel 6.3 staan de resultaten van de analyse van de antwoorden van 170 leerkrachten op de items van toets E3. De opbouw van deze tabel is identiek aan de opbouw van tabel 6.1. Het verschil met tabel 6.1 is dat tabel 6.3 de resultaten van de analyse van de items betreft en niet van de subcategorieën.

De verdeling van de p-waarden van de items is, net als bij de subcategorieën, scheef verdeeld: de items zijn grotendeels als maakbaar beoordeeld. Slechts een klein deel van de items — 15 van de 53 items — is door minder dan 80% van de leerkrachten als maakbaar beoordeeld. Twee items worden door minder dan helft van de leerkrachten als maakbaar beoordeeld. Dit zijn item zes van deel A en item 29 van deel B. Vijftien items voldoen aan het in paragraaf 6.2.1 beschreven kwaliteitscriterium dat de p-waarde van het item tussen de .29 en de .79 moet liggen (zie tabel 6.3). Bij deze items staat een ‘\*’.

### Decielscores van de items

Bij de subcategorieën levert de indeling in decielklassen een ander beeld op over de maakbaarheid van toets E3 dan het overzicht van p-waarden. Het is de vraag of dit bij de items wederom het geval is. In tabel 6.4 staan toetsscores van leerkrachten op de items ingedeeld in decielklassen. De opbouw van tabel 6.4 is gelijk aan de opbouw van tabel 6.2. De indeling in decielklassen lijkt inderdaad een ongunstiger beeld op te leveren. Uit kolom vier van tabel 6.3 blijkt dat leerkrachten de items van toets E3 in het algemeen als maakbaar beoordelen. Uit tabel 6.3 blijkt dat 38 items door tenminste 80% van de leerkrachten als maakbaar zijn bevonden, en 15 door een lager percentage.

Uit tabel 6.4 daarentegen blijkt een ander beeld. Tachtig procent van de leerkrachten vindt vier tot maximaal 29 items niet maakbaar. De andere twintig procent vindt maximaal drie items niet maakbaar. In de praktijk kan dit gevolgen hebben voor de meetfout bij het schatten van de vaardigheid van leerlingen op basis van toets E3, omdat deze schatting gebaseerd is op items die bij sommige leerkrachten wel voorgelegd mogen worden en bij anderen niet.

Tabel 6.3  
Analyse van de items

| Deel | Item | Missing | P-waarde | Sd. | R <sub>it</sub> | R <sub>ir</sub> |
|------|------|---------|----------|-----|-----------------|-----------------|
| A    | 1    | 1       | .96      | .19 | .14             | .11             |
|      | 2    | 0       | .99      | .11 | .27             | .26             |
|      | 3    | 0       | .98      | .13 | .13             | .11             |
|      | 4    | 0       | .99      | .11 | .26             | .25             |
|      | 5    | 0       | .91      | .28 | .33             | .29             |
|      | 6*   | 1       | .41      | .49 | .35             | .28             |
|      | 7    | 0       | .95      | .21 | .37             | .34*            |
|      | 8    | 0       | .84      | .37 | .54             | .50*            |
|      | 9    | 0       | .92      | .27 | .35             | .31*            |
|      | 10   | 0       | .91      | .28 | .32             | .28             |
|      | 11*  | 1       | .77      | .42 | .57             | .52*            |
|      | 12   | 0       | .85      | .35 | .40             | .35*            |
|      | 13   | 0       | .84      | .37 | .43             | .38*            |
|      | 14   | 1       | .91      | .28 | .39             | .35*            |
|      | 15*  | 0       | .78      | .41 | .56             | .51*            |
|      | 16*  | 1       | .75      | .44 | .60             | .56*            |
|      | 17   | 1       | .86      | .34 | .45             | .40*            |
|      | 18   | 1       | .83      | .38 | .36             | .30*            |
|      | 19   | 0       | .92      | .27 | .15             | .11             |
|      | 20*  | 1       | .63      | .48 | .36             | .29             |
|      | 21   | 0       | .95      | .22 | .17             | .14             |
|      | 22   | 1       | .83      | .37 | .23             | .18             |
|      | 23*  | 0       | .76      | .42 | .16             | .10             |
|      | 24*  | 1       | .65      | .48 | .35             | .29             |
| B    | 1    | 1       | .99      | .11 | .14             | .13             |
|      | 2    | 1       | .99      | .11 | .14             | .13             |
|      | 3    | 1       | .98      | .13 | .21             | .19             |
|      | 4    | 1       | .98      | .13 | .21             | .19             |
|      | 5    | 1       | .98      | .15 | .12             | .9              |
|      | 6    | 2       | .96      | .20 | .29             | .26             |
|      | 7    | 2       | .98      | .13 | .21             | .19             |
|      | 8    | 1       | .83      | .37 | .46             | .41*            |
|      | 9*   | 1       | .78      | .41 | .53             | .48*            |
|      | 10   | 1       | .80      | .40 | .47             | .42*            |
|      | 11   | 0       | .91      | .28 | .29             | .25             |
|      | 12   | 0       | .97      | .17 | .34             | .32*            |
|      | 13   | 0       | .95      | .21 | .24             | .21             |
|      | 14   | 0       | .99      | .08 | -.1             | -.2             |
|      | 15   | 2       | .85      | .36 | .52             | .48*            |
|      | 16*  | 1       | .76      | .43 | .56             | .51*            |
|      | 17*  | 1       | .77      | .42 | .50             | .44*            |
|      | 18   | 0       | .92      | .27 | .33             | .30*            |
|      | 19*  | 0       | .79      | .40 | .55             | .51*            |
|      | 20   | 0       | .99      | .08 | .10             | .8              |
|      | 21   | 1       | .84      | .37 | .44             | .39*            |
|      | 22*  | 3       | .58      | .49 | .65             | .60*            |
|      | 23*  | 2       | .63      | .48 | .62             | .57*            |
|      | 24   | 0       | .94      | .24 | .28             | .24             |
|      | 25   | 0       | .81      | .39 | .41             | .36*            |
|      | 26   | 2       | .88      | .33 | .22             | .17             |
|      | 27   | 1       | .86      | .34 | .30             | .25             |
|      | 28*  | 2       | .62      | .49 | .45             | .39*            |
|      | 29*  | 1       | .47      | .50 | .54             | .48*            |

Tabel 6.4  
Indeling van de items in decielklassen

| Decielklasse | Decielscore                                                          | Toetsscore |
|--------------|----------------------------------------------------------------------|------------|
| 1            | 35.8<br>38.9<br>40.9<br>43.7<br>46.3<br>48.1<br>49.7<br>50.8<br>52.3 | 24-35      |
| 2            |                                                                      | 36-38      |
| 3            |                                                                      | 39-40      |
| 4            |                                                                      | 41-43      |
| 5            |                                                                      | 44-46      |
| 6            |                                                                      | 47-48      |
| 7            |                                                                      | 49         |
| 8            |                                                                      | 50         |
| 9            |                                                                      | 51-52      |
| 10           |                                                                      | 53         |

Samenvattend blijkt de mate van Toets Curriuculum Overlap afhankelijk van welke gegevens men gebruikt. Uit de proportie maakbare vragen krijgt men een goede indruk van de maakbaarheid van de items, maar uit de indeling van de antwoorden van leerkrachten in decielklassen blijkt dat een groot deel van de leerkrachten veel items als niet maakbaar beoordeelt. Daarnaast blijken 15 van de 53 een p-waarde tussen .29 en .79 te hebben en dus zijn er in verhouding meer kwalitatief sterke items dan kwalitatief sterke subcategorieën.

### 6.2.3 TCO bij de lesmethoden

Zoals in paragraaf 3.3 al is beschreven, is het meten van TCO door het vragen naar lesmethode van een geheel andere orde dan het vragen naar subcategorieën en items, omdat bij de lesmethode niet de maakbaarheid van de toets wordt beoordeeld. Bij de lesmethoden is de volgende aanpak gevolgd. Eerst is onderzocht of er bij het vaststellen van TCO met verstorende variabelen rekening gehouden moet worden. Deze variabelen zijn het 'anders gebruiken van de lesmethode' (gewoonte) en het 'meer of minder behandelen van onderwerpen (planning)' (zie paragraaf 3.3). Met een t-toets is berekend of deze variabelen een onbedoeld effect hebben op de rapportage van TCO. Daarna is de frequentie van het gebruik van de verschillende lesmethoden nagegaan. En ten slotte is onderzocht of en hoe de lesmethoden samenhangen met TCO. Dit is op twee manieren gedaan. Eerst door het verschil in gemiddelde TCO-score bij de lesmethoden te onderzoeken. Daarna door verschillen tussen de proporties maakbare subcategorieën en items bij de lesmethoden te onderzoeken.

#### **Twee verstorende variabelen: planning en gewoonte**

De manier waarop een leerkracht een lesmethode gebruikt, heeft mogelijk effect op de gerapporteerde TCO. Om deze reden zijn de variabelen 'planning' en 'gewoonte' (zie 6.2.3 en 3.3) opgenomen in de vragenlijst. Het betreft V-2 en V-3 van de vragenlijst (zie bijlage 1). Door naar de frequentie van voorkomen te kijken en verschillen in gemiddelde TCO te toetsen, is onderzocht of er bij het meten van TCO met deze variabelen rekening gehouden moet worden.

In tabel 6.5 is af te lezen hoeveel leerkrachten gebruik maken van de acht voorgecodeerde lesmethodemogelijkheden, uitgesplitst naar gewoonte (*gebruikelijk* en *uitzondering*) en planning (*alles, meer, minder, of meer en minder* behandeld). De waarnemingen zijn gebaseerd op de antwoorden van 167 leerkrachten.

In de eerste kolom van de tabel staan de lesmethoden. In de kolommen twee tot en met vijf staat het aantal leerkrachten dat alles, meer, minder, of meer en minder behandelen en dit op dezelfde manier deden als in het voorafgaande jaar. In de kolommen zes tot en met tien staat het aantal over dezelfde vier planningscategorieën maar dan voor die leerkrachten die het anders deden dan het voorafgaande jaar. In de elfde kolom staan de gesommeerde frequenties.

Het eerste wat opvalt aan tabel 6.5 is dat de opzet om de meest voorkomende lesmethoden voor te coderen, geslaagd is. Slechts vijf leerkrachten gebruikten een andere methode dan de voorgedecodeerde methoden (zie 'geen van deze' in de tabel). Twee leerkrachten begrepen de instructie niet en omcirkelden twee cijfers (zie bijlage 1). Verder geven met betrekking tot gewoonte 19 van de 167 leerkrachten aan niet dezelfde planning te volgen als het jaar daarvoor (zie kolommen 6 tot en met 9). Ongevraagd gaven enkele leerkrachten hiervoor een reden. Vier leerkrachten gaven aan voor het eerst aan groep drie les te geven en konden zich daarom niet aan de planning houden en één leerkracht was net met een nieuwe lesmethode begonnen. Er lijkt dus geen sprake van een 'uitzondering' te zijn omdat men de planning of het lesmateriaal niet geschikt vond, maar eerder om praktische redenen.

Tabel 6.5  
Gebruik van lesmethoden

| Lesmethode                 | Gebruikelijk |      |        |               | Uitzondering |      |        |               | Totaal |
|----------------------------|--------------|------|--------|---------------|--------------|------|--------|---------------|--------|
|                            | Alles        | Meer | Minder | Meer & Minder | Alles        | Meer | Minder | Meer & Minder |        |
| Wereld in getallen (oud)   | 8            | 5    |        | 3             |              | 1    |        | 2             | 18     |
| Wereld in getallen (nieuw) | 33           | 8    |        | 3             | 3            |      |        |               | 47     |
| Operatoir Rekenen (oud)    | 2            | 3    | 2      | 2             | 1            |      |        |               | 10     |
| Operatoir Rekenen (nieuw)  | 3            |      |        | 2             | 2            |      |        |               | 7      |
| Pluspunt                   | 29           | 7    | 1      | 4             | 3            |      |        |               | 44     |
| Rekenen en Wiskunde        | 12           | 10   |        | 7             |              | 2    |        | 2             | 33     |
| Geen van deze              | 1            | 1    |        | 1             | 2            |      |        |               | 5      |
| Dubbel antwoord            | 1            |      |        |               | 1            |      |        |               | 2      |
| Totaal                     | 89           | 34   | 3      | 22            | 12           | 3    |        | 4             | 167    |

Hoewel antwoorden op gewoonte niet gelijkmatig verdeeld zijn — weinig leerkrachten vonden dat er sprake was van een uitzondering — is toch een t-toets uitgevoerd om een indruk te kunnen vormen van het effect van deze variabele op TCO. Het verschil in gemiddelde maakbaarheid is berekend tussen leerkrachten die de lesmethode op dezelfde manier gebruikten (gebruikelijk) en leerkrachten die de lesmethode niet op dezelfde manier gebruikten (uitzonderlijk) als het jaar ervoor. Er bleek geen verschil [ $F(1) = .007$ ,  $p = .932$ ;  $t(23) = -.119$ ,  $p = .906$ ] in maakbaarheid. Gewoonte is daarom niet in verdere analyses opgenomen.

Met betrekking tot de 'planning' is in tabel 6.5 te zien dat de meeste leerkrachten zich aan de planning van de methode houden (60.5%). Zij behandelen *alles*. Een aantal behandelt *meer* (22.2%), bijna niemand behandelt *minder* (1.8%) en een aantal doet zowel *meer* als *minder* (15.6%). Opvallend is dat leerkrachten die de lesmethode Rekenen en wiskunde gebruiken, zich het minst aan de planning van de lesmethode lijken te houden.

Het verschil in maakbaarheidsscore, dat wil zeggen de totaalscore op de items, tussen de vier planningscategorieën is met zes univariate t-toetsen bepaald. Er is geen verschil gevonden tussen het volgens de planning (a), het meer (b), het minder (c), of het zowel meer als minder (d) behandelen van lesstof. De resultaten van de analyse staan in tabel 6.6.

Tabel 6.6  
Resultaten van de t-toets over de vier planningscategorieën

|       | f-toets | p-waarde | t-toets | df | Sign. |
|-------|---------|----------|---------|----|-------|
| a — b | .436    | .497     | .000    | 75 | 1.00  |
| a — c | .234    | .630     | .677    | 02 | .556  |
| a — d | .798    | .373     | 1.197   | 36 | .239  |
| b — c | 1.114   | .298     | .672    | 02 | .567  |
| b — d | .025    | .876     | -.306   | 02 | .785  |
| c — d | 2.620   | .111     | 1.107   | 45 | .274  |

Geen *overall* verschil betekent niet dat er als er gericht naar de data gekeken wordt nog steeds geen verschillen in maakbaarheid gevonden worden. Er zou zowel bij specifieke items als binnen methoden een effect kunnen zijn. Dat kan echter nauwelijks getoetst worden vanwege het geringe aantal waarnemingen. Toch is bij de lesmethode met de meeste waarnemingen in de groepen — bij Rekenen en wiskunde — het verschil in TCO getoetst. Er bleek geen significant verschil.

Vanwege het geringe aantal waarnemingen heeft verder onderzoek naar mogelijke effecten van de variabelen planning en gewoonte geen zin. Voorlopig wordt op basis van de wel uitgevoerde analyses aangenomen dat de nevenvariabelen geen of geen noemenswaardig effect hebben op TCO. In verdere analyses met TCO wordt daarom voor gewoonte, noch voor planning gecontroleerd.

### TCO bij de lesmethoden

Om een verschil in maakbaarheid tussen de lesmethoden te kunnen vaststellen, is eerst naar de gemiddelde TCO-scores gekeken. Daarna is op vraagniveau de mate van TCO tussen de lesmethoden onderzocht.

De gemiddelde maakbaarheidsscore, de spreiding en de betrouwbaarheid (Cronbach's alpha) van de lesmethoden staan in tabel 6.7. Bij bepaalde lesmethoden — Operator rekenen (oud & nieuw) en Pluspunt — is de door de leerkrachten gerapporteerde overlap groter dan bij andere lesmethoden — Wereld in getallen (oud en nieuw) en Rekenen en wiskunde. Groot zijn deze verschillen echter niet.

Tabel 6.7  
Gemiddelde maakbaarheidsoordelen per methode

| Lesmethoden                   | Gem. | Sd. | Alpha |
|-------------------------------|------|-----|-------|
| De wereld in getallen (oud)   | 43.3 | 6.3 | .87   |
| De wereld in getallen (nieuw) | 44.5 | 6.6 | .89   |
| Operator rekenen (oud)        | 47.8 | 4.2 | .84   |
| Operator rekenen (nieuw)      | 47.9 | 4.9 | .86   |
| Pluspunt                      | 47.0 | 5.5 | .87   |
| Rekenen en wiskunde           | 42.1 | 6.9 | .88   |

Om op vraagniveau — de vragen van de TCO-vragenlijst — verschillen in de mate van TCO tussen de lesmethoden te ontdekken, is een overzicht gemaakt van de p-waarde van de items bij de zes voorgecodeerde lesmethoden. Dit overzicht is voor de subcategorieën te vinden in en voor de items in bijlage 2. Aan de hand van deze bijlage zijn tabellen 6.8 en 6.9 opgesteld, waarin items staan die voor bepaalde lesmethoden *opvallend minder* maakbaar zijn dan voor de andere lesmethoden. De volgende regel is toegepast: een item is bij een (of meerdere) lesmethode opvallend minder maakbaar als dit item (of: deze items) bij die lesmethode een p-waarde heeft die tenminste ééntiende lager is dan de p-waarde bij de overige lesmethoden. Deze regel is toegepast, tenzij de p-waarde tussen de lesmethoden zo sterk uiteenloopt dat niet beweerd kan worden dat een(of twee) item(-s) bij een lesmethode opvallend minder maakbaar is dan een andere. Eerst zullen opvallende verschillen tussen lesmethoden op subcategorieën besproken worden en daarna op items.

In de eerste kolom van tabel 6.8 staan de bij bepaalde lesmethoden minder maakbare subcategorieën aangeduid met een '★'. Alleen subcategorieën waar tenminste één van de lesmethoden opvallend minder maakbaar is bevonden, zijn opgenomen in de tabel. Slechts enkele subcategorieën worden bij een of meerdere lesmethoden minder maakbaar bevonden (zie tabel 6.8). De lesmethode Rekenen en wiskunde valt in de tabel het meeste op.

Tabel 6.8  
Overzicht van minder maakbare subcategorieën

| Subcategorie | Wereld in<br>getallen<br>(oud)<br>n= 19 | Wereld in<br>getallen<br>(nieuw)<br>n= 47 | Operator<br>rekenen<br>(oud)<br>n= 10 | Operator<br>rekenen<br>(nieuw)<br>n= 7 | Pluspunt<br>n= 46 | Rekenen<br>en wis-<br>kunde<br>n= 34 |
|--------------|-----------------------------------------|-------------------------------------------|---------------------------------------|----------------------------------------|-------------------|--------------------------------------|
| Aanvullen    | ★                                       | ★                                         |                                       |                                        |                   |                                      |
| Splitsen     |                                         |                                           |                                       |                                        |                   | ★                                    |
| Meten        |                                         |                                           |                                       |                                        |                   | ★                                    |
| Tijd         |                                         |                                           |                                       |                                        |                   | ★                                    |

Tabel 6.9 is op een vergelijkbare manier opgezet. Anders is dat de eerste twee kolommen het toetsdeel en het itemnummer bevatten.

Een voorbeeld van een item dat verschillend beoordeeld wordt, is item 8 uit deel A. Item 8 was met een p-waarde van .84 in tabel 6.3 niet opvallend, hoewel de spreiding aan de hoge kant was. Negenentwintig items zijn volgens de gehanteerde regel voor één of meer lesmethoden opvallend minder maakbaar te noemen. Uit de tabel blijkt dat leerkrachten die Wereld in getallen gebruiken meer items als minder maakbaar beoordelen dan leerkrachten die andere lesmethoden gebruiken. Daarna volgen Rekenen en wiskunde en Pluspunt. Gezien het kleine aantal waarnemingen bij Operator rekenen, zijn er geen conclusies over de meer of minder maakbaarheid van deze lesmethode te trekken.

Samenvattend, blijken in verhouding meer items minder als maakbaar beoordeeld worden dan subcategorieën. Als men de twee tabellen 6.8 en 6.9 met elkaar vergelijkt, blijken de als minder maakbaar beoordeelde items wel min of meer overeen te stemmen met de als minder maakbaar beoordeelde subcategorieën. Zo blijken bijvoorbeeld leerkrachten die Wereld in getallen (oud en nieuw) gebruiken de maakbaarheid van subcategorie 'aanvullen' minder maakbaar te vinden dan de leerkrachten van andere lesmethoden en dit is in de items die tot deze subcategorie behoren (items 10 en 11 van deel A en 19 van deel B) terug te zien. Waarom sommige lesmethoden als minder maakbaar worden beoordeeld, is niet onderzocht.

Een inhoudelijke analyse van de subcategorieën, de items en de lesmethoden zou hierover uitsluitsel kunnen geven. Deze analyse, valt echter buiten het bereik van deze studie.

Tabel 6.9  
Overzicht van minder maakbare items per lesmethode

| Deel van<br>toets E3 | Item | Wereld in<br>getallen<br>(oud)<br>n= 19 | Wereld in<br>getallen<br>(nieuw)<br>n= 47 | Operator<br>rekenen<br>(oud)<br>n= 10 | Operator<br>rekenen<br>(nieuw)<br>n= 7 | Pluspunt<br>n= 46 | Rekenen en<br>wiskunde<br>n= 34 |
|----------------------|------|-----------------------------------------|-------------------------------------------|---------------------------------------|----------------------------------------|-------------------|---------------------------------|
| A                    | 05   |                                         |                                           |                                       | ★                                      |                   |                                 |
|                      | 08   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 10   |                                         |                                           |                                       |                                        |                   | ★                               |
|                      | 11   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 12   |                                         |                                           |                                       | ★                                      |                   |                                 |
|                      | 13   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 14   | ★                                       |                                           |                                       |                                        |                   |                                 |
|                      | 15   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 16   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 17   | ★                                       |                                           |                                       |                                        |                   | ★                               |
|                      | 20   |                                         |                                           |                                       |                                        | ★                 | ★                               |
|                      | 22   |                                         |                                           | ★                                     | ★                                      |                   | ★                               |
|                      | 23   |                                         |                                           |                                       |                                        | ★                 | ★                               |
|                      | 24   |                                         |                                           |                                       |                                        |                   | ★                               |
| B                    | 08   |                                         |                                           |                                       |                                        | ★                 | ★                               |
|                      | 09   |                                         |                                           |                                       |                                        | ★                 | ★                               |
|                      | 10   |                                         |                                           |                                       |                                        | ★                 | ★                               |
|                      | 15   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 16   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 19   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 21   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 22   | ★                                       | ★                                         |                                       |                                        |                   | ★                               |
|                      | 23   | ★                                       | ★                                         |                                       |                                        |                   |                                 |
|                      | 25   |                                         |                                           |                                       |                                        |                   | ★                               |
|                      | 26   |                                         |                                           | ★                                     |                                        |                   |                                 |
|                      | 27   |                                         |                                           |                                       |                                        |                   | ★                               |
|                      | 28   |                                         |                                           |                                       | ★                                      |                   | ★                               |
|                      | 29   | ★                                       | ★                                         |                                       |                                        | ★                 | ★                               |

### Overeenstemming in TCO tussen de subcategorie- en de itemmethode

In de voorafgaande paragrafen is de overlap in TCO tussen lesmethoden en subcategorieën en tussen lesmethoden en items besproken. Wat de overeenkomsten en verschillen in TCO tussen de subcategorieën en items zijn, is nauwelijks aan de orde gekomen. In deze paragraaf zal dit besproken worden. Hiertoe is allereerst de Pearsoncorrelatie tussen de TCO van de subcategorieën en de items berekend en daarna zijn de overeenkomsten en verschillen met een tabel van p-waarden zichtbaar gemaakt.

De overlap tussen de maakbaarheid van de subcategorieën en de maakbaarheid van de items van toets E3 lijkt niet erg groot. De Pearsoncorrelatie tussen de subcategorieën en de items is .64. Indien de twee methoden hetzelfde construct TCO zouden meten, TCO, zouden ze perfect moeten correleren. Aangezien de correlatie niet perfect is, meten de methoden òf verschillende constructen, òf meerdere constructen, òf zijn de methoden behept met grote meetfouten.

Om de verschillen tussen de twee meetmethoden te achterhalen, is een overzicht gemaakt van de p-waarden van de subcategorieën van de items (zie tabel 6.10). In de eerste kolom staan de subcategorieën en hun p-waarden in de TCO-vragenlijst. In de tweede kolom staan de items en het toetsonderdeel waartoe ze behoren. In de derde kolom staan de p-waarden van de vragen van de itemmethode van de TCO-vragenlijst. In de vierde kolom staan de p-waarden van de items op toets E3. In de tabel zijn de subcategorieën die niet inhoudelijk tot dezelfde categorieën behoren door dubbele strepen van elkaar gescheiden.

Als men de p-waarden van de subcategorieën en de daartoe behorende items vergelijkt, ziet men dat de overlap in het algemeen groot is. Dat wil zeggen: items die inhoudelijk behoren tot de subcategorieën die door een grote proportie leerkrachten als erg maakbaar beoordeeld zijn, worden ook door een grote proportie leerkrachten als maakbaar beoordeeld en items die behoren tot de minder maakbare subcategorieën worden gemiddeld minder als maakbaar beoordeeld. Er zijn slechts enkele items die van dit algemene patroon afwijken. Dat zijn items **6** en **20** van deel A en items **22**, **23** en **28** van deel B van toets E3 (zie tabel 6.10). De reden waarom deze items afwijken is niet meteen duidelijk. Deze items zouden vanwege hun moeilijkheid — ze vergen meer vaardigheid — of vanwege hun inhoud — ze meten een ander construct — kunnen afwijken. In de tabel is te zien dat deze items gemiddeld moeilijker voor de leerlingen zijn. Inhoudelijk redenen voor het afwijken van de items liggen niet voor de hand. Een alternatieve verklaring is dat deze items opvallen door de scheve verdeling van p-waarden in de steekproef. De items vallen op omdat deze items de enige zijn die door een relatief groot aantal leerkrachten niet als maakbaar zijn beoordeeld. Een implicatie hiervan is dat wanneer de verdeling van TCO-oordelen niet zo scheef is als bij toets E3 het geval was, de overeenstemming tussen de vragen van de subcategorie- en de itemmethode veel lager zou kunnen worden.

#### **6.2.4 Conclusies over de mate van TCO**

Uit de voorafgaande analyses kunnen enkele voorlopige conclusies over de mate van maakbaarheid van de drie methoden getrokken worden. Een eerste belangrijk resultaat was dat de leerkrachten de subcategorieën gemiddeld meer maakbaar vonden dan de items zelf. Een tweede resultaat was dat de maakbaarheidsoordelen verschillen tussen leerkrachten die verschillende lesmethoden gebruiken. Een derde resultaat, ten slotte, is dat er niet alleen tussen maar ook binnen de lesmethoden verschillen zichtbaar zijn tussen de maakbaarheidsoordelen over de subcategorieën en de items.

Uit deze conclusies blijkt dat de methode die gebruikt wordt om TCO te meten van belang is. Dit is omdat de mate waarin TCO vastgesteld wordt tussen de meetmethoden verschilt. De itemmethode lijkt volgens de voorgaande paragrafen (6.2 tot en met 6.2.3) het meest perspectief te bieden, omdat deze methode de meeste kwalitatief sterke vragen heeft. Als men naar de p-waarde kijkt ( $.29 < p < .79$ ) is dat de itemmethode: 28.3% van de items zijn kwalitatief goed tegenover 14.3% van de subcategorieën. Bij de lesmethoden is het niet zinnig om over p-waarde te praten. Toch lijkt het vaststellen van de gebruikte lesmethode van belang, omdat de gerapporteerde TCO zowel tussen als binnen lesmethoden lijkt te verschillen.

Tabel 6.10  
P-waarden van de subcategorie- en de itemmethode

| Subcategorie (& p-waarde TCO)             | Deel & Itemnr. | p-waarde TCO | p-waarde LVS |
|-------------------------------------------|----------------|--------------|--------------|
| Structuur van de telrij (.98)             | A_01           | .96          | .83          |
|                                           | A_02           | .99          | .95          |
|                                           | A_04           | .99          | .95          |
|                                           | B_12           | .97          | .88          |
|                                           | B_14           | .99          | .91          |
| Resultatief en structurerend tellen (.97) | A_03           | .98          | .91          |
|                                           | B_11           | .91          | .69          |
| Vergelijkingen en ordenen (.95)           | A_05           | .91          | .87          |
|                                           | <b>A_06</b>    | <b>.41</b>   | <b>.54</b>   |
|                                           | B_13           | .95          | .96          |
| Splitsen (.79)                            | A_07           | .95          | .87          |
|                                           | A_08           | .84          | .85          |
|                                           | B_15           | .85          | .79          |
|                                           | B_16           | .76          | .73          |
| Samenstellen (.89)                        | A_09           | .92          | .92          |
| Aanvullen (.81)                           | A_10           | .91          | .91          |
|                                           | A_11           | .77          | .77          |
|                                           | B_17           | .77          | .85          |
|                                           | B_18           | .92          | .83          |
|                                           | B_19           | .79          | .90          |
| Optellen (.96)                            | A_12           | .85          | .88          |
|                                           | A_13           | .84          | .81          |
|                                           | A_14           | .91          | .96          |
|                                           | B_20           | .99          | .95          |
|                                           | B_21           | .84          | .81          |
| Aftrekken (.92)                           | A_15           | .78          | .78          |
|                                           | A_16           | .75          | .65          |
|                                           | A_18           | .83          | .83          |
|                                           | <b>B_22</b>    | <b>.58</b>   | <b>.63</b>   |
|                                           | <b>B_23</b>    | <b>.63</b>   | <b>.62</b>   |
|                                           | B_24           | .94          | .86          |
|                                           | B_25           | .81          | .74          |
| Diversen (.81)                            | A_17           | .86          | .82          |
|                                           | A_19           | .92          | .85          |
|                                           | <b>A_20</b>    | <b>.63</b>   | <b>.66</b>   |
| Optellen (.97)                            | B_01           | .99          | .99          |
|                                           | B_02           | .99          | .98          |
|                                           | B_03           | .98          | .95          |
| Aftrekken (.96)                           | B_04           | .98          | .94          |
|                                           | B_05           | .98          | .95          |
|                                           | B_06           | .96          | .86          |
|                                           | B_07           | .98          | .92          |
| Splitsen (.83)                            | B_08           | .83          | .75          |
|                                           | B_09           | .78          | .72          |
|                                           | B_10           | .80          | .78          |
| Meten (.88)                               | A_21           | .95          | .86          |
|                                           | A_22           | .83          | .82          |
|                                           | B_26           | .88          | .82          |
|                                           | B_27           | .86          | .80          |
|                                           | <b>B_28</b>    | <b>.62</b>   | <b>.72</b>   |
| Tijd (.59)                                | A_23           | .76          | .91          |
|                                           | A_24           | .65          | .75          |
|                                           | B_29           | .47          | .43          |

### 6.3 Discriminerend vermogen

De item-testcorrelatie ( $R_{it}$ ) is de product-momentcorrelatie tussen de score op een vraag en de totaalscore op een toets. Deze maat wordt gebruikt om een indruk te krijgen van de mate waarin een vraag onderscheid maakt tussen een hoge en een lage toetsscore. Een hoge  $R_{it}$ -score betekent dat leerkrachten met een hoge toetsscore de vraag als maakbaar beoordelen en dat de vraag relatief veel bijdraagt aan de betrouwbaarheid van de toets. Een vergelijkbare maat is de item-restcorrelatie ( $R_{ir}$ ). Deze maat is de produkt-momentcorrelatie tussen de vraag en *de rest* van de toets.

Van de subcategorieën en de items van de TCO-vragenlijst zijn de  $R_{it}$ 's en de  $R_{ir}$ 's berekend. Onderzocht is hoeveel items van een methode een *goed* ( $.30 < R_{ir}$  of  $R_{it} < .39$ ) of *zeer goed* ( $R_{ir}$  of  $R_{it} > .40$ ) discriminerend vermogen hebben.

Het discriminerend vermogen van de subcategorieën is te vinden in de kolommen zes en zeven van tabel 6.1. Volgens de  $R_{it}$ -maat, die een enigszins geflatteerd beeld van het discriminerend vermogen geeft, hebben vijf subcategorieën een goed en zeven een zeer goed discriminerend vermogen. Volgens de  $R_{ir}$ -maat geven drie subcategorieën een goed en twee subcategorieën een zeer goed discriminerend vermogen. Als geëist wordt dat de p-waarde tussen .29 en .79 ligt en de  $R_{ir}$  boven .30, blijkt slechts één van de veertien subcategorieën kwalitatief goed. Deze subcategorie is 'splitsen'.

Het discriminerend vermogen van de items is te vinden in kolom zes en zeven van tabel 6.3. De  $R_{it}$  van twaalf items is goed en van 20 items zeer goed. De  $R_{ir}$  van elf items is goed en van 15 items zeer goed. Wanneer uitgegaan wordt van de eerder genoemde kwaliteitseisen over de p-waarde en het discriminerend vermogen, blijven elf van de 53 items over. Deze items zijn: items 11, 15 en 16 van deel A van toets E3 en items 9, 16, 17, 19, 22, 23, 28 en 29 van deel B. Het blijkt dus dat twee subcategorieën, en 21 items afvallen als men van de vragen van een meetmethode een  $R_{it}$ -waarde boven de .30 eist. Bij de items zijn dat groten-deels de items met een extreme p-waarde (d.w.z.  $> .90$ ).

Samenvattend valt er dus veel voor te zeggen alleen items met een niet extreme p-waarde te gebruiken indien de itemmethode als TCO-instrument gebruikt wordt. Een noodzakelijk vervolgonderzoek is dan: het verwijderen van items met een extreme p-waarde ( $> .90$ ) en opnieuw het discriminerend vermogen te bepalen om te achterhalen of er dan nog items zijn die wat betreft het discriminerend vermogen benedenmaats zijn.

### 6.4 Betrouwbaarheid

De betrouwbaarheid van de drie meetmethoden is op twee manieren vastgesteld. Eerst is van elke methode de betrouwbaarheid (Cronbach's alpha) berekend. Daarna is de bijdrage aan de betrouwbaarheid van de afzonderlijke subcategorieën (en items) berekend. Dit is gedaan door het verschil te bepalen tussen de waarde van alpha gegeven alle subcategorieën (of items) en alpha gegeven dat de subcategorie (of items) niet in het onderdeel wordt opgenomen. Van de afzonderlijke subcategorieën en items is vervolgens geëist dat zij een positieve bijdrage aan Cronbach's alpha leveren.

De betrouwbaarheid van alle items gezamenlijk is met een alpha van .88 hoger dan de alpha van .66 van de subcategorieën. De standaardmeetfouten van de methoden, zijn respectievelijk 2.26 en 1.08. In tabel 6.7 kan men zien dat de betrouwbaarheden van de diverse lesmethoden elkaar niet veel ontlopen. Met betrekking tot de betrouwbaarheid van de drie methoden kan geconcludeerd worden dat met de itemmethode betrouwbaarder gemeten wordt dan met de subcategoriemethode.

Van de subcategorieën is geëist dat zij een positieve bijdrage leveren aan de betrouwbaarheid. De subcategorieën 2, 13 en 14 doen dat niet. Wat betreft de items leveren items 6, 19, 21, 22, en 24 van deel A en items 5, 14, 20 en 26 van deel B een negatieve bijdrage aan de betrouwbaarheid. Voor een deel kan de negatieve bijdrage aan de betrouwbaarheid van de vragen van de twee meetmethoden aan de geringe variantie van de vragen toegeschreven worden. Omdat de bijdrage aan de betrouwbaarheid, net als het discriminerend vermogen, verandert als de samenstelling van een toets verandert, zou deze na het beslissen voor een bepaalde toetssamenstelling opnieuw bepaald moeten worden.

## 6.5 Dimensionaliteit

Bij de constructie van een instrument is het van belang om de dimensionaliteit van het instrument te kennen. Idealiter meten de vragen van een instrument TCO slechts één dimensie: maakbaarheid. Om de dimensionaliteit van de vragen van de subcategorie- en de item-methode te achterhalen is eerst een twee-factormodel gepast, daarna is een homogeniteitsanalyse en een Principale Componenten Analyse (PCA) uitgevoerd.

Voorafgaande aan de beschrijving van de analyses dien enkele opmerkingen over de dataset gemaakt te worden. Als input voor de analyses zijn tetrachorische correlatiematrices gebruikt omdat deze geschikt zijn voor variabelen die op intervalniveau liggen en die kunstmatig gedichotomiseerd zijn. Daarnaast leveren Pearsoncorrelaties bij dichotome gegevens een onderschatting van de correlaties op. Bij tetrachorische correlaties is dat niet het geval. Verder was er sprake van verschillende vormen van de onderlinge afhankelijkheid bij de vragen van de meetmethoden. Bij de itemmethode waren er twee itemparen waarbij het antwoordpatroon van alle leerkrachten gelijk was. Eén van de twee items van de twee itemparen is uit de datamatrix verwijderd. Niet alle afhankelijkheid was daarmee verwijderd. De resterende afhankelijkheid is uit de datamatrix verwijderd door de dichtst bijzijnde grammiaanse correlatiematrix bij een niet-positief definitie matrix te zoeken en deze minimaal te veranderen. Het programma POSDEF is hiervoor gebruikt (Knol & Ten Berge, 1989).

### 6.5.1 Twee-factormodel

Ten behoeven van het onderzoek naar de dimensionaliteit van de subcategorieën en de items is met een *Principal Axes Factoring* (PAF)-programma (Engelen, eigen programma) een twee factormodel op de tetrachorische correlatiematrix gepast. De resultaten van PAF ervan identiek aan de uitvoer van de SAS procedure FACTOR. Eerst is het twee-factormodel op de veertien subcategorieën gepast en daarna op de items van de TCO-vragenlijst.

Een twee-factormodel is gepast op de veertien subcategorieën. De eerste factor verklaart 40% ( $\lambda_1 = 5.657$ ) van de variantie en de tweede factor 16% ( $\lambda_2 = 2.198$ ). De factorladingen van de subcategorieën staan in tabel 6.11. In de eerste en tweede kolom van de tabel staan de categorieën en subcategorieën, in de derde kolom staan de ladingen op de eerste factor en in de vierde kolom staan de ladingen op de tweede factor. De kritieke waarde voor een correlatie coëfficiënt op  $\alpha = .01$  voor een tweezijdige toets is gegeven in tabel 11.1 in Stevens (1992). Hieruit blijkt dat bij een steekproef van 170 ladingen boven .397 significant zijn. Alleen de significante ladingen zijn in de tabel opgenomen.

De categorieën waartoe de subcategorieën behoren geven een aanwijzing voor een interpretatie van de factoren. Bijna alle subcategorieën laden sterk op de eerste factor en mogelijk is het gemeenschappelijke wat ze meten het construct maakbaarheid. Op de tweede factor laden twee verschillende subcategorieën positief en de subcategorieën van categorie

‘rekendictee’ negatief. Mogelijk hebben deze subcategorieën iets inhoudelijks gemeenschappelijk waardoor ze inhoudelijk geïnterpreteerd kunnen worden. ‘Meten en tijd’ laden noch op de eerste, noch op de tweede factor en lijken daarom iets geheel anders te meten.

Tabel 6.11  
Factorladingen van de subcategorieën

| Categorie         | Subcategorie                        | Factor 1 | Factor 2 |
|-------------------|-------------------------------------|----------|----------|
| Tellen en ordenen | Structuur van de telrij             | .647     |          |
|                   | Resultatief en structurerend tellen | .617     |          |
|                   | Vergelijkingen en ordenen           | .610     |          |
| Structureren      | Splitsen                            | .802     |          |
|                   | Samenstellen                        | .697     |          |
|                   | Aanvullen                           | .667     | .502     |
| Bewerkingen       | Optellen                            | .696     | .598     |
|                   | Aftrekken                           | .733     |          |
|                   | Diversen                            | .583     |          |
| Rekendictee       | Optellen                            | .726     | -.496    |
|                   | Aftrekken                           | .720     | -.588    |
|                   | Splitsen                            | .611     | -.635    |
| Meten en tijd     | Meten                               |          |          |
|                   | Tijd                                |          |          |

Drie verschillende twee-factormodellen zijn op de items van de TCO-vragenlijst gepast. Binnen deze analyses zijn steeds strengere eisen aan de items gesteld. Bij de eerste analyse is alleen de afhankelijkheid uit de correlatiematrix verwijderd, dat wil zeggen: identiek beantwoorde items zijn verwijderd en de scores op enkele items zijn minimaal veranderd zodat de negatieve eigenwaarde uit de matrix is gehaald. Van de vijftig overgebleven items, verklaarde de eerste factor 30% variantie [ $\lambda_1=15.127$ ] en de tweede 20% [ $\lambda_2=9.762$ ].

Bij de tweede analyse zijn items met extreme p-waarden — dat wil zeggen,  $p>.90$  — verwijderd om de 28 voldoende discriminerende items in de analyse te hebben. Bij deze items verklaarde de eerste factor 35% [ $\lambda_1=9.938$ ] van de variantie en de tweede factor 19% [ $\lambda_2=5.316$ ].

In de derde analyse ten slotte zijn alleen items met een p-waarde lager dan .80 en met een item-restcorrelatie groter of gelijk aan .30 opgenomen. Van deze 15 items verklaarde de eerste factor 44% [ $\lambda_1=6.628$ ] en de tweede 14% [ $\lambda_2=2.029$ ]. De resultaten van deze analyse staan in tabel 6.12. In de eerste twee kolommen staan het toetsdeel en het item. In de derde en vierde kolom staan de significante ladingen op de eerste en tweede factor. De items zijn geordend op basis van de grootte van de eerste factor.

Net als bij de subcategorieën blijkt het bij de items ingewikkeld te zijn de betekenis van de factoren te achterhalen. Elf items laden significant op de eerste factor, vier laden negatief op de tweede factor en drie items meten iets anders. Voor deze factoren een inhoudelijke verklaring geven, blijkt niet eenvoudig. De elf items die significant op de eerste factor laden behoren tot verschillende subcategorieën en de vier items die negatief op de tweede factor laden horen inhoudelijk bij de subcategorieën ‘splitsen’ en ‘meten’. De eerste factor lijkt daarom de maakbaarheid van de items representeren en de tweede lijkt een meer inhoudelijke betekenis te moeten krijgen.

In de paragrafen 3.3 en 6.1.1 is het gevaar uitgesproken dat leerkrachten zich eerder door de moeilijkheid van de items laten leiden dan door de maakbaarheid. De proportie correcte ant-

woorden van de leerlingen staat in kolom vijf van tabel 6.12. Hieruit blijkt dat de eerste, noch de tweede factor als 'moeilijkheidsgraad' geïnterpreteerd kan worden.

Tabel 6.12  
Factorladingen van items met  $p < .80$  en  $R_{it} \geq .30$

| Deel | Item | Factor 1 | Factor 2 | p-waarde |
|------|------|----------|----------|----------|
| B    | 22   | .874     |          | .62      |
| B    | 16   | .855     |          | .73      |
| A    | 11   | .832     |          | .77      |
| B    | 23   | .827     |          | .62      |
| B    | 19   | .824     |          | .90      |
| A    | 16   | .821     |          | .65      |
| A    | 15   | .741     |          | .78      |
| B    | 17   | .665     |          | .82      |
| B    | 09   | .595     | -.489    | .72      |
| B    | 29   | .594     | -.624    | .43      |
| B    | 10   | .528     | -.486    | .78      |
| B    | 28   |          |          | .72      |
| A    | 06   |          |          | .54      |
| A    | 20   |          |          | .66      |
| A    | 24   |          | -.672    | .75      |

Samenvattend blijken de uitgevoerde factoranalyses enig bewijs van een maakbaarheidsdimensie op te leveren. Bij de subcategorieën blijkt er een dimensie te bestaan die 40% van de variantie verklaart en die mogelijk als 'maakbaarheid' geïnterpreteerd kan worden. Bij de items waarvan is geëist dat deze een p-waarde lager dan .80 hebben en een  $R_{it}$  groter of gelijk aan .30 is eveneens een dimensie gevonden die 44% van de variantie verklaart en mogelijk als maakbaarheid geïnterpreteerd kan worden.

### 6.5.2 Homogeniteitsanalyse

Met een homogeniteitsanalyse zoekt men naar afwijkingen van het algemene patroon in de data door de eigenwaarde van de te passen dimensies te optimaliseren. Het aantrekkelijke van het uitvoeren van deze analyse met SPSS is dat het één figuur op kan leveren waar de subcategorieën en de items op dezelfde dimensies staan afgebeeld.

De volgende stappen zijn ondernomen. Eerst is afzonderlijk voor de subcategorieën en de items onderzocht hoeveel variantie door twee dimensies verklaard worden als verschillende eisen aan de data gesteld worden. Daarna zijn de items in een figuur geplaatst, zodat aan de hand van de clustering tussen de items de betekenis van de dimensies achterhaald kan worden.

Wanneer alle 14 subcategorieën in de analyse betrokken worden kunnen de dimensies de antwoorden van de leerkrachten nauwelijks van elkaar onderscheiden. De eerste dimensie verklaart 21% en de tweede 17% van de variantie. Wanneer de 7 subcategorieën met een p-waarde lager dan .90 opgenomen worden, verklaren de twee dimensies aanzienlijk meer variantie. De eerste verklaart 33% en de tweede 20%. Een homogeniteitsanalyse waar zowel aan de p-waarde als aan de  $R_{it}$ -waarde eisen zijn gesteld ( $.29 < p < .79$  en  $R_{it} > .30$ ) is niet uitgevoerd omdat er maar twee subcategorieën zijn die aan deze eisen voldoen.

Wanneer alle 53 items in de analyse betrokken worden, levert dat een soortgelijk resultaat als bij de subcategorieën op: de antwoorden van de leerkrachten onderscheiden zich nauwelijks op beide dimensies. De eerste dimensie verklaart 15% van de variantie en de tweede 12%. Wanneer de 28 items met een p-waarde lager dan .90 in de analyse betrokken worden, levert dat een beter beeld op: de eerste dimensie verklaart 35% van de variantie en de tweede 14%. Het stellen van eisen aan zowel de p-waarde als de  $R_{it}$ -waarde levert geen aanzienlijke winst in passing op.

Besloten is de vragen met een p-waarde lager dan .90 te gebruiken voor een spreidingsdiagram. Het stellen van strengere eisen aan de vragen bleek zowel bij de subcategorie als bij de itemmethode niet zinvol. Bij de subcategoriemethode was dat omdat het slechts twee vragen betreft en bij de itemmethode leverde het stellen van strengere eisen geen significante winst in passing op.

In Bijlage 3 (figuur A) staan de items en de subcategorieën met een p-waarde onder de .90 geplott in een spreidingsdiagram. De eerste dimensie zou een maakbaarheidsdimensie kunnen zijn, omdat alle maakbare subcategorieën en items laag op deze dimensie laden en alle niet maakbare subcategorieën en items hoog. De betekenis van tweede dimensie valt niet direct uit de figuur op te maken. Voor de betekenis moet mogelijk wederom een inhoudelijke verklaring gezocht worden omdat de items en de subcategorieën van onder andere structureren laag en van meten en tijd hoog op deze dimensie laden. Misschien moet deze dimensie als een 'kennis versus inzicht'-dimensie geïnterpreteerd worden. Een andere mogelijkheid is dat de items en subcategorieën die laag laden op de tweede dimensie door leerkrachten moeilijker bevonden worden dan items en subcategorieën die er hoog op laden. Om dit te weten te komen moet naar de resultaten van leerlingen op de items van toets E3 gekeken worden.

Figuur B van Bijlage 3 is gelijk aan figuur A, behalve dat de items die bij een bepaalde subcategorie behorend zijn omcirkeld. In de figuur is te zien dat de items die tot een bepaalde subcategorie behoren, in het algemeen ook daadwerkelijk in de buurt van de subcategorie clusteren. Dit lijkt erop te wijzen dat de subcategorieën en de bij de subcategorie behorende items op dezelfde manier — dat wil zeggen op dezelfde dimensies — beoordeeld zijn.

In figuur C van Bijlage 3 staan de leerkrachten die verschillende lesmethoden gebruiken, afgezet tegen dezelfde dimensies als de subcategorieën en de items in figuren A en B. Er valt in de figuur een aantal dingen op. Allereerst is er een vergelijkbare clustering te zien als bij de maakbare items. Dat wil zeggen dat de maakbare antwoorden negatief laden op de eerste dimensie en de niet maakbare antwoorden positief. Verder valt op dat de leerkrachten die bepaalde lesmethoden gebruiken zich zowel op de eerste als op de tweede dimensie onderscheiden. Bijvoorbeeld: de antwoorden van leerkrachten die beide uitgaven van Wereld in Getallen gebruiken, laden positief op de eerste dimensie, terwijl de antwoorden bij de overige lesmethoden er negatief op laden. Op de tweede dimensie laden beide uitgaven van Wereld in getallen negatief, terwijl de overige lesmethoden er positief op laden. Beide uitgaven van Wereld in getallen nemen dus op beide dimensies een uitzonderingspositie in.

Samenvattend wijst de homogeniteitsanalyse op een maakbaarheidsdimensie die zowel een deel van de variantie binnen de items als binnen de subcategorieën verklaart. De clustering in de data wijst er mogelijk op dat bepaalde onderwerpen maakbaarder zijn dan andere onderwerpen. De betekenis van de tweede dimensie is niet duidelijk: mogelijk is het een 'kennis versus inzicht'-dimensie, of een moeilijkheidsdimensie. Het vergelijken van de

moeilijkheidsgraad van de items voor de leerlingen (de p-waarde) met de ladingen op de eerste factor kan over de moeilijkheidsdimensie uitsluitsel geven.

### 6.5.3 Principale Componenten Analyse

Na de voorafgaande analyses is een principale componentenanalyse (PCA) met SPSS uitgevoerd. Met PAF- en de homogeniteitsanalyse zijn steeds twee factoren gepast, met PCA is exploratief nagegaan hoeveel dimensies — of componenten — de meeste variantie verklaren. Achtereenvolgens wordt de dimensionaliteit van de subcategorieën en de items besproken.

#### Dimensionaliteit van de subcategorieën

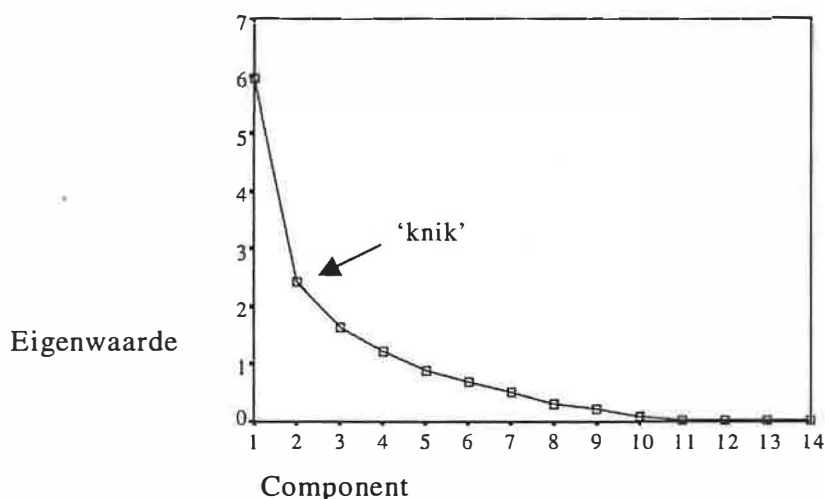
De beslissing over het aantal componenten dat behouden mag worden, is afhankelijk van het gebruikte criterium. Verschillende uitkomsten volgens verschillende criteria worden hier besproken.

Het Kaiser criterium behoudt componenten met een eigenwaarde groter dan één. Uit de studie van Cattell en Jaspers (1967; gerefereerd door Stevens, 1992) blijkt het Kaiser criterium accuraat als het aantal variabelen klein is en de communaliteiten hoog zijn. Zonder in details te treden, is dit het geval bij de subcategorieën en mag het criterium dus gebruikt worden. Als men vervolgens het Kaiser criterium gebruikt, blijven er vier componenten over die samen 80% van de variantie verklaren (zie tabel 6.13).

Tabel 6.13  
Uitkomst van de PCA: de eigenwaarden en het percentage verklaarde variantie en het cumulatieve percentage verklaarde variantie

| Component | Eigenwaarde | % verklaarde variantie | Cum. % verkl. variantie |
|-----------|-------------|------------------------|-------------------------|
| 1         | 5.959       | 42.6                   | 42.6                    |
| 2         | 2.430       | 17.4                   | 59.9                    |
| 3         | 1.631       | 11.7                   | 71.6                    |
| 4         | 1.206       | 8.6                    | 80.2                    |
| 5         | .881        | 6.3                    | 86.5                    |
| 6         | .692        | 5.0                    | 91.4                    |
| 7         | .506        | 3.6                    | 95.0                    |
| 8         | .311        | 2.2                    | 97.3                    |
| 9         | .210        | 1.5                    | 98.8                    |
| 10        | .093        | .7                     | 99.4                    |
| 11        | .021        | .1                     | 99.6                    |
| 12        | .020        | .1                     | 99.7                    |
| 13        | .020        | .1                     | 99.9                    |
| 14        | .019        | .1                     | 100                     |

Bij het tweede criterium, de screeplot, worden eigenwaarden van de componenten afgezet tegen hun ordinale nummers. Een knik in de plot geeft aan dat het sterk dalen van de eigenwaarde gestopt is. Uit figuur 2 blijkt de knik veel eerder dan de vierde component te komen. Al na de eerste component is een sterke daling in de winst van de eigenwaarde te zien (zie figuur 2). Uit de screeplot blijkt slechts één dimensie van belang en deze verklaart 43% van de variantie (zie tabel 6.13).



Figuur 6.1  
Screeplot van de PCA over de 14 subcategorieën

Welke van deze oplossingen men verkiest is een kwestie van smaak. Besloten is daarom te kijken of er inhoudelijke redenen zijn om voor één of vier componenten te kiezen.

Ten behoeve van de interpretatie is de componentenmatrix varimax-geroteerd. De gerooteerde matrix staat in tabel 6.14. In de eerste twee kolommen van de tabel staan de categorieën en subcategorieën en in kolom drie tot en met zes staan de factorladingen van de subcategorieën op de eerste tot-en-met vierde component. Alleen de significante ladingen ( $\geq .397$ ) zijn in de tabel opgenomen.

Tabel 6.14  
De varimax-geroteerde componentenmatrix van de subcategorieën

| Categorie        | Subcategorie                       | 1    | 2    | 3    | 4    |
|------------------|------------------------------------|------|------|------|------|
| Tellen & ordenen | Structuur van de telrij            | .531 |      |      | .564 |
|                  | Resultatief & structurerend tellen | .481 |      | .677 |      |
|                  | Vergelijkingen & ordenen           | .479 | .481 |      |      |
| Structuren       | Splitsen                           |      | .854 |      |      |
|                  | Samenstellen                       |      | .692 |      |      |
|                  | Aanvullen                          |      | .884 |      |      |
| Bewerkingen      | Optellen                           |      | .408 | .859 |      |
|                  | Aftrekken                          |      |      | .823 |      |
|                  | Diversen                           |      | .468 | .520 | .402 |
| Rekendictee      | Optellen                           | .937 |      |      |      |
|                  | Aftrekken                          | .911 |      |      |      |
|                  | Splitsen                           | .804 |      |      |      |
| Meten & tijd     | Meten                              |      |      |      | .730 |
|                  | Tijd                               |      |      |      | .911 |

Alleen de significante varimax-geroteerde factorladingen worden voor interpretatie van de componenten gebruikt. Het blijkt dat de eerste component — de enige component volgens de screeplot — grotendeels door de subcategorieën van de categorieën Tellen & ordenen en Rekendictee verklaard wordt. De tweede component wordt grotendeels door de subcate-

gorieën van Structureren verklaard. De derde component wordt met name door de subcategorieën van Bewerkingen verklaard en de vierde ten slotte met name door Meten & tijd.

De componenten blijken bijzonder goed aan te sluiten bij de vijf categorieën waar de inhoud van toets E3 in onderverdeeld is. Dus er is wel wat voor de vier componenten te zeggen: ze verklaren gezamenlijk 80% van de variantie, hebben een eigenwaarde boven de één en zijn goed te interpreteren. De vier componenten binnen deze PCA lijken dus niet de maakbaarheid van toets E3 (TCO) te weerspiegelen, maar eerder de categorieën waar de inhoud van de toets in onderverdeeld is.

### **Dimensionaliteit van de items**

Op dezelfde wijze als bij de subcategorieën is over de items een exploratieve principale componenten analyse uitgevoerd. Anders is dat alleen items met een p-waarde lager dan .90 zijn opgenomen. Met het Kaiser criterium worden zeven componenten ( $\lambda_1 = 10.120$ ,  $\lambda_2 = 5.748$ ,  $\lambda_3 = 2.460$ ,  $\lambda_4 = 1.811$ ,  $\lambda_5 = 1.418$ ,  $\lambda_6 = 1.125$ ,  $\lambda_7 = 1.002$ ) geaccepteerd, die samen 85% van de variantie verklaren. Zeven componenten zijn er te veel om nog goed te kunnen interpreteren. Met de scree test blijven er twee componenten over die samen 57% van de variantie verklaren.

Voor de interpretatie van de principale componenten is uitgegaan van significante varimax geroteerde ladingen van twee componenten (zie tabel 6.15). Een groot deel van de items laadt significant op de eerste component. Dit wijst erop dat de eerste component als een maakbaarheidsdimensie geïnterpreteerd kan worden. De betekenis van de tweede component is niet meteen duidelijk. Enkele items laden negatief op de tweede component en enkele positief. Om te kunnen achterhalen of de tweede component een moeilijkheidsdimensie representeert, is in de vijfde kolom de moeilijkheid van de items (in p-waarde) opgenomen. Uit de tabel kan geen aanwijzing voor een moeilijkheidsdimensie gevonden worden. Mogelijk biedt een analyse van de inhoud van de items meer inzicht in de betekenis van de tweede principale component.

### **Conclusies met betrekking tot de dimensionaliteit**

Bij de constructie van een instrument is het van belang te bepalen of één of meerdere constructen gemeten worden, omdat men dan een indruk heeft van de begripsvaliditeit van een instrument. Hiertoe zijn enkele analyses gedaan. Hoewel bij ieder van deze analyses andere data gebruikt zijn, waardoor de resultaten moeilijk te vergelijken zijn, zijn er wel enkele algemene conclusies te trekken.

Een eerste conclusie is dat hoewel vaak meerdere dimensies een groot deel van de variantie verklaren, steeds de eerste dimensie de belangrijkste dimensie blijkt te zijn. Zowel de PAF, de homogeniteitsanalyses, als de PCA wijzen erop dat de eerste dimensie als een maakbaarheidsdimensie geïnterpreteerd kan worden. Toch zijn er ook resultaten die hiermee in strijd zijn. Bij de PCA over de subcategorieën, bijvoorbeeld, lijken de dimensies de subcategorieën in (hoofd-)categorieën te ordenen. Dit wijst erop dat de betekenis van de eerste component inhoudelijk geduid moet worden. De diverse betekenissen van de eerste dimensie — maakbaarheid, moeilijkheid, of iets inhoudelijks, hoeven overigens niet met elkaar in strijd te zijn. Bijvoorbeeld, als bepaalde onderwerpen nog te moeilijk zijn voor een klas, kan de leerkracht het behandelen van het onderwerp mogelijk even uitstellen.

Een tweede conclusie is dat met deze analyses enkele afwijkingen van het algemene patroon, dat wil zeggen afwijkende subcategorieën, items, lesmethoden en leerkrachten,

zichtbaar zijn gemaakt. Het meest duidelijk zijn deze afwijkingen zichtbaar in de drie figuren met resultaten van de homogeniteitsanalyse.

Tabel 6.15  
De varimax-geroteerde component matrix van de TCO-items

| Deel | Item | Component |       | Moeilijkheid |
|------|------|-----------|-------|--------------|
|      |      | 1         | 2     |              |
| B    | 22   | .913      |       | .63          |
| B    | 23   | .881      |       | .62          |
| A    | 16   | .862      |       | .65          |
| A    | 11   | .845      |       | .77          |
| A    | 08   | .841      |       | .85          |
| B    | 16   | .841      |       | .73          |
| B    | 15   | .828      | -.397 | .79          |
| A    | 15   | .822      |       | .78          |
| B    | 19   | .809      |       | .90          |
| B    | 21   | .774      | -.455 | .81          |
| A    | 13   | .766      | -.487 | .81          |
| B    | 27   | .672      |       | .80          |
| A    | 22   | .669      | .516  | .82          |
| A    | 23   | .669      |       | .62          |
| B    | 17   | .652      | .421  | .82          |
| A    | 24   | .633      | -.423 | .75          |
| B    | 26   | .560      | .503  | .82          |
| B    | 29   | .533      | .541  | .43          |
| B    | 09   | .519      | .518  | .72          |
| A    | 12   | .451      |       | .88          |
| A    | 17   | .420      | .585  | .82          |
| B    | 28   | .417      | .554  | .72          |
| B    | 25   | .416      | .476  | .74          |
| B    | 10   | .407      | .526  | .78          |
| B    | 08   | .399      | .609  | .75          |
| A    | 18   |           | .554  | .83          |
| A    | 20   |           | .450  | .66          |
| A    | 06   |           |       | .54          |

## 6.6 Schaalbaarheid van de drie meetmethoden

In paragraaf 6.5 is met verschillende analyses de dimensionaliteit van de vragen van de TCO-vragenlijst onderzocht. Bij het schalen van de vragen van de subcategorie- en item-methode wordt een latente trekmodellen ook onderzocht of de vragen alle hetzelfde construct — maakbaarheid — meten. Met diverse itemreponsmodellen is geprobeerd de vragen van de twee meetmethoden op één schaal te plaatsen. De modellen betreffen het Mokkenmodel, het Raschmodel en het OPLM. De eigenschappen van de modellen en hun passing bij de methoden worden in de paragrafen 6.6.1 tot en met 6.6.3 beschreven.

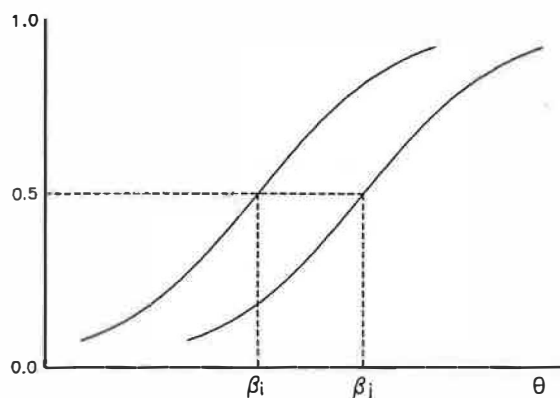
### 6.6.1 Het Mokkenmodel

Het Mokkenmodel, is een niet-parametrisch, probabilistisch IRT model. In dit model worden een drietal aannamen gemaakt: (1) de itemrespons wordt bepaald door een latente trek, (2) de vragen van de test zijn eendimensionaal, en (3) de regressiefunctie is een monotoon stijgende functie van de latente trek. Voor een uitgebreidere omschrijving van het Mokkenmodel, zie Van den Brink & Mellenbergh (1998), of Molenaar, Debets, Sijtsma & Hemker (1994).

Het Mokkenmodel over zowel de 14 subcategorieën als de 53 items is niet houdbaar. Aan Loevinger's H is te zien dat de schaal zwak is ( $H = .27$ ). Hoewel de modelpassing slecht is, levert de analyse wel enige informatie over de clustering van de subcategorieën en de items. De subcategorieën die iets gemeenschappelijks lijken te meten zijn: subcategorieën 4, 5, 10, 11 en 12; subcategorieën 6, 7, 8, en 9; en subcategorieën 13 en 14. Deze resultaten lijken sterk op de resultaten van de principale componentenanalyse. De clustering bij de items is niet informatief en wordt hier daarom verder niet besproken.

### 6.6.2 Het Raschmodel

Het tweede model dat gepast is, is het Raschmodel. De eerste twee aannamen van het Raschmodel zijn gelijk aan de aannamen van het Mokkenmodel. Het Raschmodel verschilt van het Mokkenmodel in de derde aanname: een parametrische regressiefunctie wordt aangenomen. In deze functie wordt één itemparameter,  $\beta$ , gedefinieerd. Bèta wordt de moeilijkheidsparameter van het item genoemd. Voor een uitgebreide omschrijving, zie Eggen & Sanders (1993), Van den Brink & Mellenbergh (1998) en Verhelst (1992).



Figuur 6.2  
Twee itemresponsfuncties in het Raschmodel  
[Bron: Eggen & Sanders (1993), p. 90]

In figuur 6.2 staan twee itemresponsfuncties afgebeeld. Figuur 6.2 geeft aan dat item j moeilijker is dan item i, omdat er meer vaardigheid ( $\theta$ ) nodig is om item j goed te maken dan om item i goed te maken. In dit onderzoek krijgt de moeilijkheidsparameter echter een andere betekenis, omdat het in dit onderzoek niet gaat om een vaardigheid die gemeten wordt, maar om de maakbaarheid van een item. Bèta is in dit onderzoek de maakbaarheidsparameter. Item i is in figuur 6.2 maakbaarder dan item j, omdat er meer maakbaarheid nodig is om item j te kunnen maken dan om item i te kunnen maken.

In de paragrafen 6.6.2.1 en 6.6.2.2 worden respectievelijk de passing van het Raschmodel op de veertien subcategorieën en de passing van het Raschmodel op de 53 items, besproken.

### Passing van het Raschmodel op de subcategorieën

Voor de algehele passing van het model is de  $R_{1c}$ -statistiek gebruikt. Deze statistiek is een *overall*-toets om een reeks modelschendingen op itemniveau te kunnen ontdekken ( $s_i$ -toetsen). Uit de  $R_{1c}$  blijkt dat de passing van het Raschmodel op de subcategorieën niet erg goed is ( $R_{1c} = 41.790$ ,  $v = 26$ ,  $p = .256$ ). Subcategorieën die een grote bijdrage aan de *misfit* leveren zijn 4, 13 en 14. Zonder deze subcategorieën — even voorbijgaand aan de vraag of dit inhoudelijk te rechtvaardigen is — wordt de passing nog slechter ( $R_{1c} = 14.527$ ,  $v = 10$ ,  $p = .150$ ).

### Passing van het Raschmodel op de items

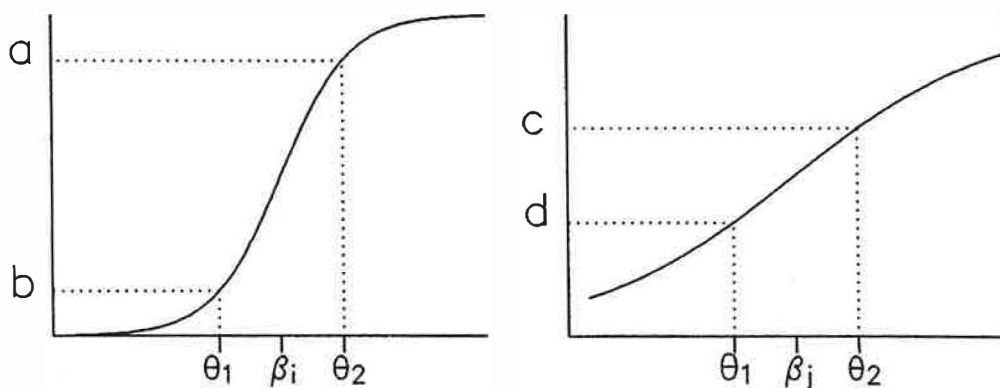
De passing van het Raschmodel op de 53 items is eveneens niet erg goed ( $R_{1c} = 209.931$ ,  $v = 156$ ,  $p = .002$ ). Een erg grote bijdrage aan de misfit komt van de items 5, 8, 22, 23 en 24 van deel A. Het verwijderen van deze items levert wederom geen betere passing op ( $R_{1c} = 215.362$ ,  $v = 144$ ,  $p = .000$ ). Daarnaast is onderzocht of het verwijderen van items met extreme  $p$ -waarden ( $p > .90$ ), dat wil zeggen de items met weinig itemvariantie, een betere modelpassing oplevert. Ook dit bleek niet het geval ( $R_{1c} = 178.66$ ,  $v = 81$ ,  $p = .000$ ).

Samenvattend blijkt bij zowel de subcategorieën als de items het Raschmodel geen bevredigende passing op te leveren.

### 6.6.3 Het OPLM

Het OPLM-, of het Verhelst- en Eggenmodel, heeft naast een moeilijkheidsparameter,  $\beta_i$  een grootte  $a_i$ , de discriminatie-index. Hoe hoger de waarde van de discriminatie-index van een item, hoe steiler de regressiefunctie is, en hoe beter het item discrimineert tussen personen met latente vaardigheden in de buurt van de itemmoeilijkheid. Voor een uitgebreide omschrijving van het OPLM, zie Eggen & Sanders (1993) of Verhelst (1992).

Zoals reeds bij het Raschmodel ter sprake kwam, krijgt in dit onderzoek bèta de betekenis van een maakbaarheidsparemeter (zie paragraaf 6.6.2). Voor de discriminatie-index geldt dat hoe hoger de waarde van deze index, hoe beter het item discrimineert tussen personen met maakbaarheidsscores in de buurt van  $\beta$  (zie figuur 6.3).



Figuur 6.3  
Twee itemresponsfuncties in het OPLM  
[Bron: Eggen & Sanders (1993), p. 122]

## Passing van het OPLM op de subcategorieën

De passing van het OPLM op de subcategorieën is goed ( $R_{1c} = 25.819$ ,  $v = 24$ ,  $p = .362$ ). In bijlage 5 staat een overzicht van de parameterschattingen. De discriminatie-indices zijn handmatig aangepast om passing van het OPLM te verbeteren. Hiertoe is gekeken naar de individuele passing van de vragen (chi-kwadraat toetsen), naar de M-toetsen (deze toetsen geven aan of er een over- dan wel een onderschatting van  $a_i$  is). Voor een uitvoerige beschrijving van de toetsen, zie Eggen & Sanders (1993) of Verhelst (1992). Figuren waarin geobserveerd en voorspeld zijn weergegeven, zijn eveneens gebruikt.

De subcategorieën van de TCO-vragenlijst zijn met het OPLM eenvoudig op één schaal te plaatsen. Ten behoeve van de interpretatie is gekozen de schaal van nul tot en met 100 te laten lopen. In bijlage 5 staat een overzicht van de toetsscores, de geschatte maakbaarheidsscores en de bijbehorende schaalesscores.

## Passing van het OPLM op de items

Door de geringe variantie tussen de itemantwoorden was het vinden van een passend OPLM voor de items van de vragenlijst ingewikkelder dan voor de subcategorieën. Om de passing van het model te verbeteren zijn er verschillende mogelijkheden. Naast de mogelijkheid om de passing te vergroten door de discriminatie-indices van de items aan te passen of door de gemiddelde discriminatie-index te verhogen, is er de mogelijkheid tot het verwijderen van items. Men zou items kunnen verwijderen die slecht voldoen aan het OPLM, zodat alleen items overblijven die voldoende passen. Een probleem daarbij is het grote aantal items met een hele hoge p-waarde (weinig itemvariantie). Wanneer het model dan gepast wordt, dan zijn het juist de items met de geringe itemvariantie die goed passen, omdat er daar zo veel van zijn. De items met voldoende itemvariantie, die men in het model wil hebben, passen niet omdat ze niet op de bulk lijken. Het is niet de bedoeling dat deze items buiten het model vallen en alleen 'slechte' items in het model opgenomen worden. Deze mogelijkheid valt daarom af.

Een tweede mogelijkheid is dat men zich niet laat leiden door de passing van het item maar door een vooraf opgesteld criterium, bijvoorbeeld alleen items met een p-waarde lager dan .90 opnemen. Een nadeel hiervan is dat de onderzoeker alleen informatie heeft over hoe de leerkracht op een selectie van de items scoort. Van de overige items moet worden aangenomen dat deze als maakbaar beoordeeld zijn.

Een derde mogelijkheid om de modelpassing te verbeteren is door leerkrachten die zich niet volgens het model gedragen (de uitbijters) te verwijderen. Er waren drie uitbijters. Deze drie zijn ook in Figuur C van de homogeniteitsanalyse terug te vinden (zie Bijlage 3). Het bleek praktisch niet uitvoerbaar om deze leerkrachten buiten de analyse te houden zonder verdere items uit de analyse te verwijderen, kan het model niet gepast worden. Zij waren namelijk als enigen die twee items als niet maakbaar beoordeelden. De uitbijters kunnen dus niet verwijderd worden zonder meteen ook items te verwijderen. Omdat men voorzichtig moet zijn met het verwijderen van uitbijters is besloten de uitbijters in het model te behouden.

De tweede mogelijkheid om de passing van het model te verbeteren, het verwijderen van items op basis van een vooropgesteld criterium, lijkt daarom de beste strategie. Het meest voor de hand liggende criterium is het weglaten van items met een extreme p-waarde. Het OPLM is op twee verschillende data gepast. Eerst zijn alle items in de analyse betrokken en daarna alleen items met een p-waarde lager dan .95.

Wanneer alle items in de analyse betrokken worden, is de passing goed ( $R_{1c} = 178.889$ ,  $v = 156$ ,  $p = .100$ ). De items die afwijken van het model zijn item **19**, **22**, **23**, en **24** van deel A en item **26** van deel B. Het geometrisch gemiddelde van de discriminatie-indices is bij dit model 3.009. Het verlagen van de gemiddelde discriminatie-index verslechtert de passing. Wanneer de 38 items met een p-waarde kleiner dan .95 in de analyse betrokken worden is de passing van het model slechter dan wanneer alle items in de analyse betrokken worden ( $R_{1c} = 147.084$ ,  $v = 111$ ,  $p = .011$ ). Aanpassing van de discriminatie-indices levert geen verbetering van de passing. Items **19**, **21**, **22** en **23** van deel A wijken af van het model. Pas wanneer deze items verwijderd wordt, is de passing veel beter ( $R_{1c} = 99.084$ ,  $v = 95$ ,  $p = .367$ ).

Zowel alle items als een selectie van items (items met een p-waarde kleiner dan .95), zou gebruikt kunnen worden voor het meetinstrument. Immers, bij beide mogelijkheden past het OPLM goed. Uit de analyses is verder gebleken dat het steeds (bijna) dezelfde items zijn die niet, of net niet, in het OPLM passen. Dus wanneer een deel van de items opgenomen wordt, lijkt dezelfde latente trek — maakbaarheid — gemeten te worden als wanneer ze allemaal opgenomen worden.

De schattingen van de itemparameters van het OPLM met alle 53 items zijn te vinden in bijlage 6. In deze bijlage zijn ook de persoonsparameters en de schaalscores te vinden. De persoonsparameter zijn met het programma Theta met WML geschat (Verhelst & Engelen, in press).

De keuze tussen een maakbaarheidsschaal met alle items of een selectie van de items is van toetstechnische aard. Een schaal waarin alle items opgenomen zijn, heeft het voordeel dat geen extra aannamen gemaakt hoeven te worden over de maakbaarheid van opgaven die buiten de selectie vallen. Een schaal die gebaseerd is op een selectie van vragen heeft het voordeel dat alleen de vragen met voldoende itemvariantie opgenomen zijn en er dus een potentieel 'stabielere' schaal gemaakt kan worden. Dat wil zeggen dat de schaal niet opvallend verandert als een of twee items uit de toets verwijderd worden. Een ander voordeel is dat wanneer minder vragen voorgelegd worden, minder vragen hoeven geanalyseerd te worden.

## **6.7 Validering van TCO met toetsscores van leerlingen als criterium**

De validiteit van een meetinstrument gaat in op de vraag of het toetsresultaat geïnterpreteerd mag worden in termen van het begrip dat het instrument bedoelt te meten (Van der Brink & Mellenbergh, 1998, p.271). Om een indruk te krijgen van het verzamelbegrip validiteit van de subcategorie-, item- en lesmethodemethode is een aantal stappen ondernomen. Enkele analyses naar de validiteit van de meetmethoden zijn al beschreven. De begripsvaliditeit van de drie meetmethoden, is hiervoor onderzocht met verschillende dimensionaliteits- en schalingstechnieken. Ook zijn eerder reeds enkele analyses beschreven die betrekking hadden op de criteriumvaliditeit van de meetmethoden. Om de criteriumvaliditeit van de vragen van de drie meetmethoden vast te stellen is in eerste instantie de overeenstemming in maakbaarheidsscores tussen de vragen van de methoden onderzocht. Dit is gedaan door de scores op de subcategorieën en de items te correleren (zie 6.2.3.3) en door de p-waarde van de vragen van de meetmethoden te vergelijken (zie 6.2.3.2 en 6.2.3.3). Een derde manier om de criteriumvaliditeit van de methoden om TCO te meten te bepalen, is door het vaststellen van een verband tussen TCO-oordelen over toets E3 (TCO-scores) en de prestaties van leerlingen op dezelfde toets (LVS-scores). De prestaties van de leerlingen op toets E3 zouden beter moeten zijn als de leerkracht de toets maakbaar beoordeelt dan wanneer deze de toets als niet maakbaar beoordeelt. Het valideren van de

drie methoden om TCO te meten door middel van de scores van leerlingen wordt in deze paragraaf beschreven.

Anders dan in de voorafgaande paragrafen zullen eerst de in de validering van de items ondernomen stappen aan bod komen en daarna volgen pas de subcategorieën en de lesmethoden, omdat bij de validering van deze twee meetmethoden van de itemmethode gebruik wordt gemaakt. Bij de itemmethode wordt met een t-toets het verschil in gemiddelde toetsscore getoetst tussen leerlingen van leerkrachten die het item als maakbaar en leerlingen van leerkrachten die het item niet als maakbaar beoordelen. Bij de subcategorieën en de lesmethoden is geen t-toets uitgevoerd omdat geen antwoorden van leerlingen op de subcategorieën of de lesmethoden verzameld kunnen worden. Bij deze meetmethoden is daarom een overzicht gemaakt van de maakbare vragen en de proportie correcte items. Een laatste stap in het vaststellen van de criteriumvaliditeit van de meetmethoden, wordt in paragraaf 6.8 beschreven. In deze paragraaf wordt met de itemresponsstheorie onderzocht of leerlingen een voor- of een nadeel van een item ondervinden.

De onderstaande analyses hebben betrekking op de antwoorden van een steekproef van 3216 leerlingen. De antwoorden van 49 leerlingen worden niet gebruikt, omdat twee leerkrachten minder dan de helft van toets E3 maakbaar vonden.

### **6.7.1 Validering van de items**

Ter validering van de items van de TCO-vragenlijst is een aantal stappen ondernomen. Eerst is de proportie — de p-waarde — leerkrachten die een vraag als maakbaar beoordeelde, afgezet tegen de proportie leerlingen dat het item correct beantwoordde. De verwachting is dat de mate van maakbaarheid van een item en het correct beantwoorden van een item door leerlingen geen één-op-één relatie heeft. Het mag deze één-op-één relatie ook niet hebben, omdat een instrument dat de maakbaarheid van een item moet meten, niet de moeilijkheid van een item mag meten. Het zijn immers twee verschillende constructen.

Toch zouden leerlingen van leerkrachten die een vraag maakbaar vinden het gemiddeld beter moeten doen dan leerlingen van leerkrachten die een vraag niet maakbaar vinden. Een t-toets is uitgevoerd om na te gaan of de gemiddelde prestaties van deze twee groepen leerlingen verschillen.

Hierna worden de toetsscores van de leerkrachten op de vragen van de TCO-vragenlijst 'TCO-scores' genoemd en de toetsscores van de leerlingen op toets E3 'LVS-scores'.

In de eerste twee kolommen van tabel 6.15 staan het toetsdeel en het itemnummer. In de derde en vierde kolom staan de p-waarde en de spreiding van de maakbaarheidsoordelen van het item — aangeduid met TCO — en in de vijfde en zesde kolom staan de p-waarde en de standaardafwijking van de p-waarde van de leerlingen die het item correct beantwoord hebben — aangeduid met LVS.

In de tabel is te zien dat een hoge TCO-score gemiddeld gepaard gaat met een hoge score op toets E3 en een lage TCO-score met een lage score op toets E3. Zoals verwacht is deze relatie niet één-op-één.

Een meer informatieve manier om te achterhalen of het construct 'maakbaarheid' goed is gemeten, is door te onderzoeken of leerlingen van leerkrachten die een opgave als maakbaar beoordelen gemiddeld beter presteren dan leerlingen van leerkrachten die een opgave niet als maakbaar beoordelen. In tabel 6.16 staan de resultaten van een t-toets voor het verschil in de gemiddelde proportie correcte antwoorden van deze verschillende groepen leerlingen.

In de eerste twee kolommen van tabel 6.16 staan zoals gewoonlijk het toetsonderdeel en de opgavenummers van de items. Onder de kop 'Niet maakbaar' staat in de derde kolom voor elk item het aantal leerlingen van leerkrachten die het item als *niet maakbaar* beoordeelden en in de vierde kolom staat de proportie correcte antwoorden van deze leerlingen. Onder de kop 'Maakbaar' staat in kolom vijf het aantal leerlingen van leerkrachten die het item als *maakbaar* beoordeelden en in kolom zes de proportie correcte antwoorden van deze leerlingen. De kolommen drie en vijf tellen niet altijd op tot 3216, omdat zowel leerkrachten als leerlingen wel eens items overslaan. In de kolommen zeven tot en met tien staan voor elke opgave de t-statistiek, het aantal vrijheidsgraden, de significantiewaarde en het verschil in gemiddelde p-waarde.

Een bekend verschijnsel in de statistiek is dat het vinden van een significant verschil groter wordt naarmate het aantal uitgevoerde toetsen op dezelfde data toeneemt. Aangezien dit in het onderzoek het geval is, is een Bonferroni-correctie op zijn plaats. In dit stadium van het verslag wordt de correctie nog niet toegepast, maar is besloten voorlopig een redelijk conservatief significantieniveau te hanteren, dat wil zeggen:  $\alpha = .01$ . Gezien het aantal waarnemingen kan dit geen probleem voor de *power* van de toets opleveren. In de conclusie en discussie (Hoofdstuk 7) zal op dit onderwerp teruggekomen worden.

Uit tabel 6.16 valt op te maken dat bij een  $\alpha$  van .01, de prestaties van de leerlingen bij 27 van de 53 opgaven significant verschillen. Bij zeven andere items is er een trendmatig verschil bij een  $\alpha$  van .05. In de tabel staat een significant verschil met '\*\*' aangegeven en een trendmatig verschil met '\*'. In de meeste gevallen is het verschil in de verwachte richting. Dat wil zeggen, dat de gemiddelde proportie correcte antwoorden van de leerlingen lager is bij een leerkracht die een opgave als *niet maakbaar* beoordeelde, dan bij een leerkracht die een opgave als *maakbaar* beoordeelde. In drie gevallen echter — bij opgave 2, 14 en 20 van deel B — is het effect omgekeerd: leerlingen geven vaker het correct antwoord wanneer het item als niet maakbaar is beoordeeld, dan wanneer het item als maakbaar is beoordeeld. Voor dit effect zijn verschillende verklaringen te geven. Een verklaring is dat er met het item iets bijzonders aan de hand is. Bijvoorbeeld, omdat het item vaardigheden toetst die veelal niet op school, maar thuis worden aangeleerd, zoals bij klok kijken het geval zou kunnen zijn. Een andere — waarschijnlijker — verklaring is dat er een beslissingsfout bij de indeling van leerkrachten in TCO-groepen (maakbaar en niet maakbaar) gemaakt is. De leerkrachten beoordeelden het item als 'niet maakbaar', terwijl het in werkelijkheid maakbaar is. De mogelijkheid dat dit effect heeft gehad op de resultaten van de t-toets bestaat, omdat deze resultaten niet alleen gebaseerd zijn op de antwoorden van de leerlingen, maar ook op de antwoorden van de leerkrachten. Eén beoordelingsfout van leerkracht kan een resultaat van een t-toets substantieel beïnvloeden, omdat hiermee tot-en-met 30 leerlingen van bijvoorbeeld 'maakbaar' naar 'niet maakbaar' kunnen verschuiven. Dit kan verklaren waarom sommige items door 'niet maakbare' leerlingen beter worden gemaakt dan door 'maakbare' leerlingen. Bij de items 2, 14 en 20 van deel B kan het gevonden verschil op toeval gebaseerd zijn.

De itemmethode lijkt wel valide als de LVS-scores als criteriummaat genomen worden. Uit de t-toets blijkt immers een duidelijk verband tussen de maakbaarheid van toets E3 en de mate waarin leerlingen de toets correct beantwoorden. Bij een groot deel van de items wordt er door de leerlingen beter gepresteerd wanneer de leerkrachten het item als maakbaar beoordelen dan wanneer zij deze niet als maakbaar beoordelen.

Tabel 6.15

Overzicht van gemiddelde p-waarde en standaardafwijking van de leerkrachtoordelen op de Itemmethode en van de leerlingen op de items van toets E3

| Deel | Item | TCO      |     | LVS      |     |
|------|------|----------|-----|----------|-----|
|      |      | p-waarde | Sd. | p-waarde | Sd. |
| A    | 1    | .96      | .19 | .83      | .38 |
|      | 2    | .99      | .11 | .95      | .21 |
|      | 3    | .98      | .13 | .91      | .28 |
|      | 4    | .99      | .11 | .95      | .22 |
|      | 5    | .91      | .28 | .87      | .34 |
|      | 6    | .41      | .49 | .54      | .50 |
|      | 7    | .95      | .21 | .87      | .34 |
|      | 8    | .84      | .37 | .85      | .35 |
|      | 9    | .92      | .27 | .92      | .28 |
|      | 10   | .91      | .28 | .91      | .28 |
|      | 11   | .77      | .42 | .77      | .42 |
|      | 12   | .85      | .35 | .88      | .32 |
|      | 13   | .84      | .37 | .81      | .39 |
|      | 14   | .91      | .28 | .96      | .20 |
|      | 15   | .78      | .41 | .78      | .42 |
|      | 16   | .75      | .44 | .65      | .48 |
|      | 17   | .86      | .34 | .82      | .39 |
|      | 18   | .83      | .38 | .83      | .37 |
|      | 19   | .92      | .27 | .85      | .36 |
|      | 20   | .63      | .48 | .66      | .47 |
|      | 21   | .95      | .22 | .86      | .35 |
|      | 22   | .83      | .37 | .82      | .39 |
|      | 23   | .76      | .42 | .91      | .28 |
|      | 24   | .65      | .48 | .75      | .43 |
| B    | 1    | .99      | .11 | .99      | .12 |
|      | 2    | .99      | .11 | .98      | .14 |
|      | 3    | .98      | .13 | .95      | .22 |
|      | 4    | .98      | .13 | .94      | .24 |
|      | 5    | .98      | .15 | .95      | .21 |
|      | 6    | .96      | .20 | .86      | .35 |
|      | 7    | .98      | .13 | .92      | .27 |
|      | 8    | .83      | .37 | .75      | .43 |
|      | 9    | .78      | .41 | .72      | .45 |
|      | 10   | .80      | .40 | .78      | .42 |
|      | 11   | .91      | .28 | .69      | .46 |
|      | 12   | .97      | .17 | .88      | .32 |
|      | 13   | .95      | .21 | .96      | .20 |
|      | 14   | .99      | .08 | .91      | .29 |
|      | 15   | .85      | .36 | .79      | .41 |
|      | 16   | .76      | .43 | .73      | .44 |
|      | 17   | .77      | .42 | .85      | .36 |
|      | 18   | .92      | .27 | .83      | .38 |
|      | 19   | .79      | .40 | .90      | .31 |
|      | 20   | .99      | .08 | .95      | .21 |
|      | 21   | .84      | .37 | .81      | .39 |
|      | 22   | .58      | .49 | .63      | .48 |
|      | 23   | .63      | .48 | .62      | .48 |
|      | 24   | .94      | .24 | .86      | .35 |
|      | 25   | .81      | .39 | .74      | .44 |
|      | 26   | .88      | .33 | .82      | .38 |
|      | 27   | .86      | .34 | .80      | .40 |
|      | 28   | .62      | .49 | .72      | .45 |
|      | 29   | .47      | .50 | .43      | .50 |

Tabel 6.16  
T-toets voor het verschil tussen niet maakbare en maakbare items

| Deel     | Item  | Niet maakbaar |      | Maakbaar |     | t      | df       | sign. | gem.  |
|----------|-------|---------------|------|----------|-----|--------|----------|-------|-------|
|          |       | n1            | p1   | n2       | p2  |        |          |       |       |
| <b>A</b> | 1     | 103           | .78  | 3069     | .83 | -1.203 | 107.669  | .232  | -.050 |
|          | 2     | 4             | .50  | 3207     | .95 | -1.567 | 3.001    | .215  | -.450 |
|          | 3     | 65            | .92  | 3145     | .91 | .265   | 66.912   | .792  | 0.089 |
|          | 4     | 10            | .80  | 3191     | .95 | -1.107 | 9.016    | .297  | -.150 |
|          | 5 **  | 163           | .68  | 3000     | .88 | -5.264 | 170.842  | .000  | -.200 |
|          | 6 **  | 1804          | .50  | 1237     | .61 | -6.156 | 2695.990 | .000  | -.110 |
|          | 7 **  | 1876          | .85  | 1282     | .90 | -4.198 | 3019.394 | .000  | -.050 |
|          | 8 **  | 539           | .80  | 2632     | .86 | -3.387 | 709.475  | .001  | -.063 |
|          | 9     | 206           | .93  | 2984     | .92 | .635   | 238.561  | .526  | .012  |
|          | 10    | 252           | .89  | 2931     | .92 | -1.167 | 286.501  | .244  | -.023 |
|          | 11 ** | 713           | .72  | 2406     | .79 | -3.528 | 1085.761 | .000  | -.066 |
|          | 12    | 387           | .89  | 2796     | .88 | .740   | 510.783  | .460  | .012  |
|          | 13    | 519           | .79  | 2671     | .82 | -1.410 | 710.237  | .159  | -.027 |
|          | 14    | 273           | .97  | 2902     | .96 | 1.170  | 348.431  | .243  | .013  |
|          | 15 ** | 679           | .74  | 2520     | .79 | -2.583 | 1016.747 | .010  | -.048 |
|          | 16 ** | 756           | .58  | 2332     | .67 | -4.405 | 1229.451 | .000  | -.090 |
|          | 17 *  | 381           | .77  | 2798     | .82 | -2.077 | 470.822  | .038  | -.047 |
|          | 18 *  | 457           | .79  | 2715     | .84 | -2.315 | 588.456  | .021  | -.047 |
|          | 19 ** | 197           | .71  | 2989     | .86 | -4.396 | 211.714  | .000  | -.150 |
|          | 20 ** | 1069          | .61  | 2013     | .68 | -3.795 | 2093.967 | .000  | -.069 |
|          | 21 ** | 130           | .75  | 3077     | .86 | -2.871 | 135.932  | .005  | -.110 |
|          | 22 ** | 439           | .71  | 2758     | .84 | -5.648 | 534.144  | .000  | -.130 |
|          | 23 ** | 750           | .84  | 2420     | .94 | -6.871 | 962.704  | .000  | -.099 |
|          | 24 ** | 1168          | .68  | 1970     | .79 | -6.736 | 2193.301 | .000  | -.110 |
| <b>B</b> | 1     | 52            | .98  | 3129     | .98 | -.218  | 52.313   | .829  | -.042 |
|          | 2 **  | 52            | 1.00 | 3119     | .98 | 8.210  | 3118.000 | .00   | .021  |
|          | 3 *   | 64            | .98  | 3103     | .95 | 2.229  | 71.393   | .029  | .036  |
|          | 4     | 64            | .91  | 3100     | .94 | -.878  | 64.746   | .383  | -.033 |
|          | 5 *   | 77            | .86  | 3098     | .95 | -2.407 | 77.338   | .018  | -.097 |
|          | 6 *   | 142           | .79  | 2960     | .86 | -2.169 | 150.595  | .032  | -.076 |
|          | 7     | 147           | .88  | 2987     | .92 | -1.551 | 155.897  | .123  | -.043 |
|          | 8 **  | 475           | .64  | 2577     | .77 | -5.262 | 617.478  | .000  | -.120 |
|          | 9     | 574           | .71  | 2482     | .72 | -.446  | 851.657  | .656  | -.093 |
|          | 10    | 515           | .76  | 2561     | .78 | -1.050 | 720.585  | .294  | -.022 |
|          | 11 *  | 252           | .48  | 2921     | .71 | -6.963 | 287.603  | .000  | -.230 |
|          | 12    | 32            | .78  | 3162     | .88 | -1.387 | 31.364   | .175  | -.100 |
|          | 13 *  | 91            | .89  | 3121     | .96 | -2.068 | 92.115   | .041  | -.069 |
|          | 14 ** | 14            | 1.00 | 3152     | .91 | 18.004 | 3151.000 | .000  | .093  |
|          | 15    | 454           | .77  | 2653     | .79 | -1.313 | 602.643  | .190  | -.028 |
|          | 16 ** | 763           | .68  | 2346     | .75 | -3.745 | 1216.973 | .000  | -.072 |
|          | 17 *  | 698           | .82  | 2447     | .86 | -2.504 | 1043.006 | .012  | -.041 |
|          | 18    | 194           | .82  | 2981     | .83 | -.303  | 217.746  | .762  | -.087 |
|          | 19    | 569           | .87  | 2596     | .90 | -1.768 | 782.607  | .077  | -.027 |
|          | 20 ** | 15            | 1.00 | 3195     | .95 | 12.761 | 3194.000 | .000  | .049  |
|          | 21 ** | 517           | .77  | 2631     | .82 | -2.612 | 693.526  | .009  | -.052 |
|          | 22 ** | 1277          | .57  | 1772     | .67 | -5.225 | 2668.985 | .000  | -.093 |
|          | 23 ** | 1200          | .57  | 1862     | .66 | -4.813 | 2477.457 | .000  | -.087 |
|          | 24    | 151           | .81  | 3045     | .86 | -1.397 | 162.020  | .164  | -.045 |
|          | 25 ** | 585           | .68  | 2569     | .75 | -3.095 | 830.435  | .002  | -.065 |
|          | 26 ** | 377           | .73  | 2817     | .84 | -4.434 | 448.607  | .000  | -.110 |
|          | 27 ** | 386           | .65  | 2818     | .82 | -6.856 | 454.564  | .000  | -.170 |
|          | 28 ** | 1223          | .66  | 1880     | .76 | -6.057 | 2415.427 | .000  | -.100 |
|          | 29 ** | 1722          | .40  | 1434     | .48 | -4.462 | 3028.093 | .000  | -.079 |

## 6.7.2 Validering van de subcategorieën

Het valideren van de subcategorieën kan alleen maar op een indirecte manier, omdat toets E3 alleen items en geen subcategorieën heeft. Ter validering van de subcategorieën zijn twee stappen ondernomen. De eerste stap is het aantonen van een verband tussen TCO gemeten met de subcategorieën en de TCO gemeten met de items. Dit is in paragraaf 6.2.3.3 en paragraaf 6.5 beschreven. De tweede stap is het aantonen van een verband tussen TCO gemeten met de subcategorieën en de LVS-scores — de resultaten van de leerlingen op toets E3 — gemeten met de items. Dit is gedaan door de gemiddelde maakbaarheidsoordelen van de leerkrachten op de subcategorieën te vergelijken met de gemiddelde score van de leerlingen *op de bij de subcategorieën behorende items*. Tabel 6.10 bevat informatie over welke items inhoudelijk bij welke subcategorieën horen.

In tabel 6.17 staan over de leerkrachten gemiddelde maakbaarheidsoordelen van de subcategorieën, de gemiddelde maakbaarheidsoordelen van de bij de subcategorieën behorende items en de proportie correcte antwoorden op deze items. In de eerste kolom van de tabel staan de subcategorieën van toets E3. In de tweede kolom staat de proportie leerkrachten die de subcategorie als maakbaar beoordelen. In de derde kolom staat de gemiddelde proportie leerkrachten die de bij de subcategorie behorende items als maakbaar beoordelen. In de vierde kolom staat de gemiddelde proportie correcte antwoorden over de bij de subcategorieën behorende items.

Tabel 6.17  
Gemiddelde p-waarde over de subcategorieën

| Subcategorie                           | TCO | gem.<br>TCO | gem.<br>LVS |
|----------------------------------------|-----|-------------|-------------|
| 1. Structuur van de telrij             | .98 | .98         | .91         |
| 2. Resultatief en structurerend tellen | .97 | .95         | .80         |
| 3. Vergelijkingen en ordenen           | .95 | .79         | .76         |
| 4. Splitsen                            | .79 | .85         | .81         |
| 5. Samenstellen                        | .89 | .92         | .92         |
| 6. Aanvullen                           | .81 | .83         | .85         |
| 7. Optellen                            | .96 | .89         | .88         |
| 8. Aftrekken                           | .92 | .89         | .85         |
| 9. Diversen                            | .81 | .81         | .78         |
| 10. Optellen                           | .97 | .99         | .97         |
| 11. Aftrekken                          | .96 | .98         | .92         |
| 12. Splitsen                           | .83 | .81         | .75         |
| 13. Meten                              | .88 | .88         | .81         |
| 14. Tijd                               | .59 | .63         | .70         |

Aan de gemiddelde proporties — kolommen drie en vier — is te zien dat de maakbaarheidsoordelen van de subcategorieën sterk overeenkomt met de moeilijkheid van de subcategorieën.

In paragraaf 6.5, waar de dimensionaliteit van de subcategorieën werd onderzocht, bleken de resultaten van de diverse analyses niet alle in dezelfde richting te wijzen. Toch leken bij deze analyses de factoren of de componenten eerder de onderliggende onderwerpen (de categorieën) te representeren dan één onderliggende maakbaarheidsdimensie. De overeenkomst tussen de TCO- en de LVS-scores is daarom wellicht te verklaren door gemeenschappelijke onderwerpen in beide toetsen.

### 6.7.3 Validiteit van de lesmethoden

Het vaststellen van de criteriumvaliditeit van de derde methode om TCO te meten, de lesmethode, moet net als bij de subcategorieën indirect gebeuren. Om de criteriumvaliditeit van de lesmethoden vast te stellen, is eerst onderzocht of de gemiddelde toetsscores van leerlingen die bepaalde lesmethoden gebruiken, verschillen. Vervolgens is gekeken of het patroon van minder maakbare items bij bepaalde lesmethoden (zie tabel 6.8) terug te vinden is in de prestaties van de leerlingen. Deze analyses worden achtereenvolgens besproken.

In paragraaf 6.2.3.2 bleek dat de maakbaarheidsoordelen tussen leerkrachten die diverse lesmethoden gebruiken, verschillen. In tabel 6.18 staan de prestaties van de leerlingen die diverse lesmethoden gebruiken. In de eerste kolom staan de lesmethoden, in de tweede kolom het gemiddelde aantal correct beantwoorde vragen, in de derde kolom de standaardafwijking van het aantal correcte antwoorden en in de vierde kolom de betrouwbaarheid van de toets (Cronbach's alpha). Het blijkt dat het gemiddelde aantal correcte antwoorden van de verschillende lesmethoden, net als de maakbaarheidsoordelen (zie tabel 6.7), nauwelijks verschilt. Dit *overall*-resultaat impliceert wederom niet dat er op itemniveau geen verschillen bestaan.

Tabel 6.18  
Prestatie van leerlingen onder de diverse lesmethoden

| Lesmethode                 | Gem. | Sd. | Alpha |
|----------------------------|------|-----|-------|
| Wereld in getallen (oud)   | 42.9 | 7.4 | .88   |
| Wereld in getallen (nieuw) | 43.5 | 6.9 | .87   |
| Operator rekenen (oud)     | 42.1 | 8.1 | .90   |
| Operator rekenen (nieuw)   | 44.5 | 6.4 | .86   |
| Pluspunt                   | 43.7 | 7.9 | .90   |
| Rekenen en wiskunde        | 42.0 | 8.0 | .89   |

Om te achterhalen waar de overeenkomsten en verschillen zijn tussen de TCO-scores en de toetsscores van leerlingen is een overzicht gemaakt van de proportie correcte antwoorden op iedere vraag bij de zes meest gebruikte methoden. Dit overzicht is opgenomen in bijlage 4. In tabel 6.8 in paragraaf 6.2.3.2 staat een overzicht van enkele TCO-items die bij bepaalde lesmethoden als *opvallend minder maakbaar* waren beoordeeld. Tabel 6.19 heeft een vergelijkbare opzet als tabel 6.8, maar in plaats van maakbaarheidsoordelen staan in tabel 6.19 de toetsscores van leerlingen. De tabel kan daarom gebruikt worden om te kijken of de qua TCO-score opvallende lesmethoden ook qua de LVS-scores opvallen. Een plus (+) is in tabel 6.19 gebruikt als de LVS-scores van leerlingen, net als de TCO-score van leerkrachten inderdaad lager is bij de betreffende lesmethode. Een min (-) is gebruikt als dat niet zo is. Dus hoe meer plussen de tabel bevat, hoe groter de criteriumvaliditeit van de lesmethoden is als methode om TCO te meten. Let wel op: het gaat in deze tabel niet om *significante verschillen*, maar om een lagere toetsscore bij een lesmethode.

Het blijkt dat op itemniveau de LVS-scores tussen leerlingen die andere lesmethoden gebruiken verschillen. Deze verschillen lopen echter niet gelijk op met de TCO-scores van de leerkrachten. In de tabel is immers te zien dat een opvallend lage TCO-score op een item bij een lesmethode zeker niet altijd met lage LVS-scores gepaard is gegaan.

Samenvattend is in de drie voorafgaande paragrafen aangetoond dat er een verband is tussen de maakbaarheidsoordelen van de leerkrachten en de toetsscores van de leerlingen op toets E3. Het verloop van het verband blijkt echter erg ingewikkeld en lijkt tussen de drie methoden te verschillen.

Als men itemgewijs kijkt zijn de itemscores van leerlingen bij de meeste items hoger wanneer de leerkracht het item als maakbaar beoordeeld dan wanneer deze het item als niet maakbaar beoordeeld. Bij de subcategorieën daarentegen gaan hoge maakbaarheidsscores bij alle subcategorieën gepaard met hoge toetsscores. Bij de lesmethoden, ten slotte, gaan opvallend lage TCO-scores lang niet altijd gepaard met lage toetsscores.

Omdat het verband tussen TCO en de toetsscores van de leerlingen zo complex blijkt, is het ingewikkeld de keuze voor een TCO-instrument op basis van de validiteit van de meetmethoden te bepalen. Geen van de methoden afzonderlijk lijken de complexiteit van het verband echt te vangen, dus een combinatie van methoden lijkt de voorkeur te verdienen. Het lijkt het van belang om lesmethode in een TCO-instrument mee te nemen, omdat de verschillen tussen toetsscores van leerlingen deels door verschillen tussen lesmethoden te verklaren zijn. Daarnaast lijkt de itemmethode de beste keuze, omdat deze het de genoemde complexiteit beter weerspiegelt dat de subcategoriemethode.

Tabel 6.19

Overzicht van opvallende TCO-scores bij diverse lesmethoden, die wel (+) of niet (-) zijn terug te vinden in de toetsscores van leerlingen

| Deel | Item | Wereld in<br>getallen<br>(oud)<br>n= 19 | Wereld in<br>getallen<br>(nieuw)<br>n= 47 | Operator<br>rekenen<br>(oud)<br>n= 10 | Operator<br>rekenen<br>(nieuw)<br>n= 7 | Pluspunt<br>n= 46 | Rekenen<br>en<br>wiskunde<br>n= 34 |
|------|------|-----------------------------------------|-------------------------------------------|---------------------------------------|----------------------------------------|-------------------|------------------------------------|
| A    | 05   |                                         |                                           |                                       | —                                      |                   |                                    |
|      | 08   | —                                       | —                                         |                                       |                                        |                   |                                    |
|      | 10   |                                         |                                           |                                       |                                        |                   | +                                  |
|      | 11   | +                                       | +                                         |                                       |                                        |                   |                                    |
|      | 12   |                                         |                                           |                                       | —                                      |                   |                                    |
|      | 13   | +                                       | +                                         |                                       |                                        |                   |                                    |
|      | 14   | +                                       |                                           |                                       |                                        |                   |                                    |
|      | 15   | —                                       | —                                         |                                       |                                        |                   |                                    |
|      | 16   | —                                       | —                                         |                                       |                                        |                   |                                    |
|      | 17   | —                                       |                                           |                                       |                                        |                   | —                                  |
|      | 20   |                                         |                                           |                                       |                                        | —                 | —                                  |
|      | 22   |                                         |                                           | +                                     | —                                      |                   | +                                  |
|      | 23   |                                         |                                           |                                       |                                        | +                 | +                                  |
|      | 24   |                                         |                                           |                                       |                                        |                   | —                                  |
| B    | 08   |                                         |                                           |                                       |                                        | —                 | +                                  |
|      | 09   |                                         |                                           |                                       |                                        | —                 | —                                  |
|      | 10   |                                         |                                           |                                       |                                        | +                 | +                                  |
|      | 15   | —                                       | —                                         |                                       |                                        |                   |                                    |
|      | 16   | +                                       | +                                         |                                       |                                        |                   |                                    |
|      | 19   | +                                       | —                                         |                                       |                                        |                   |                                    |
|      | 21   | +                                       | +                                         |                                       |                                        |                   |                                    |
|      | 22   | +                                       | —                                         |                                       |                                        |                   | —                                  |
|      | 23   | +                                       | —                                         |                                       |                                        |                   |                                    |
|      | 25   |                                         |                                           |                                       |                                        |                   | —                                  |
|      | 26   |                                         |                                           | —                                     |                                        |                   |                                    |
|      | 27   |                                         |                                           | +                                     | +                                      |                   | +                                  |
|      | 28   |                                         |                                           |                                       | —                                      |                   | —                                  |
|      | 29   | +                                       | —                                         |                                       |                                        | —                 | +                                  |

## 6.8 Detectie van vraagonzuiverheid

In paragraaf 6.6 is getracht de vragen van de TCO-vragenlijst met behulp van itemresponsmodellen op één schaal te plaatsen. In de laatste paragrafen van dit memorandum wordt beschreven hoe is getracht de items van toets E3 op één schaal te plaatsen met het doel om *differential item functioning* (DIF) op te sporen. Bij het onderzoeken van DIF wordt nagegaan of de gemiddelde itemscore bij even vaardige subgroepen van een populatie gelijk is (Holland & Wainer, 1993). Wanneer dit niet het geval is, functioneert het item anders tussen de subgroepen. Men spreekt van *itembias* of vraagonzuiverheid wanneer de verschillen tussen de groepen veroorzaakt worden door de aspecten van de vraag die niet relevant zijn voor wat men beoogt te meten.

Vraagonzuiverheid wordt in de paragrafen 6.8.2 tot en met 6.8.4 opgespoord door leerlingen aan verschillende groepen toe te kennen en vervolgens te kijken of leerlingen met een gelijke vaardigheid een ongelijke kans hebben het item goed te maken. Gezocht wordt daarbij naar twee vormen van vraagonzuiverheid: vraagonzuiverheid samenhangend met TCO en vraagonzuiverheid samenhangend met het gebruik van verschillende lesmethoden.

Vraagonzuiverheid wordt in een viertal analyses onderzocht. In de eerste analyse is onderzocht of er sprake is van vraagonzuiverheid wanneer scholen verschillende lesmethoden gebruiken. Deze analyse wordt in paragraaf 6.8.2 besproken. In de tweede analyse in paragraaf 6.8.3 wordt vraagonzuiverheid samenhangend met TCO onderzocht door de leerlingen in vier groepen te verdelen met betrekking tot de maakbaarheidsoordelen van hun leerkrachten. In de derde analyse in paragraaf 6.8.4 is vraagonzuiverheid samenhangend met TCO itemgewijs onderzocht. De algemene calibratie van toets E3 wordt eerst in paragraaf 6.8.1 besproken.

### 6.8.1 Algemene calibratie van de items van toets E3

Om de items van toets E3 op één schaal te plaatsen, zijn drie stappen ondernomen. Allereerst is gekeken of de oude itemparameterschattingen van een afname van de toets in 1989 een goede modelpassing opleveren en is een Raschmodel gepast. Daarna is naar een passend OPLM gezocht.

De oude itemparameterschattingen van een afname van toets E3 in 1989 bleken niet meer goed op de huidige data te passen. Wanneer de oude schattingen gebruikt worden, past het OPLM slecht ( $R_{1c} = 514.342$ ,  $v = 154$ ,  $p = .000$ ). Drieëntwintig van de 53 items weken significant af van het model. De moeilijkheidsgraad van de items is veranderd en bestaande normtabellen zijn verouderd en mogen niet meer gebruikt worden. Voldoende reden daarom om een nieuwe calibratie uit te voeren. Eerst is een Raschmodel gepast. De passing van dit model was nog slechter dan het model van 1989 ( $R_{1c} = 860$ ,  $v = 534$ ,  $p = .000$ ). Daarna is het OPLM gepast. Het OPLM blijkt een goede passing ( $R_{1c} = 157.443$ ,  $v = 156$ ,  $p = .456$ ) op te leveren. Toch is de gemiddelde discriminatie-index van dit model met een waarde van 5.463, behoorlijk hoog. Verhelst (1992) waarschuwt dat men door het veelvuldig aanpassen van de discriminatie-indices van het model soms grote waarden voor de discriminatie-indices moet kiezen om ook kleine nuances in de data te ondervangen. Hij wijst in dit verband op het gevaar voor kanskapitalisatie. Hoewel bij het beschreven model de  $a_i$ -waarden hoog *zijn*, valt de spreiding daarentegen wel mee. Toch moet men bij het trekken van conclusies over dit model wel rekening houden met het feit dat de items van dit model erg sterk discrimineren.

De schattingen van de itemparameters voor de items van toets E3 van dit éénparametermodel staan in bijlage 7. De schattingen van de persoonsparameters en de bijbehorende schaalwaarden staan in dezelfde bijlage.

## 6.8.2 Vraagonzuiverheid samenhangend met de gebruikte lesmethode

Vraagonzuiverheid samenhangend met de gebruikte lesmethode is achterhaald door vier groepen leerlingen te creëren:

1. leerlingen die Wereld in getallen gebruiken (n=1307),
2. leerlingen die Operator rekenen gebruiken (n=347),
3. leerlingen die Pluspunt gebruiken (n=809) en
4. leerlingen die Rekenen en wiskunde (n=620) gebruiken.

De 133 leerlingen die een andere methode gebruiken, en de leerlingen van leerkrachten die de vraag naar lesmethodegebruik in de TCO-vragenlijst verkeerd hebben ingevuld, zijn buiten beschouwing gelaten. Wanneer er geen onzuiverheid in een item is, mag de kans het item correct te beantwoorden tussen leerlingen die verschillende lesmethoden gebruiken, maar een gelijke vaardigheid hebben, niet verschillen. Om dit te toetsen is naar significante afwijkingen tussen de vier groepen in de itemresponscurven gezocht.

In tabel 6.20 staat een overzicht van onzuivere items. Een plus in de tabel geeft aan dat de lesmethode op de meeste vaardigheidsniveaus wordt bevoordeeld. Een min geeft aan dat de lesmethode op de meeste vaardigheidsniveaus wordt benadeeld. Een plus/min geeft aan dat de lesmethode op ongeveer de helft van de vaardigheidsniveaus wordt bevoordeeld en op andere helft wordt benadeeld. Voor alle items geldt dat ze niet voor alle vaardigheidsniveaus — niet uniform — onzuiver zijn.

Een voorbeeld van een onzuiver item is item 23 van deel A. In bijlage 8 zijn van dit item vier figuren opgenomen waarin de verwachte proportie correcte antwoorden (de lijn met het horizontale streepjes) en de geobserveerde proportie (de lijn met de kruisjes) van elke lesmethode staan afgebeeld. Te zien is dat de proportie correcte antwoorden bij Operator rekenen voor alle vaardigheidsniveaus groter is dan voorspeld. Leerlingen die Wereld in getallen gebruiken hebben ook een voordeel, maar het voordeel wordt minder naarmate de vaardigheid van de leerlingen groter wordt. De leerlingen die Rekenen en wiskunde en Pluspunt gebruiken worden benadeeld bij dit item, maar dat nadeel is in het algemeen het grootst bij de minder vaardige leerlingen.

Tabel 6.20  
Overzicht van lesmethode-onzuivere items

| Deel | Item | Wereld in getallen | Operator rekenen | Pluspunt | Rekenen en wiskunde |
|------|------|--------------------|------------------|----------|---------------------|
| A    | 04   | -                  | -                | +        | +/-                 |
|      | 05   | +                  | +                | -        | -                   |
|      | 07   | +                  | -                | -        | +                   |
|      | 10   | +                  | +                | +/-      | -                   |
|      | 19   | +                  | -                | -        | -                   |
|      | 22   | +/-                | -                | +        | -                   |
|      | 23   | +                  | +                | -        | -                   |
| B    | 05   | +                  | +                | -        | -                   |
|      | 10   | +/-                | +/-              | -        | +/-                 |
|      | 11   | -                  | +                | +        | -                   |
|      | 12   | +                  | +                | +/-      | +/-                 |
|      | 16   | +/-                | +                | +/-      | +                   |
|      | 26   | +                  | -                | +        | -                   |
|      | 27   | +/-                | -                | +        | +/-                 |

Uit tabel 6.20 is af te lezen in welke richting items onzuiver zijn bij het gebruik van bepaalde lesmethoden. Uit het overzicht blijkt dat het niet altijd dezelfde lesmethoden zijn die worden benadeeld. Wel valt op dat leerlingen die Wereld in getallen als lesmethode gebruiken vaker een voordeel van een onzuiver item ondervinden dan leerlingen die andere lesmethoden gebruiken. Om gerichte uitspraken te doen over de relatie tussen het onzuivere item en de lesmethode, kan men de figuren, zoals uit bijlage 8 raadplegen.

Geconcludeerd moet worden dat er sprake lijkt van vraagonzuiverheid door het gebruik van verschillende lesmethoden. Op veertien van de 53 items van toets E3 hadden leerlingen die een specifieke lesmethode gebruiken een voordeel boven de leerlingen die een andere lesmethode gebruiken. Leerlingen die Wereld in getallen rekenen gebruiken, zijn op de meeste items bevoordeeld. Daarna volgen leerlingen die Operator rekenen gebruiken. Leerlingen die Rekenen en wiskunde gebruiken zijn het meest frequent benadeeld. Leerlingen die Pluspunt gebruiken zijn minder vaak benadeeld. Toch geldt dit niet voor alle leerlingen die een lesmethode gebruiken, omdat het effect afhankelijk is van de vaardigheid van leerlingen. Wanneer men geïnteresseerd is in waar de verschillen exact zitten, moet men van elk item de plaatjes zoals van bijlage 8 raadplegen.

### 6.8.3 Vraagonzuiverheid samenhangend met TCO

Vraagonzuiverheid samenhangend met TCO is op twee manieren gemeten. De eerste manier om te achterhalen of de vraagonzuiverheid bij de items van toets E3 samenhangt met TCO is onderzocht door de leerkrachten in vier TCO-groepen van ongeveer gelijke grootte in te delen en te toetsen of er sprake is van vraagonzuiverheid tussen de groepen. Bij de tweede manier wordt vraagonzuiverheid samenhangend met TCO itemsgewijs onderzocht. Hoe de leerkrachten in groepen zijn ingedeeld, is in tabel 6.21 af te lezen. De verwachting is dat als er sprake is van vraagonzuiverheid leerlingen in de eerste twee groepen (waar minder TCO is), worden benadeeld door een item en leerlingen in de laatste twee groepen (waar meer TCO is) worden bevoordeeld.

De indeling van leerkrachten in vier groepen naar maakbaarheid kan op verschillende manieren geoperationaliseerd worden. Men kan de ruwe TCO-score van leerkrachten of de getransformeerde schaalscore voor de indeling gebruiken. Nadeel van beide manieren is dat de groepen niet van gelijke grootte zijn, waardoor een eventueel gevonden effect moeilijk te interpreteren is. Gekozen is daarom de leerkrachten te ordenen naar hun maakbaarheids-oordelen en hen vervolgens in vier ongeveer gelijke groepen in te delen. De exacte indeling is in tabel 6.21 te zien. In de eerste kolom staat het groepsnummer, in de tweede kolom staat het percentage leerkrachten dat tot de groep behoort en in de derde en vierde kolom staan respectievelijk de ruwe TCO-score en de schaalscore van de groepen.

Tabel 6.21  
De indeling van leerkrachten in vier TCO-groepen

| Groep | %    | Ruwe score | Schaalscore |
|-------|------|------------|-------------|
| 1     | 25.6 | 30-40      | 55-66       |
| 2     | 23.8 | 41-46      | 67-75       |
| 3     | 27.4 | 47-50      | 79-85       |
| 4     | 32.2 | 51-53      | 86-100      |

Volgens de tabel 6.22 is er bij vier items sprake van vraagonzuiverheid. Behalve dat de groepen anders zijn, is de indeling van deze tabel vergelijkbaar met die van tabel 6.21.

Tabel 6.22  
Overzicht van onzuivere items door TCO

| Deel | Item | Groep 1 | Groep 2 | Groep 3 | Groep 4 |
|------|------|---------|---------|---------|---------|
| A    | 07   | +       | +/-     | +/-     | -       |
|      | 24   | +       | -       | -       | +       |
| B    | 08   | -       | +       | +       | +/-     |
|      | 10   | +/-     | +/-     | -       | +/-     |

Zoals eerder gezegd was de verwachting dat als er sprake is van vraagonzuiverheid, de groepen met relatief weinig TCO benadeeld worden door een item en groepen met relatief veel TCO bevoordeeld wordt. Dit blijkt echter niet uit de tabel. Minder maakbaar impliceert niet perse een slechtere prestatie. Toch hoeft met dit resultaat niet meteen geconcludeerd te worden dat TCO geen rol speelt bij toets E3. Het gegeven dat de verschillen in ruwe score tussen groepen 2, 3 en 4 erg klein waren, kan verklaren dat er nauwelijks vraagonzuiverheid samenhangend met TCO gevonden is. Het gegeven dat de groepen zijn ingedeeld naar de TCO-scores van de leerkrachten *op de gehele toets*, kan verklaren dat de items de groepen niet in de verwachte richting bevoor- of benadelen. Door per item afzonderlijk naar vraagonzuiverheid te zoeken tussen leerkrachten die het item als maakbaar en als niet maakbaar beoordeelden, kunnen effecten van TCO zichtbaar worden, die eerder niet zichtbaar waren. Dit is in de volgende paragraaf onderzocht.

#### 6.8.4 Vraagonzuiverheid samenhangend met TCO, itemgewijs onderzocht

De procedure bij het itemgewijs onderzoeken van vraagonzuiverheid samenhangend met TCO is als volgt. Als uitgangspunt zijn de resultaten van een t-toets (tabel 6.15) gebruikt. Bij 26 items verschilden de resultaten van leerlingen significant tussen een als maakbaar en als niet maakbaar beoordeelde vraag. Per item is een databestand gemaakt waarin de maakbaarheidsoordelen van de leerkrachten op dat item staan. Elk databestand met de unieke informatie van de maakbaarheidsoordelen van de leerkrachten op één item is vervolgens afzonderlijk gecalibreerd. Daarna is onderzocht of er op dat specifieke item sprake is van vraagonzuiverheid tussen leerlingen wiens leerkracht het item als maakbaar beoordeelde en leerlingen wiens leerkracht het item als niet maakbaar beoordeelde. En indien er sprake was van vraagonzuiverheid welke groep het item bevoordeelde en welke groep het benadeelde.

In tabel 6.23 ligt het patroon binnen onzuivere items wel in de lijn der verwachting. Leerlingen ondervinden worden door een item benadeeld wanneer de leerkrachten het item als maakbaar beoordelen en worden bevoordeeld wanneer leerkrachten het item niet als maakbaar beoordelen.

Tabel 6.23  
Overzicht van onzuivere items door TCO

| Deel     | Item      | Niet<br>maakbaar | Maakbaar |
|----------|-----------|------------------|----------|
| <b>A</b> | <b>05</b> | -                | +        |
|          | <b>16</b> | -                | +/-      |
|          | <b>22</b> | -                | +        |
|          | <b>23</b> | -                | +        |
| <b>B</b> | <b>11</b> | -                | +        |
|          | <b>22</b> | -                | +        |
|          | <b>23</b> | -                | +        |
|          | <b>27</b> | -                | +        |
|          | <b>29</b> | -                | +        |

Geconcludeerd kan worden dat TCO wel degelijk effect heeft op de prestaties van leerlingen op toets E3. Een totaaloordeel van leerkrachten over TCO is echter niet voldoende om dit vast te stellen. Pas wanneer men de items van toets E3 afzonderlijk analyseert — maakbaar versus niet maakbaar — wordt het effect zichtbaar.

#### 6.8.5 Conclusies over vraagonzuiverheid

Er is in paragraaf 6.8.4 aangetoond dat er bij toets E3 sprake is van vraagonzuiverheid die samenhangt met TCO. In paragraaf 6.8.2 is aangetoond dat er sprake is van vraagonzuiverheid die samenhangt met de gebruikte lesmethoden. Onduidelijk is of er nog steeds sprake is van vraagonzuiverheid samenhangend met de gebruikte lesmethoden wanneer er voor TCO gecontroleerd wordt. Pogingen die ondernomen zijn om het voorgaande te onderzoeken, zijn mislukt door beperkingen binnen het besturingsprogramma Dos.

De opzet van de analyse zal besproken worden zodat de analyse later, bij uitkomen van het OPLM op een ander besturingsprogramma, wel uitgevoerd kan worden.

Het opsporen van vraagonzuiverheid samenhangend met lesmethode waarbij voor TCO gecontroleerd wordt, heeft de volgende procedure. Van elke leerkracht is een patroon bekend van maakbare en niet maakbare opgaven van toets E3. De 170 leerkrachten uit de steekproef hebben gezamenlijk 133 verschillende combinaties van maakbare en niet maakbare opgaven opgeleverd. De toetsresultaten van leerlingen zijn voor TCO gecorrigeerd door alleen de maakbare opgaven te gebruiken. In OPLM-taal: de leerlingen hebben 133 verschillende boekjes gekregen. Vervolgens kan onderzocht worden of er nog steeds sprake is van onzuiverheid samenhangend met het gebruik van verschillende lesmethoden als voor TCO gecorrigeerd is.

## 7. Conclusie en discussie

Drie vraagstellingen stonden centraal bij constructie van het instrument Toets Curriculum Overlap (TCO). De eerste vraag was in welke mate er sprake is van TCO. De tweede vraag was of het mogelijk is een valide en betrouwbaar instrument TCO te ontwikkelen. Een derde onderzoeksvraag was de vraag hoe een instrument TCO vormgegeven moest worden.

De eerste onderzoeksvraag gaat in op de noodzaak van een TCO-instrument. In andere woorden: is er een zodanig *gebrek* aan TCO dat het afnemen van een instrument TCO noodzakelijk is? Om deze eerste onderzoeksvraag te beantwoorden is voor de drie methoden van het TCO-instrument — subcategorie, item, lesmethode — een overzicht gemaakt van de p-waarde en van de decielscores van de leerkrachten op de vragen van deze methoden. Gezien de hoge p-waarde van de meeste vragen leek de toets goed aan te sluiten bij het onderwijs van de leerkrachten. Toch blijkt het afnemen van de vragenlijst noodzakelijk, omdat de opvattingen tussen leerkrachten nogal kunnen verschillen. Een deel van de leerkrachten vindt toets E3 grotendeels maakbaar, maar een ander deel vindt dat niet. In het bijzonder bij deze laatste groep leerkrachten zou de vaardigheid van de leerlingen onderschat kunnen worden omdat enkele van de items uit de toets in het onderwijs aan de orde zijn geweest.

De tweede onderzoeksvraag is of het mogelijk is een valide en betrouwbaar TCO-instrument te ontwikkelen. Om deze onderzoeksvraag te kunnen beantwoorden zijn drie meetmethoden in de TCO-vragenlijst geoperationaliseerd: het stellen van vragen over de maakbaarheid van subcategorieën waar de lesstof van toets E3 in onderverdeeld kan worden, het stellen van vragen over de maakbaarheid van de items van toets E3 en het stellen van vragen over de lesmethoden die leerkrachten in het onderwijs gebruiken. De betrouwbaarheid en de validiteit van twee meetmethoden zijn met verschillende analyses vastgesteld.

Zowel met de subcategorie- als de itemmethode wordt TCO betrouwbaar gemeten, hoewel de betrouwbaarheid van de itemmethode wel beter is dan van de subcategorie-methode.

De validiteit van de vragen van de meetmethoden is op verschillende plaatsen in het memorandum aan de orde gekomen. Eerst zijn de meetmethoden zelf als maten voor de criteriumvaliditeit gebruikt. Het blijkt dat de drie meetmethoden een ander beeld van TCO opleveren. Zo lijkt TCO veel groter wanneer men de subcategoriemethode gebruikt dan wanneer men de itemmethode gebruikt. Ook verschilt de mate waarin TCO gerapporteerd wordt tussen de lesmethoden.

De begripsvaliditeit van de methoden is met de PAF, de homogeniteitsanalyse en de PCA onderzocht. Hoewel niet alle resultaten eenduidig zijn, is er steeds één dimensie die de meeste variantie verklaart en die mogelijk als TCO of 'maakbaarheid' geïnterpreteerd kan worden. De tweede dimensie heeft mogelijk een inhoudelijke betekenis. Volgens de PCA lijken bij de subcategoriemethode meerdere dimensies van belang. Deze dimensies lijken als categorieën, waar de subcategorieën in onderverdeeld worden, geïnterpreteerd moeten worden. Gezien de vragenlijst is dat eigenlijk ook niet verwonderlijk. Immers de subcategorieën staan na de categorieën, waardoor het leerkrachten uitnodigt een hele categorie als maakbaar of als niet maakbaar te beoordelen. Bij de items is deze clustering rond de categorieën niet terug te vinden. Na de dimensionaliteitsanalyses is met IRT getracht de subcategorieën en de items op één schaal te plaatsen. Voor beide methoden is een goed passend OPLM gevonden.

Door de uitkomsten van de drie methoden met de resultaten van de leerlingen op toets E3 in verband te brengen, kan de criteriumvaliditeit van de methoden onderzocht worden. De criteriumvaliditeit van zowel de subcategorie- als de itemmethode lijkt goed. Er is een verband tussen TCO en de toetsscores van leerlingen gevonden. Toch is het verband nogal ingewikkeld en lijkt deze niet gelijk te zijn voor de drie methoden. Verder is bij negen items DIF samenhangend met TCO aangetoond. Dat wil zeggen dat leerlingen van leerkrachten die de items als niet maakbaar beoordeelden een nadeel van de items ondervinden ten opzichte van leerlingen van leerkrachten die de items wel als maakbaar beoordeelden.

De derde onderzoeksvraag is hoe een TCO-instrument vormgegeven moet worden. Een combinatie van de itemmethode en een vraag naar de gebruikte lesmethode lijkt als TCO-instrument het meeste perspectief te bieden. Uit zowel de analyse van de proportie maakbare vragen (Paragraaf 6.2 tot en met 6.2.4) als de analyse van de dimensionaliteit van de meetmethoden (Paragraaf 6.5 tot en met 6.5.3) is gebleken dat de keuze van het instrument gevolgen heeft voor het beeld dat men van TCO krijgt. De lesmethode- en subcategorie-methode lijken daarbij een te eenvoudig beeld van TCO te bieden. Dit is niet het geval bij de itemmethode. Omdat de gerapporteerde TCO en de toetsscores van leerlingen op toets E3 tussen de gebruikte lesmethode blijkt te verschillen, lijkt een combinatie van deze methoden het beste. Met betrekking tot de lesmethoden is het alleen noodzakelijk de vraag naar de gebruikte lesmethode voor te leggen. De twee vragen die opgenomen waren om de twee verstorende variabelen 'planning' en 'gewoonte' te meten hoeven niet voorgelegd te worden omdat deze variabelen de gerapporteerde TCO niet beïnvloeden. Met betrekking tot de itemmethode moet overwogen worden of alle items om een selectie voor het TCO-instrument te gebruiken. De geringe variantie in de data — veel leerkrachten rapporteerden de items als maakbaar — pleit voor het selecteren van items met voldoende itemvariantie. Daarna kunnen nog items verwijderd worden die bijvoorbeeld een negatieve bijdrage aan de betrouwbaarheid leveren. Toch kan er ook voor gekozen worden om alle items te gebruiken. Als men een selectie van items gebruikt, moet men van de items die men niet opneemt aannemen dat zij maakbaar waren. Dit kan voor de populatie van leerkrachten zo zijn, maar hoeft niet voor een individuele leerkracht te gelden.

Op welke manier kan het instrument het best gebruikt worden? Drie mogelijke toepassingen kunnen overwogen worden. De eerste mogelijke toepassing is dat de leerkracht zichzelf het TCO-instrument twee of drie maanden *voor* de afname van toets E3 afneemt, zodat de leerkracht de tijd heeft lesstof die nog niet is onderwezen, alsnog te onderwijzen. Het instrument wordt dan als 'entree-toets' gebruikt. Het voordeel hiervan is, dat geen correctie voor de vaardigheidsschattingen van leerlingen hoeft plaats te vinden, terwijl toch met TCO rekening is gehouden. Een gevaar bestaat dan wel dat leerkrachten letterlijk de items uit de toets aan de kinderen gaan voorleggen, of wel *teaching to the test*. Dit gevaar treedt op als de itemmethode als instrument TCO gekozen wordt. Een oplossing voor het probleem van *teaching to the test* zou zijn om voor het beoordelen van TCO een paralleltoets van toets E3 te gebruiken. Er is dan geen *teaching to the test*, maar *teaching to the construct*.

De tweede mogelijke toepassing is dat leerkrachten zichzelf het instrument afnemen net *voor* dat toets E3 wordt afgenomen en dat de leerlingen vervolgens alleen die items van toets E3 maken die de leerkracht maakbaar vond. Bij deze mogelijkheid verschilt de toetslengte en -samenstelling tussen leerlingen van verschillende leerkrachten. Hierdoor zal de geschatte vaardigheid van leerlingen ook verschillen. Onderzocht zal moeten worden of deze afwijkingen van de geschatte vaardigheid acceptabel zijn, of niet.

De derde mogelijke toepassing is om toets E3 in zijn geheel af te nemen en de toetsscores van leerlingen vervolgens *achteraf* te corrigeren voor TCO. Correctie vereist dan wel dat de relatie tussen de TCO- en de LVS-score bekend is teneinde te kunnen bepalen hoe groot die correctie moet zijn.

Ten slotte nog een kritische kanttekening. Aangetoond is dat het construct TCO bestaat en een advies is gegeven hoe het construct het beste gemeten kan worden. In het verslag is echter ook reeds op het gevaar voor kanskapitalisatie gewezen. Op de TCO-data, dat wil zeggen de oordelen van de leerkrachten, zijn veel analyses uitgevoerd. Zo zijn de waarden van de discriminatie-indices frequent veranderd om een goed passend OPLM-model te vinden. Het gevaar bestaat dan ook dat enkele resultaten specifiek voor de steekproef zijn.



## 8. Literatuur

- Cito-leerlingvolgsysteem (1992). *Rekenen-Wiskunde 1*. Arnhem: Cito.
- Eggen, T.J.H.M. & Sanders, P.F. (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Goffree, F., & Frowijn, R. (1996). *Op weg naar zelfevaluatie in het basisonderwijs. Katern 1: Toetsen en kerndoelen*. Enschede: SLO.
- Haan, D.M., de (1992). *Measuring Test-Curriculum Overlap*. Enschede: Thesis.
- Heuvelmans, A.P.J.M. (1998). *Constructie en verwerking van vragenlijsten*. OPD memorandum. Arnhem: Cito.
- Holland, P.W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Knol, D.L., & Ten Berge, J.M.F. (1989). Least squares approximation of an improper correlation matrix by a proper one. *Psychometrika*, 54, 53-61.
- Brink, W.P. van den, & Mellenbergh, G.J. (1998). *Testleer en testconstructie*. Amsterdam: Boom.
- Molenaar, I. W., Debets, P., Sijtsma, K., & Hemker, B.T. (1994). *MPS: A program for Mokken Scale Analysis for polytomous items*. Groningen: Iec ProGamma.
- Oosterveld, P., & Vorst, H.C.M. (1996) Methoden voor vragenlijstconstructie. *Nederlands Tijdschrift voor de Psychologie*, 51 (1).
- Scheerens, J. (1992). *Effective schooling: research, theory, and practice*. London/New York: Cassell.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Thorndike, R.L. (1982). *Applied Psychometrics*. Teachers College, Columbia University.
- Verhelst, N. D. (1992). *Het eenparameter logistisch model (OPLM). Een theoretische handleiding bij het computerprogramma*. Arnhem: Cito.
- Verhelst, N.D., & Engelen, R. (in press) *An ability estimator in the two parameter logistic model based on raw scores*.



## Bijlagen

|           |                                                                                                                                 |       |
|-----------|---------------------------------------------------------------------------------------------------------------------------------|-------|
| Bijlage 1 | Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en scoreformulieren voor de antwoorden van de leerlingen op toets E3 | I     |
| Bijlage 2 | De proportie maakbare vragen van de subcategorie- en de itemmethode bij de zes lesmethoden                                      | X     |
| Bijlage 3 | Drie spreidingsdiagrammen van homogeniteitsanalyses                                                                             | XIII  |
| Bijlage 4 | De proportie correcte antwoorden op toets E3 bij de zes lesmethoden                                                             | XV    |
| Bijlage 5 | Parameters voor de subcategoriemethode van de TCO-vragenlijst                                                                   | XVI   |
| Bijlage 6 | Parameters voor de itemmethode van de TCO-vragenlijst                                                                           | XVII  |
| Bijlage 7 | Parameters van toets E3                                                                                                         | XVIII |
| Bijlage 8 | Itemkarakteristieke curven van een lesmethode-onzuiver item                                                                     | XIX   |



## Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

### Toets Curriculum Overlap

#### Instructie

##### Algemeen

U neemt deel aan het onderzoek Toets Curriculum Overlap (TCO). Het onderzoek richt zich op leerkrachten die lesgeven aan groep 3. Voorwaarde voor deelname is dat u de toets E3 van het pakket Rekenen-Wiskunde 1 van het Cito-leerlingvolgsysteem (Cito-LVS) aan het einde van het leerjaar afneemt. Voor het onderzoek is het steeds van belang gegevens over het onderwijs van één klas te verzamelen. We verzoeken u telkens een eigen oordeel te geven over wat u in uw klas onderwezen hebt. Wat uw collega's in andere klassen onderwijzen kunt u buiten beschouwing houden.

##### Instructie bij de vragenlijst

Het Cito is momenteel bezig met het ontwikkelen van een instrument 'Toets Curriculum Overlap'. Met dit instrument kan gemeten worden in hoeverre de lesstof in de Cito-leerlingvolgsysteem toetsen aansluiten bij het verzorgde onderwijs. Naar onze opvatting is TCO op drie manieren vast te stellen. Het doel van dit onderzoek is na te gaan welke manier de voorkeur verdient.

De manieren zijn:

- het stellen van vragen over de **lesmethode** die leerkrachten gebruiken
- het stellen van vragen per **categorie** waarin de lesstof onderverdeeld kan worden
- het stellen van vragen per **opgave**

Hieronder volgen enkele instructies die voor het invullen van de vragenlijst van belang zijn:

- De vragen worden ingeleid. Wij verzoeken u elke inleiding steeds zorgvuldig door te nemen.
- De meeste vragen zijn gesloten. Antwoorden op deze vragen kunt u geven door het cijfer voor het antwoord van uw keuze te omcirkelen. Eventuele toelichtingen kunt u in de daarvoor gereserveerde ruimte geven.
- Neem het toetsboekje E3 en de handleiding uit het pakket Rekenen-Wiskunde 1 bij de hand. In de vragenlijst staat aangegeven wanneer u deze nodig hebt.
- Van belang is dat het meten van TCO leerkrachten zo weinig mogelijk tijd kost. Om een idee te krijgen hoeveel tijd het invullen van de vragenlijst per onderdeel — lesmethode (vraag 1 tot en met 3), categorie (vraag 5) en opgave (vraag 7) — kost, vragen we u dit aan het einde van elk onderdeel aan te geven. Wij attenderen u hierop bij aanvang van elk onderdeel.

##### Instructie bij de scoreformulieren

Voor het onderzoek zijn de antwoorden van uw leerlingen op de toets E3 van het pakket Rekenen-Wiskunde 1 nodig. Deze antwoorden noteert u op de twee scoreformulieren.

Op het eerste scoreformulier noteert u de antwoorden van uw leerlingen op DEEL 1 van de toets E3. Op het tweede, de antwoorden van uw leerlingen op DEEL 2 van de toets E3. Op beide formulieren treft u aan: een tabel met rijen (horizontaal) en kolommen (verticaal).

In de linker kolom staan volgnummers voor uw leerlingen. U hoeft dus niet de namen van uw leerlingen op te schrijven. In de overige kolommen staan de volgnummers van de opgaven van de toets E3. Per regel noteert u de antwoorden van uw leerlingen op de opgaven van toets E3 Rekenen-Wiskunde 1. Zorg dat u het juiste deel voor u heeft, voordat u met het invullen van het scoreformulier begint. Hanteer bij het invullen van de twee scoreformulieren dezelfde ordening voor uw leerlingen. Mocht u meer dan 35 leerlingen in één klas hebben, dan kunt u doornummers en de resultaten van deze leerlingen onderaan zetten.

##### Terugsturen materiaal

Alvast hartelijk dank voor uw medewerking! Wilt u de ingevulde vragenlijst en het ingevulde scoreformulier vóór 1 juli terug sturen? U kunt daarbij gebruik maken van bijgevoegde portvrije enveloppe. Mocht er iets gebeuren waardoor u de vragenlijst en de scoreformulieren niet op tijd aan ons ingevuld kunt terugsturen, dan vragen wij u in ieder geval de ingevulde vragenlijst TCO terug te sturen.

Ons adres is: Cito Instituut voor Toetsontwikkeling, Antwoordnummer 573, 6800 VB Arnhem

##### Vragen?

Voor vragen kunt u contact opnemen met: Henk Moelands (Tel: 026-3521562) of Sandra van Abswoude (Tel: 026-3521162).



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

**Vragenlijst Toets Curriculum Overlap**

Zoals reeds in de instructie staat, vragen we u in vraag 4 aan te geven hoeveel tijd het invullen u heeft gekost. Het eerste onderdeel (lesmethode) bestaat uit de vragen één tot en met drie.

**Onderdeel 1: Lesmethode**

- V-1 Voor het rekenonderwijs bij groep 3 maakt u gebruik van een lesmethode. Hieronder treft u een aantal gangbare lesmethoden aan. Welke lesmethode gebruikt u tijdens uw lessen in groep 3? (Omcirkel het cijfer voor het antwoord van uw keuze. Omcirkel niet meer dan één cijfer.)

- |   |                                                                                 |
|---|---------------------------------------------------------------------------------|
| 0 | De wereld in getallen (Malmberg, oude uitgave van 1981-1986)                    |
| 1 | De wereld in getallen (Malmberg, vernieuwde uitgave van 1991)                   |
| 2 | Operatoir rekenen (Uitgeverij Zwijsen, oude uitgave van 1980-1987)              |
| 3 | Operatoir rekenen (Uitgeverij Zwijsen, vernieuwde uitgave vanaf 1989)           |
| 4 | De rekenboom (Uitgeverij Wolters Noordhoff B.V., uitgave 1984)                  |
| 5 | Rekenwerk (Educatieve Partners Nederland, uitgave van 1985-1990)                |
| 6 | Naar zelfstandig rekenen (Uitgeverij Wolters Noordhoff B.V., uitgave 1968-1979) |
| 7 | Pluspunt (Uitgeverij Malmberg (vanaf 1991)                                      |
| 8 | Rekenen en wiskunde (Bekadidakt, uitgave vanaf 1983)                            |
| 9 | Geen van deze                                                                   |

- V-2 Iedere lesmethode kent een bepaalde planning. U kunt in uw onderwijs deze planning strikt volgen, of niet. U volgt de planning niet als u onderwerpen behandelt die niet in de lesmethode staan of als u onderwerpen niet behandelt die wel in de lesmethode staan. Een onderwerp is naar onze opvatting behandeld wanneer opgaven hierover maakbaar zijn, d.w.z.: **dat de lesstof onderwezen is en de leerlingen er oefeningen mee hebben gemaakt.**

In verrijkingstof (remelca e.a.) en in extra oefeningen wordt geen nieuwe lesstof behandeld en worden dus niet meegerekend!

Welk alternatief beschrijft uw situatie dit schooljaar het beste? (Omcirkel het cijfer voor het antwoord van uw keuze.)

- |   |                                                                                                                                                                              |
|---|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Ik behandel <b>alleen</b> die onderwerpen die in de methode staan.                                                                                                           |
| 2 | Ik behandel meer. Dat wil zeggen, dat ik onderwerpen die niet in de methode staan <b>wel</b> behandel.                                                                       |
| 3 | Ik behandel minder. Dat wil zeggen, dat ik onderwerpen die in de methode staan <b>niet</b> behandel.                                                                         |
| 4 | Ik behandel meer en minder. Dat wil zeggen, dat ik onderwerpen die niet in de methode staan <b>wel</b> behandel en onderwerpen die in de methode staan <b>niet</b> behandel. |

- V-3 Bij vraag twee gaf u aan of u dit schooljaar de planning van uw lesmethode al dan niet gevolgd hebt. Geldt uw antwoord ook voor andere jaren of is er dit jaar sprake van een uitzondering? (Omcirkel het cijfer voor het antwoord van uw keuze.)

- |   |                                                      |
|---|------------------------------------------------------|
| 1 | Mijn antwoord op vraag 2 geldt ook voor andere jaren |
| 2 | Nee, er is dit jaar sprake van een uitzondering      |

- V-4 Hoeveel minuten heeft het invullen van het onderdeel 'lesmethode' u gekost? (\*Vul het aantal in als een 2-cijferig getal, bijvoorbeeld 3 minuten als '03'.)

|  |  |
|--|--|
|  |  |
|--|--|



**Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3**

Aan het einde van het tweede onderdeel (Categorie) bij vraag 6 vragen wij u aan te geven hoeveel tijd u voor het invullen nodig heeft gehad.

**Onderdeel 2: Categorie**

De te toetsen lesstof in boekje E3 van het pakket Rekenen-Wiskunde 1 bestaat uit de categorieën: *tellen en ordenen*, *structuren*, *bewerkingen*, *rekendictee*, en *meten en tijd*. Deze categorieën bestaan vervolgens weer uit een aantal subcategorieën.

Wilt u per subcategorie aangeven of u vindt dat opgaven over deze subcategorieën voor uw leerlingen maakbaar zijn? **Met maakbaar wordt bedoeld dat de lesstof onderwezen is en dat de leerlingen er oefeningen mee gemaakt hebben.** Het is de bedoeling dat u uw oordeel alleen baseert op de omschrijving van de subcategorieën die in bijlage 1 gegeven is.

- V-5 Hieronder treft u de in toets E3 onderscheiden categorieën en subcategorieën aan. Hoe luidt **per subcategorie** uw oordeel over de **maakbaarheid**? (Omcirkel achter elke subcategorie het cijfer dat correspondeert met uw oordeel.)

| Subcategorieën:                        | Uw oordeel:   |              |
|----------------------------------------|---------------|--------------|
|                                        | Niet maakbaar | Wel maakbaar |
| <b>Tellen en Ordenen</b>               |               |              |
| 1. Structuur van de telrij             | 1             | 2            |
| 2. Resultatief en structurerend tellen | 1             | 2            |
| 3. Vergelijkingen en ordenen           | 1             | 2            |
| <b>Structuren</b>                      |               |              |
| 4. Splitsen                            | 1             | 2            |
| 5. Samenstellen                        | 1             | 2            |
| 6. Aanvullen                           | 1             | 2            |
| <b>Bewerkingen</b>                     |               |              |
| 7. Optellen                            | 1             | 2            |
| 8. Aftrekken                           | 1             | 2            |
| 9. Diversen                            | 1             | 2            |
| <b>Rekendictee</b>                     |               |              |
| 10. Optellen                           | 1             | 2            |
| 11. Aftrekken                          | 1             | 2            |
| 12. Splitsen                           | 1             | 2            |
| <b>Meten en Tijd</b>                   |               |              |
| 13. Meten                              | 1             | 2            |
| 14. Tijd                               | 1             | 2            |

- V-6 Hoeveel minuten heeft het invullen van het onderdeel 'categorie' u gekost\*? (\*Vul het aantal in als een 2-cijferig getal, bijvoorbeeld 3 minuten als '03'.)

|  |  |
|--|--|
|  |  |
|--|--|



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

Aan het einde van het derde onderdeel 3 (Opgave) bij vraag 8 vragen wij u aan te geven hoeveel tijd u voor het invullen nodig heeft gehad.

**Onderdeel 3: Opgave**

In dit onderdeel vragen wij u om de inhoud van de opgaven van toets E3 te beoordelen op de maakbaarheid ervan. **Met maakbaar wordt bedoeld dat de lesstof door u onderwezen is en dat de leerlingen er oefeningen mee hebben gemaakt.**

U moet zich **niet (!)** laten leiden door het feit dat u de ene opgave *moeilijker* voor de leerlingen vindt dan de andere. Uw oordeel bestaat alleen uit de afweging of u de lesstof, nodig voor het beantwoorden van de opgaven, onderwezen hebt. U moet zich **ook niet (!)** laten leiden door de *vaardigheid* van uw leerlingen. De ene leerling begrijpt meer dan de ander. Het gaat er niet om of uw leerlingen de opgaven kunnen maken, maar of uw leerlingen de lesstof gehad hebben.

V-7a Hieronder treft u de opgaven aan van DEEL 1 van het toetsboekje E3. Hoe luidt per **opgave** uw oordeel over de **maakbaarheid** ervan? (Omcirkel achter elke opgave het cijfer dat correspondeert met uw oordeel.)

| Toets E3, DEEL 1 | Uw oordeel:   |              |
|------------------|---------------|--------------|
|                  | Niet maakbaar | Wel maakbaar |
| Opgave 1         | 1             | 2            |
| Opgave 2         | 1             | 2            |
| Opgave 3         | 1             | 2            |
| Opgave 4         | 1             | 2            |
| Opgave 5         | 1             | 2            |
| Opgave 6         | 1             | 2            |
| Opgave 7         | 1             | 2            |
| Opgave 8         | 1             | 2            |
| Opgave 9         | 1             | 2            |
| Opgave 10        | 1             | 2            |
| Opgave 11        | 1             | 2            |
| Opgave 12        | 1             | 2            |
| Opgave 13        | 1             | 2            |
| Opgave 14        | 1             | 2            |
| Opgave 15        | 1             | 2            |
| Opgave 16        | 1             | 2            |
| Opgave 17        | 1             | 2            |
| Opgave 18        | 1             | 2            |
| Opgave 19        | 1             | 2            |
| Opgave 20        | 1             | 2            |
| Opgave 21        | 1             | 2            |
| Opgave 22        | 1             | 2            |
| Opgave 23        | 1             | 2            |
| Opgave 24        | 1             | 2            |



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

V-7b Hieronder treft u de opgaven aan van DEEL 2 van het toetsboekje E3. Hoe luidt per **opgave** uw oordeel over de **maakbaarheid** ervan? (Omcirkel achter elke opgave het cijfer dat correspondeert met uw oordeel.)

| Toets E3, <b>DEEL 2</b> | Uw oordeel:   |              |
|-------------------------|---------------|--------------|
|                         | Niet maakbaar | Wel maakbaar |
| Opgave 1                | 1             | 2            |
| Opgave 2                | 1             | 2            |
| Opgave 3                | 1             | 2            |
| Opgave 4                | 1             | 2            |
| Opgave 5                | 1             | 2            |
| Opgave 6                | 1             | 2            |
| Opgave 7                | 1             | 2            |
| Opgave 8                | 1             | 2            |
| Opgave 9                | 1             | 2            |
| Opgave 10               | 1             | 2            |
| Opgave 11               | 1             | 2            |
| Opgave 12               | 1             | 2            |
| Opgave 13               | 1             | 2            |
| Opgave 14               | 1             | 2            |
| Opgave 15               | 1             | 2            |
| Opgave 16               | 1             | 2            |
| Opgave 17               | 1             | 2            |
| Opgave 18               | 1             | 2            |
| Opgave 19               | 1             | 2            |
| Opgave 20               | 1             | 2            |
| Opgave 21               | 1             | 2            |
| Opgave 22               | 1             | 2            |
| Opgave 23               | 1             | 2            |
| Opgave 24               | 1             | 2            |
| Opgave 25               | 1             | 2            |
| Opgave 26               | 1             | 2            |
| Opgave 27               | 1             | 2            |
| Opgave 28               | 1             | 2            |
| Opgave 29               | 1             | 2            |

V-8 Hoeveel minuten heeft het invullen van onderdeel 3 (vraag 7a en 7b) u gekost? (\*Vul het aantal in als een 2-cijferig getal, bijvoorbeeld 3 minuten als '03'.)

|  |  |
|--|--|
|  |  |
|--|--|



## Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

### Hartelijk dank voor het invullen!

Wilt u de ingevulde vragenlijst en de ingevulde scoreformulieren zo spoedig mogelijk, doch uiterlijk voor 1 juli a.s. terugsturen naar het Cito? Vragenlijsten en scoreformulieren die later binnen komen, kunnen helaas niet meer worden verwerkt. Ons adres is:

Cito instituut voor toetsontwikkeling  
t.a.v.: Drs. H.A. Moelands  
Antwoordnummer 573  
6800VB Arnhem

Een postzegel plakken is niet nodig als u de bijgevoegde antwoordenveloppe gebruikt.

---

### Bijlage van de TCO-vragenlijst: Beschrijving van de subcategorieën

#### Tellen en ordenen

##### 1. Structuur van de telrij

Bij dit onderdeel wordt niet het 'opdreunen' van de telrij vanaf 1 getoetst, maar hoe goed de leerlingen flexibel kunnen tellen en of de telrij voor hen een geordende structuur is. Het gaat daarom om opgaven waarbij:

- moet worden voor- en achteruit geteld in sprongen van 1 en 2 vanaf een willekeurig getal
- moet worden bepaald welk getal voor of na een gegeven getal komt;
- moet worden geteld met sprongen van 1 of 2 op een getallenlijn die in eenheden is gestructureerd;
- contextproblemen moeten worden opgelost door gebruik te maken van het tellen en de structuur van de telrij.

Drie getallengebieden worden onderscheiden:

- de getallenreeks tot en met 10 (en de uitloop over 10) als basis voor de automatisering tot 10;
- de getallenreeks tot en met 20 als basis voor de automatisering tot 20;
- de voortzetting daarvan tot en met 100 als basis voor het rekenen tot 100.

##### 2. Resultatief en structurerend tellen

Leerlingen moeten eerst *resultatief* kunnen tellen, d.w.z. één voor één tellen. In de opgaven moeten leerlingen voorwerpen resultatief tellen.

*Structurerend* tellen is het tellen met sprongen van twee, vijf en tien om snel een aantal te bepalen. Het gaat om hoeveelheden die met de leerling één oogopslag herkent.

Bij de opgaven is het accent gelegd op het systematisch tellen van kleine en grote hoeveelheden (maximaal dertig) die soms wel en soms niet zijn geordend en die in enkele gevallen slechts gedeeltelijk zichtbaar zijn. Bij alle opgaven kunnen leerlingen gebruik maken van groepen van twee, vijf en tien als eenheid. Bekend is dat het moeilijk te bepalen valt of leerlingen daadwerkelijk gebruik maken van het tellen in groepjes.

##### 3. Vergelijken en ordenen

Bij *vergelijken en ordenen* moeten de leerlingen het volgende beheersen:

- hoeveelheden vergelijken op basis van tellen
- maten en hoeveelheden vergelijken op basis van gegeven aantallen (bijv. wat is groter: 12 of 21?)
- getallen vergelijken en ordenen met en zonder ondersteuning van een getallenlijn.

Ook naar getallen boven de twintig wordt hier gevraagd.

#### Structureren

De meeste opgaven zijn visueel verduidelijkt. De overige opgaven zijn formule-opgaven of opgaven met een afbeelding van een splitsing of een samenstelling (bijvoorbeeld middels een strook).

##### 4. Splitsen

Bij *splitsen* gaat het om opgaven waarbij getallen en hoeveelheden tussen 10 en 20 in tientallen en eenheden moeten worden gesplitst.



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

#### **5. Samenstellen**

Bij *samenstellen* gaat het om opgaven waarbij een getal tussen 10 en 20 moeten worden samengesteld door één tiental en een aantal eenheden samen te nemen.

#### **6. Aanvullen**

Bij *aanvullen* gaat het om opgaven waarbij getallen en hoeveelheden boven tussen 10 en 20 worden aangevuld op basis van de structuur van getallen.

#### **Bewerkingen**

Deze categorie wordt aan de hand van contexten en formule-opgaven getoetst. Niet alleen opgaven tot 10 maar ook opgaven uit het getallengebied tot 20 worden voorgelegd.

#### **7. Optellen**

Optellen wordt aan de hand van contexten en formule-opgaven getoetst. Optellen heeft de betekenis van samennemen, toevoegen en vergelijken. Het gaat om het optellen van getallen waarbij de uitkomst niet groter is dan 20. Opgaven zoals  $(5+4)$ ,  $(12+3)$  worden voorgelegd. Moeilijkere opgaven waar de tien gepasseerd wordt (zoals  $6+7$ ), worden niet bedoeld.

#### **8. Aftrekken**

Aftrekken wordt aan de hand van contexten en formule-opgaven getoetst. Aftrekken heeft de betekenis van eraf halen, aanvullen en verschil bepalen. Het gaat om opgaven waar het aftrektal niet groter is dan 20. Moeilijkere opgaven waar de tien gepasseerd wordt ( $15-7$ ) worden niet bedoeld.

#### **9. Diversen**

Het gaat om vermenigvuldig- en deelsituaties die op diverse manieren opgelost kunnen worden en om het bepalen wanneer welke uitgevoerd wordt.

#### **Rekendictee**

Het gaat hierbij om een auditieve aanbieding van opgaven die de leerlingen geautomatiseerd moeten hebben of in ieder geval vlot en vaardig moeten kunnen oplossen. De opgaven worden onder tijdsdruk afgenomen.

#### **10. Optellen**

Hierbij gaat het om het optellen van getallen waarbij de uitkomst niet groter is dan tien.

#### **11. Aftrekken**

Het gaat hier om het aftrekken van getallen waarbij het aftrektal niet groter is dan tien.

#### **12. Splitsen**

Het gaat hier om het splitsen van getallen tot en met tien.

#### **Meten en tijd**

##### **13. Meten**

Het gaat om een eerste oriëntatie op meten. De opgaven betreffen het vergelijken van lengtes, weegresultaten, oppervlaktes en het bepalen van het aantal blokjes van een bouwwerk.

##### **14. Tijd**

Het betreft het klokkijken in hele uren en het toepassen van die kennis, bijvoorbeeld bij het bepalen hoe lang iets duurt.



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

**Scoreformulier bij DEEL 1 van toets E3  
van het pakket Rekenen-Wiskunde 1**

Deel 1 van de toets E3 Rekenen-Wiskunde 1 bevat 24 opgaven.

Noteer met een "één" (1) als de betreffende leerling de opgave

correct beantwoord heeft. Indien de leerling de opgave niet correct beantwoord heeft, noteer dan een "nul" (0). Indien een antwoord van een leerling ontbreekt geeft u dit met een "streepje"(-) aan.

| Leerling | Opgaven van toets E3 Rekenen-Wiskunde 1 Deel 1 |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----------|------------------------------------------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|          | 1                                              | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 1        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 3        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 4        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 6        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 7        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 8        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 9        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 11       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 12       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 13       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 14       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 15       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 16       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 17       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 18       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 19       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 20       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 21       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 22       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 23       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 24       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 26       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 27       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 28       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 29       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 30       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 31       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 32       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 33       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 34       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 35       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

Vul hierna de antwoorden van uw leerlingen op deel 2 van de toets op het volgende scoreformulier in.



Bijlage 1 Testmateriaal: twee instructiepagina's, de TCO-vragenlijst en twee scoreformulieren voor de antwoorden van de leerlingen op toets E3

Scoreformulier bij DEEL 2 van toets E3  
van het pakket Rekenen-Wiskunde 1

Deel 2 van de toets E3 Rekenen-Wiskunde 1 bevat 29 opgaven.  
Noteer met een "één" (1) als de betreffende leerling de opgave correct beantwoord heeft. Indien de leerling de opgave niet correct beantwoord heeft, noteer dan een "nul" (0). Indien een antwoord van een leerling ontbreekt geeft u dit met een "streepje"(-) aan.

| Leerling | Opgaven van toets E3 Rekenen-Wiskunde 1 Deel 2 |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|----------|------------------------------------------------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
|          | 1                                              | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
| 1        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 3        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 4        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 6        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 7        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 8        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 9        |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 10       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 11       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 12       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 13       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 14       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 15       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 16       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 17       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 18       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 19       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 20       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 21       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 22       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 23       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 24       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 25       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 26       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 27       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 28       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 29       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 30       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 31       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 32       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 33       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 34       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 35       |                                                |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |

U bent nu klaar! Wilt u de ingevulde vragenlijst en de ingevulde scoreformulieren in de bijgevoegde enveloppe uiterlijk 1 juli naar ons toesturen? Na de zomer zullen wij u de resultaten van dit onderzoek toesturen. Nogmaals: Hartelijk dank voor uw medewerking!



Bijlage 2 De proportie maakbare vragen (p-waarde) van de subcategorie- en de itemmethode bij de zes lesmethoden

**A. De proportie maakbare vragen per lesmethode van de subcategoriemethode**

| <b>Subcategorie</b>                 | <b>Wereld in<br/>getallen<br/>(oud)</b> | <b>Wereld in<br/>getallen<br/>(nieuw)</b> | <b>Operatoir<br/>rekenen<br/>(oud)</b> | <b>Operatoir<br/>rekenen<br/>(nieuw)</b> | <b>Pluspunt</b> | <b>Rekenen en<br/>wiskunde</b> |
|-------------------------------------|-----------------------------------------|-------------------------------------------|----------------------------------------|------------------------------------------|-----------------|--------------------------------|
| Structuur van de telrij             | 1.000                                   | 1.000                                     | 1.000                                  | 1.000                                    | .957            | 1.000                          |
| Resultatief en structurerend tellen | .895                                    | .957                                      | 1.000                                  | 1.000                                    | 1.000           | .941                           |
| Vergelijkingen en ordenen           | .895                                    | 1.000                                     | .900                                   | 1.000                                    | .957            | .765                           |
| Splitsen                            | .737                                    | .652                                      | .800                                   | 1.000                                    | .891            | .794                           |
| Samenstellen                        | .842                                    | .848                                      | .800                                   | .857                                     | 1.000           | .882                           |
| Aanvullen                           | .684                                    | .674                                      | .900                                   | 1.000                                    | .913            | .912                           |
| Optellen                            | .947                                    | .891                                      | 1.000                                  | 1.000                                    | .978            | .912                           |
| Aftrekken                           | .895                                    | .870                                      | 1.000                                  | 1.000                                    | .957            | .853                           |
| Diversen                            | .842                                    | .696                                      | .900                                   | 1.000                                    | .848            | .882                           |
| Optellen                            | .947                                    | 1.000                                     | 1.000                                  | 1.000                                    | .935            | .912                           |
| Aftrekken                           | .947                                    | 1.000                                     | 1.000                                  | 1.000                                    | .935            | .794                           |
| Splitsen                            | .895                                    | .891                                      | .900                                   | .714                                     | .826            | .647                           |
| Meten                               | 1.000                                   | .935                                      | .800                                   | .714                                     | .913            | .441                           |
| Tijd                                | .526                                    | .717                                      | 1.000                                  | 1.000                                    | .565            | .618                           |



**B. De proportie maakbare vragen per lesmethode van de itemmethode**

| Deel | Item | Wereld in<br>getallen<br>(oud) | Wereld in<br>getallen<br>(nieuw) | Operatoir<br>rekenen<br>(oud) | Operatoir<br>rekenen<br>(nieuw) | Pluspunt | Rekenen<br>en<br>wiskunde |
|------|------|--------------------------------|----------------------------------|-------------------------------|---------------------------------|----------|---------------------------|
| A    | 01   | .947                           | .979                             | 1.000                         | 1.000                           | .957     | .912                      |
|      | 02   | 1.000                          | .979                             | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 03   | .895                           | 1.000                            | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 04   | 1.000                          | .979                             | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 05   | 1.000                          | .979                             | 1.000                         | .714                            | .870     | .824                      |
|      | 06   | .368                           | .362                             | .300                          | .571                            | .457     | .412                      |
|      | 07   | 1.000                          | .957                             | 1.000                         | 1.000                           | 1.000    | .824                      |
|      | 08   | .684                           | .723                             | 1.000                         | 1.000                           | .957     | .824                      |
|      | 09   | .842                           | .830                             | 1.000                         | 1.000                           | .978     | .971                      |
|      | 10   | .947                           | .915                             | 1.000                         | 1.000                           | .978     | .794                      |
|      | 11   | .474                           | .617                             | 1.000                         | 1.000                           | .913     | .824                      |
|      | 12   | .895                           | .872                             | 1.000                         | .571                            | .826     | .853                      |
|      | 13   | .632                           | .617                             | 1.000                         | 1.000                           | .957     | .971                      |
|      | 14   | .684                           | .830                             | 1.000                         | 1.000                           | .978     | .971                      |
|      | 15   | .421                           | .702                             | 1.000                         | .857                            | .870     | .853                      |
|      | 16   | .526                           | .553                             | 1.000                         | 1.000                           | .848     | .794                      |
|      | 17   | .789                           | .936                             | .900                          | 1.000                           | .913     | .706                      |
|      | 18   | .684                           | .915                             | .700                          | .857                            | .848     | .765                      |
|      | 19   | .947                           | .979                             | .900                          | .714                            | .913     | .912                      |
|      | 20   | .737                           | .681                             | .700                          | .857                            | .565     | .500                      |
|      | 21   | 1.000                          | 1.000                            | .900                          | .857                            | .957     | .912                      |
|      | 22   | .947                           | .957                             | .400                          | .571                            | 1.000    | .559                      |
|      | 23   | .947                           | .979                             | 1.000                         | 1.000                           | .717     | .294                      |
|      | 24   | .684                           | .872                             | .900                          | 1.000                           | .543     | .294                      |
| B    | 01   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | .971                      |
|      | 02   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | .971                      |
|      | 03   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 04   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 05   | .947                           | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 06   | .947                           | 1.000                            | 1.000                         | 1.000                           | .935     | .882                      |
|      | 07   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .941                      |
|      | 08   | .947                           | .936                             | .900                          | .857                            | .783     | .676                      |
|      | 09   | .842                           | .830                             | 1.000                         | 1.000                           | .739     | .647                      |
|      | 10   | .895                           | .830                             | 1.000                         | 1.000                           | .783     | .647                      |
|      | 11   | 1.000                          | .830                             | .900                          | .857                            | .978     | .882                      |
|      | 12   | 1.000                          | .979                             | 1.000                         | .857                            | 1.000    | .912                      |
|      | 13   | 1.000                          | .979                             | .700                          | .857                            | .978     | .941                      |
|      | 14   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | 1.000                     |
|      | 15   | .737                           | .638                             | .900                          | 1.000                           | .935     | .941                      |
|      | 16   | .579                           | .574                             | 1.000                         | 1.000                           | .848     | .853                      |
|      | 17   | .632                           | .723                             | .700                          | .857                            | .870     | .824                      |
|      | 18   | 1.000                          | .894                             | 1.000                         | .857                            | .935     | .882                      |
|      | 19   | .579                           | .660                             | 1.000                         | .857                            | .935     | .853                      |
|      | 20   | .947                           | 1.000                            | 1.000                         | 1.000                           | 1.000    | 1.000                     |
|      | 21   | .684                           | .638                             | 1.000                         | 1.000                           | .957     | .941                      |
|      | 22   | .421                           | .404                             | .800                          | .857                            | .739     | .559                      |
|      | 23   | .316                           | .404                             | .900                          | 1.000                           | .826     | .618                      |
|      | 24   | .947                           | .915                             | 1.000                         | .857                            | .978     | .941                      |
|      | 25   | .895                           | .851                             | .800                          | .857                            | .826     | .676                      |
|      | 26   | .947                           | .979                             | .400                          | .857                            | .913     | .735                      |
|      | 27   | 1.000                          | .957                             | .700                          | .714                            | 1.000    | .559                      |
|      | 28   | .579                           | .723                             | .600                          | .429                            | .674     | .471                      |
|      | 29   | .421                           | .511                             | .800                          | .714                            | .522     | .147                      |

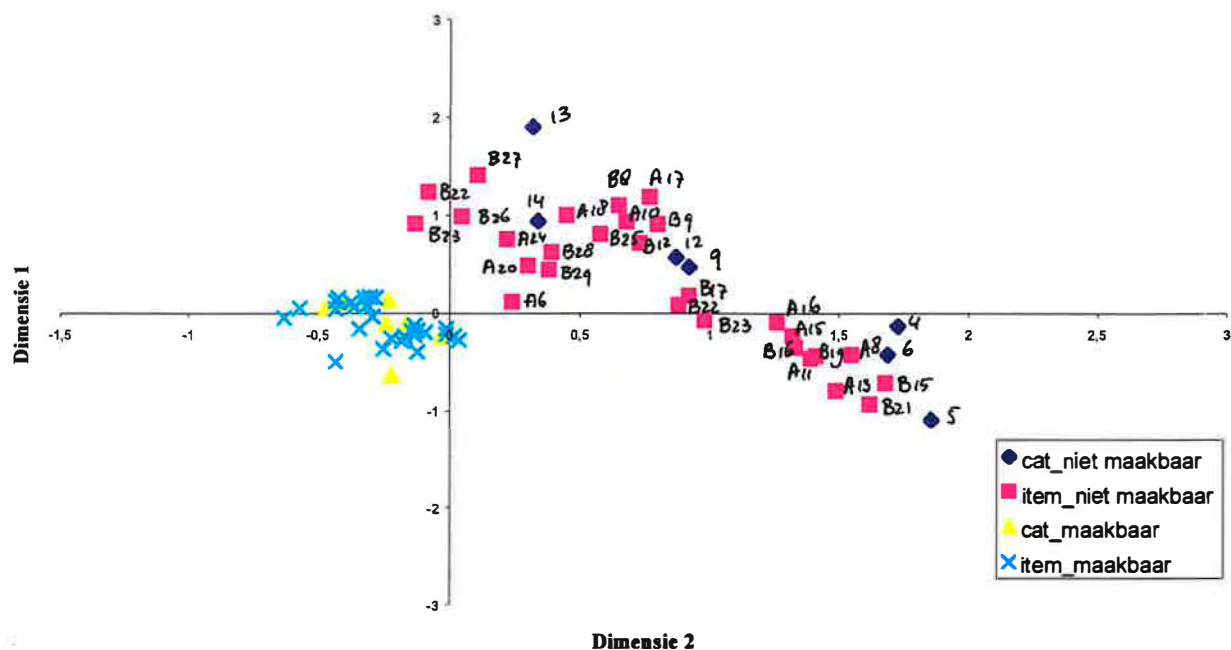


Bijlage 3: Overzicht van maakbaarheidsoordelen in p-waarden bij zes lesmethoden

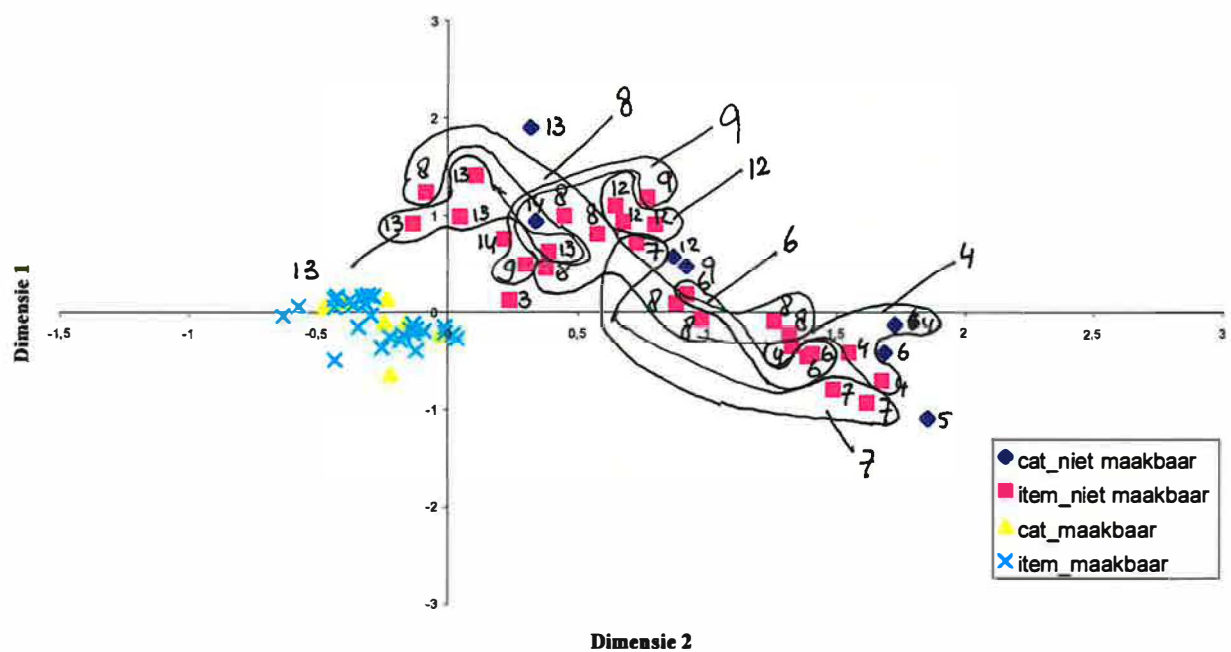
| Deel | Item | Wereld in<br>getallen<br>(oud) | Wereld in<br>getallen<br>(nieuw) | Operatoir<br>rekenen<br>(oud) | Operatoir<br>rekenen<br>(nieuw) | Pluspunt | Rekenen<br>en<br>wiskunde |
|------|------|--------------------------------|----------------------------------|-------------------------------|---------------------------------|----------|---------------------------|
| A    | 01   | .947                           | .979                             | 1.000                         | 1.000                           | .957     | .912                      |
|      | 02   | 1.000                          | .979                             | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 03   | .895                           | 1.000                            | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 04   | 1.000                          | .979                             | 1.000                         | 1.000                           | 1.000    | .971                      |
|      | 05   | 1.000                          | .979                             | 1.000                         | .714                            | .870     | .824                      |
|      | 06   | .368                           | .362                             | .300                          | .571                            | .457     | .412                      |
|      | 07   | 1.000                          | .957                             | 1.000                         | 1.000                           | 1.000    | .824                      |
|      | 08   | .684                           | .723                             | 1.000                         | 1.000                           | .957     | .824                      |
|      | 09   | .842                           | .830                             | 1.000                         | 1.000                           | .978     | .971                      |
|      | 10   | .947                           | .915                             | 1.000                         | 1.000                           | .978     | .794                      |
|      | 11   | .474                           | .617                             | 1.000                         | 1.000                           | .913     | .824                      |
|      | 12   | .895                           | .872                             | 1.000                         | .571                            | .826     | .853                      |
|      | 13   | .632                           | .617                             | 1.000                         | 1.000                           | .957     | .971                      |
|      | 14   | .684                           | .830                             | 1.000                         | 1.000                           | .978     | .971                      |
|      | 15   | .421                           | .702                             | 1.000                         | .857                            | .870     | .853                      |
|      | 16   | .526                           | .553                             | 1.000                         | 1.000                           | .848     | .794                      |
|      | 17   | .789                           | .936                             | .900                          | 1.000                           | .913     | .706                      |
|      | 18   | .684                           | .915                             | .700                          | .857                            | .848     | .765                      |
|      | 19   | .947                           | .979                             | .900                          | .714                            | .913     | .912                      |
|      | 20   | .737                           | .681                             | .700                          | .857                            | .565     | .500                      |
|      | 21   | 1.000                          | 1.000                            | .900                          | .857                            | .957     | .912                      |
|      | 22   | .947                           | .957                             | .400                          | .571                            | 1.000    | .559                      |
|      | 23   | .947                           | .979                             | 1.000                         | 1.000                           | .717     | .294                      |
|      | 24   | .684                           | .872                             | .900                          | 1.000                           | .543     | .294                      |
| B    | 01   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | .971                      |
|      | 02   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | .971                      |
|      | 03   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 04   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 05   | .947                           | 1.000                            | 1.000                         | 1.000                           | .957     | .971                      |
|      | 06   | .947                           | 1.000                            | 1.000                         | 1.000                           | .935     | .882                      |
|      | 07   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .957     | .941                      |
|      | 08   | .947                           | .936                             | .900                          | .857                            | .783     | .676                      |
|      | 09   | .842                           | .830                             | 1.000                         | 1.000                           | .739     | .647                      |
|      | 10   | .895                           | .830                             | 1.000                         | 1.000                           | .783     | .647                      |
|      | 11   | 1.000                          | .830                             | .900                          | .857                            | .978     | .882                      |
|      | 12   | 1.000                          | .979                             | 1.000                         | .857                            | 1.000    | .912                      |
|      | 13   | 1.000                          | .979                             | .700                          | .857                            | .978     | .941                      |
|      | 14   | 1.000                          | 1.000                            | 1.000                         | 1.000                           | .978     | 1.000                     |
|      | 15   | .737                           | .638                             | .900                          | 1.000                           | .935     | .941                      |
|      | 16   | .579                           | .574                             | 1.000                         | 1.000                           | .848     | .853                      |
|      | 17   | .632                           | .723                             | .700                          | .857                            | .870     | .824                      |
|      | 18   | 1.000                          | .894                             | 1.000                         | .857                            | .935     | .882                      |
|      | 19   | .579                           | .660                             | 1.000                         | .857                            | .935     | .853                      |
|      | 20   | .947                           | 1.000                            | 1.000                         | 1.000                           | 1.000    | 1.000                     |
|      | 21   | .684                           | .638                             | 1.000                         | 1.000                           | .957     | .941                      |
|      | 22   | .421                           | .404                             | .800                          | .857                            | .739     | .559                      |
|      | 23   | .316                           | .404                             | .900                          | 1.000                           | .826     | .618                      |
|      | 24   | .947                           | .915                             | 1.000                         | .857                            | .978     | .941                      |
|      | 25   | .895                           | .851                             | .800                          | .857                            | .826     | .676                      |
|      | 26   | .947                           | .979                             | .400                          | .857                            | .913     | .735                      |
|      | 27   | 1.000                          | .957                             | .700                          | .714                            | 1.000    | .559                      |
|      | 28   | .579                           | .723                             | .600                          | .429                            | .674     | .471                      |
|      | 29   | .421                           | .511                             | .800                          | .714                            | .522     | .147                      |



Figuur A Spreidingsdiagram van de subcategorieën en de items

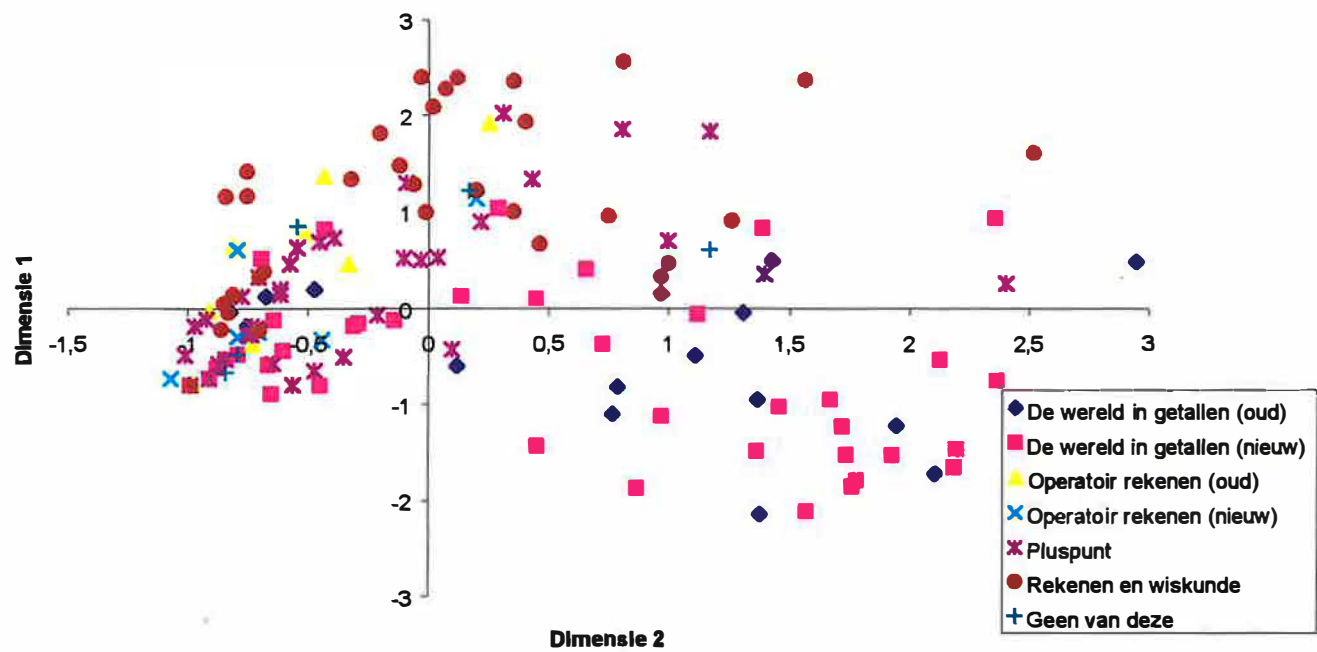


Figuur B Spreidingsdiagram van de subcategorieën en de items, met clustering





Figuur C Spreidingsdiagram met de leerkrachten





Bijlage 4 De proportie correcte antwoorden op toets E3 bij de zes lesmethoden

| Deel | Item | Wereld in<br>getallen (o)<br>N= 319 | Wereld in<br>getallen (n)<br>N= 988 | Operator<br>rekenen (o)<br>N= 242 | Operator<br>rekenen (n)<br>N= 105 | Pluspunt<br>N=809 | Rekenen en<br>wiskunde<br>N= 620 |
|------|------|-------------------------------------|-------------------------------------|-----------------------------------|-----------------------------------|-------------------|----------------------------------|
|      |      | p-waarde                            | p-waarde                            | p-waarde                          | p-waarde                          | p-waarde          | p-waarde                         |
| A    | 01   | .787                                | .831                                | .752                              | .800                              | .828              | .818                             |
|      | 02   | .956                                | .957                                | .934                              | .971                              | .947              | .939                             |
|      | 03   | .912                                | .906                                | .884                              | .914                              | .928              | .905                             |
|      | 04   | .947                                | .937                                | .905                              | .914                              | .972              | .931                             |
|      | 05   | .850                                | .911                                | .872                              | .876                              | .815              | .795                             |
|      | 06   | .495                                | .477                                | .413                              | .667                              | .576              | .518                             |
|      | 07   | .881                                | .888                                | .752                              | .886                              | .837              | .869                             |
|      | 08   | .843                                | .846                                | .789                              | .895                              | .867              | .805                             |
|      | 09   | .871                                | .900                                | .930                              | .962                              | .937              | .890                             |
|      | 10   | .940                                | .925                                | .921                              | .962                              | .902              | .835                             |
|      | 11   | .724                                | .742                                | .773                              | .838                              | .770              | .739                             |
|      | 12   | .915                                | .881                                | .831                              | .943                              | .847              | .874                             |
|      | 13   | .755                                | .791                                | .826                              | .819                              | .848              | .810                             |
|      | 14   | .940                                | .945                                | .938                              | .971                              | .969              | .963                             |
|      | 15   | .787                                | .765                                | .760                              | .819                              | .797              | .731                             |
|      | 16   | .630                                | .607                                | .620                              | .790                              | .667              | .577                             |
|      | 17   | .787                                | .826                                | .711                              | .819                              | .849              | .774                             |
|      | 18   | .787                                | .844                                | .719                              | .867                              | .836              | .818                             |
|      | 19   | .878                                | .909                                | .785                              | .867                              | .775              | .803                             |
|      | 20   | .621                                | .668                                | .607                              | .733                              | .622              | .602                             |
|      | 21   | .862                                | .860                                | .938                              | .800                              | .849              | .847                             |
|      | 22   | .777                                | .827                                | .744                              | .819                              | .883              | .755                             |
|      | 23   | .918                                | .941                                | .967                              | .971                              | .873              | .803                             |
|      | 24   | .708                                | .749                                | .740                              | .905                              | .713              | .719                             |
| B    | 1    | .981                                | .988                                | .988                              | .990                              | .977              | .984                             |
|      | 2    | .966                                | .986                                | .975                              | .981                              | .967              | .976                             |
|      | 3    | .950                                | .942                                | .950                              | .952                              | .941              | .939                             |
|      | 4    | .931                                | .936                                | .963                              | .962                              | .930              | .916                             |
|      | 5    | .962                                | .971                                | .979                              | .971                              | .921              | .926                             |
|      | 6    | .875                                | .848                                | .884                              | .867                              | .831              | .819                             |
|      | 7    | .925                                | .907                                | .946                              | .924                              | .902              | .905                             |
|      | 8    | .762                                | .713                                | .715                              | .733                              | .737              | .666                             |
|      | 9    | .755                                | .659                                | .764                              | .610                              | .706              | .666                             |
|      | 10   | .799                                | .736                                | .793                              | .781                              | .732              | .735                             |
|      | 11   | .715                                | .649                                | .756                              | .819                              | .700              | .631                             |
|      | 12   | .906                                | .906                                | .810                              | .762                              | .880              | .871                             |
|      | 13   | .969                                | .967                                | .913                              | .933                              | .959              | .947                             |
|      | 14   | .875                                | .902                                | .880                              | .876                              | .890              | .900                             |
|      | 15   | .796                                | .780                                | .707                              | .848                              | .786              | .766                             |
|      | 16   | .687                                | .690                                | .744                              | .771                              | .718              | .713                             |
|      | 17   | .781                                | .829                                | .839                              | .867                              | .848              | .832                             |
|      | 18   | .762                                | .850                                | .798                              | .838                              | .818              | .787                             |
|      | 19   | .840                                | .890                                | .868                              | .914                              | .899              | .865                             |
|      | 20   | .953                                | .950                                | .942                              | .971                              | .954              | .937                             |
|      | 21   | .771                                | .784                                | .798                              | .886                              | .811              | .815                             |
|      | 22   | .505                                | .621                                | .599                              | .648                              | .661              | .582                             |
|      | 23   | .517                                | .586                                | .599                              | .581                              | .653              | .595                             |
|      | 24   | .875                                | .873                                | .855                              | .924                              | .831              | .824                             |
|      | 25   | .759                                | .745                                | .665                              | .724                              | .728              | .677                             |
|      | 26   | .843                                | .856                                | .773                              | .781                              | .852              | .745                             |
|      | 27   | .796                                | .822                                | .562                              | .543                              | .902              | .765                             |
|      | 28   | .680                                | .745                                | .562                              | .743                              | .737              | .656                             |
|      | 29   | .367                                | .414                                | .413                              | .448                              | .483              | .400                             |



Bijlage 5 Parameters voor de subcategorieën van de TCO-vragenlijst

| A. Itemparameterschattingen van de subcategorieën bij het OPLM |       |           | B. Persoonsparameterschattingen en schaal-scores van de subcategorieën bij het OPLM |        |             |
|----------------------------------------------------------------|-------|-----------|-------------------------------------------------------------------------------------|--------|-------------|
| Subcategorie                                                   | $a_i$ | $\beta_i$ | Toetsscore                                                                          | Theta  | Schaalscore |
| 1                                                              | 2     | -.624     | 0                                                                                   | -2,186 | 0           |
| 2                                                              | 2     | -.624     | 1                                                                                   | -1,465 | 17          |
| 3                                                              | 2     | -.296     | 2                                                                                   | -1,097 | 25          |
| 4                                                              | 4     | .814      | 3                                                                                   | -,831  | 31          |
| 5                                                              | 2     | .168      | 4                                                                                   | -,607  | 36          |
| 6                                                              | 3     | .717      | 5                                                                                   | -,403  | 41          |
| 7                                                              | 2     | -.439     | 6                                                                                   | -,205  | 45          |
| 8                                                              | 2     | -.031     | 7                                                                                   | -,003  | 50          |
| 9                                                              | 2     | .625      | 8                                                                                   | ,206   | 55          |
| 10                                                             | 2     | -.524     | 9                                                                                   | ,420   | 60          |
| 11                                                             | 2     | -.363     | 10                                                                                  | ,626   | 64          |
| 12                                                             | 1     | -.126     | 11                                                                                  | ,828   | 69          |
| 13                                                             | 1     | -.515     | 12                                                                                  | 1,053  | 74          |
| 14                                                             | 1     | 1.217     | 13                                                                                  | 1,371  | 81          |
|                                                                |       |           | 14                                                                                  | 2,182  | 100         |



Bijlage 6 Parameters voor de itemmethode van de TCO-vragenlijst

| A. Itemparameterschattingen van de items |      |       |           | B. Persoonsparameterschattingen en schaalscores van de items |        |             |
|------------------------------------------|------|-------|-----------|--------------------------------------------------------------|--------|-------------|
| Deel                                     | Item | $a_i$ | $\beta_i$ | Toetsscore                                                   | Theta  | Schaalscore |
| A                                        | 1    | 2     | -.782     | 0                                                            | -2.408 | 0           |
|                                          | 2    | 4     | -.446     | 1                                                            | -1.829 | 12          |
|                                          | 3    | 3     | -.625     | 2                                                            | -1.556 | 18          |
|                                          | 4    | 4     | -.446     | 3                                                            | -1.375 | 22          |
|                                          | 5    | 2     | -.361     | 4                                                            | -1.237 | 25          |
|                                          | 6    | 1     | 1.454     | 5                                                            | -1.124 | 27          |
|                                          | 7    | 4     | -.073     | 6                                                            | -1.028 | 29          |
|                                          | 8    | 6     | .408      | 7                                                            | -.942  | 31          |
|                                          | 9    | 5     | .159      | 8                                                            | -.865  | 32          |
|                                          | 10   | 3     | -.035     | 9                                                            | -.795  | 35          |
|                                          | 11   | 6     | .522      | 10                                                           | -.730  | 35          |
|                                          | 12   | 3     | .181      | 11                                                           | -.668  | 36          |
|                                          | 13   | 5     | .379      | 12                                                           | -.611  | 38          |
|                                          | 14   | 4     | .136      | 13                                                           | -.555  | 39          |
|                                          | 15   | 5     | .476      | 14                                                           | -.502  | 40          |
|                                          | 16   | 6     | .557      | 15                                                           | -.450  | 41          |
|                                          | 17   | 3     | .163      | 16                                                           | -.400  | 42          |
|                                          | 18   | 2     | .067      | 17                                                           | -.351  | 43          |
|                                          | 19   | 2     | -.443     | 18                                                           | -.304  | 44          |
|                                          | 20   | 2     | .649      | 19                                                           | -.256  | 46          |
|                                          | 21   | 2     | -.647     | 20                                                           | -.210  | 46          |
|                                          | 22   | 1     | -.646     | 21                                                           | -.164  | 47          |
|                                          | 23   | 1     | -.223     | 22                                                           | -.119  | 48          |
|                                          | 24   | 2     | .603      | 23                                                           | -.075  | 49          |
| B                                        | 1    | 3     | -.625     | 24                                                           | -.031  | 50          |
|                                          | 2    | 3     | -.625     | 25                                                           | .011   | 51          |
|                                          | 3    | 3     | -.524     | 26                                                           | .053   | 52          |
|                                          | 4    | 3     | -.524     | 27                                                           | .094   | 52          |
|                                          | 5    | 2     | -.960     | 28                                                           | .134   | 53          |
|                                          | 6    | 3     | -.232     | 29                                                           | .173   | 55          |
|                                          | 7    | 3     | -.446     | 30                                                           | .211   | 55          |
|                                          | 8    | 3     | .250      | 31                                                           | .248   | 56          |
|                                          | 9    | 3     | .384      | 32                                                           | .283   | 56          |
|                                          | 10   | 3     | .342      | 33                                                           | .318   | 57          |
|                                          | 11   | 2     | -.361     | 34                                                           | .353   | 58          |
|                                          | 12   | 4     | -.203     | 35                                                           | .387   | 59          |
|                                          | 13   | 2     | -.711     | 36                                                           | .420   | 59          |
|                                          | 14   | 3     | -.999     | 37                                                           | .454   | 60          |
|                                          | 15   | 7     | .426      | 38                                                           | .488   | 61          |
|                                          | 16   | 6     | .540      | 39                                                           | .523   | 62          |
|                                          | 17   | 4     | .475      | 40                                                           | .559   | 62          |
|                                          | 18   | 3     | -.063     | 41                                                           | .596   | 63          |
|                                          | 19   | 6     | .476      | 42                                                           | .635   | 64          |
|                                          | 20   | 3     | -.999     | 43                                                           | .677   | 65          |
|                                          | 21   | 5     | .379      | 44                                                           | .722   | 66          |
|                                          | 22   | 5     | .808      | 45                                                           | .772   | 67          |
|                                          | 23   | 5     | .736      | 46                                                           | .828   | 68          |
|                                          | 24   | 3     | -.193     | 47                                                           | .892   | 69          |
|                                          | 25   | 2     | .111      | 48                                                           | .967   | 71          |
|                                          | 26   | 2     | -.105     | 49                                                           | 1.060  | 75          |
|                                          | 27   | 2     | -.078     | 50                                                           | 1.182  | 75          |
|                                          | 28   | 2     | .694      | 51                                                           | 1.359  | 79          |
|                                          | 29   | 3     | 1.000     | 52                                                           | 1.659  | 85          |
|                                          |      |       |           | 53                                                           | 2.366  | 100         |



Bijlage 7 Parameters van toets E3

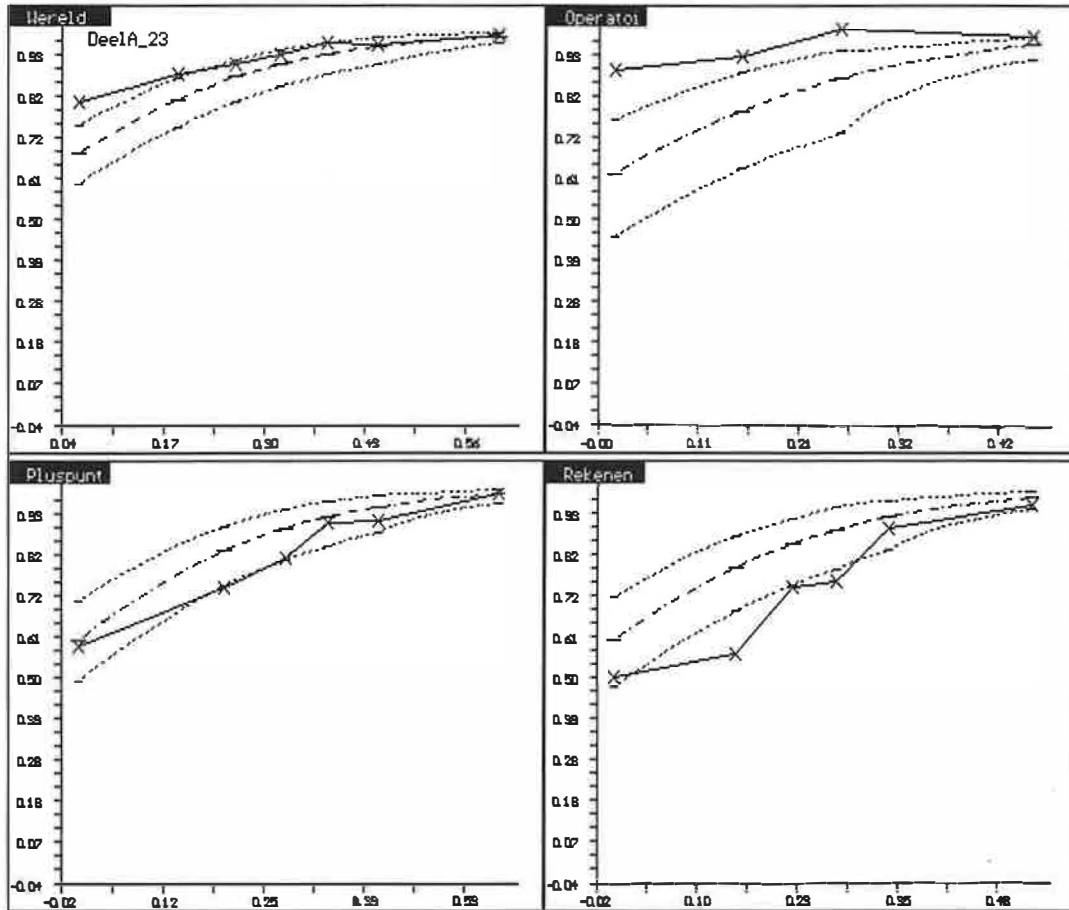
| A. Itemparameterschattingen |      |       |           | B. Persoonsparameterschattingen |        |
|-----------------------------|------|-------|-----------|---------------------------------|--------|
| Deel                        | Item | $a_i$ | $\beta_i$ | Toetsscore                      | Theta  |
| A                           | 1    | 4     | -.046     | 0                               | -1.064 |
|                             | 2    | 5     | -.298     | 1                               | -.808  |
|                             | 3    | 4     | -.272     | 2                               | -.688  |
|                             | 4    | 5     | -.266     | 3                               | -.608  |
|                             | 5    | 5     | -.036     | 4                               | -.546  |
|                             | 6    | 4     | .362      | 5                               | -.496  |
|                             | 7    | 6     | .001      | 6                               | -.453  |
|                             | 8    | 8     | .083      | 7                               | -.416  |
|                             | 9    | 6     | -.092     | 8                               | -.382  |
|                             | 10   | 8     | -.011     | 9                               | -.350  |
|                             | 11   | 7     | .158      | 10                              | -.321  |
|                             | 12   | 6     | -.023     | 11                              | -.294  |
|                             | 13   | 6     | .076      | 12                              | -.269  |
|                             | 14   | 6     | -.238     | 13                              | -.244  |
|                             | 15   | 5     | .084      | 14                              | -.221  |
|                             | 16   | 6     | .264      | 15                              | -.199  |
|                             | 17   | 7     | .102      | 16                              | -.177  |
|                             | 18   | 7     | .084      | 17                              | -.157  |
|                             | 19   | 6     | .033      | 18                              | -.137  |
|                             | 20   | 5     | .241      | 19                              | -.117  |
|                             | 21   | 3     | -.260     | 20                              | -.098  |
|                             | 22   | 4     | -.040     | 21                              | -.080  |
|                             | 23   | 6     | -.071     | 22                              | -.062  |
|                             | 24   | 5     | .131      | 23                              | -.044  |
| B                           | 1    | 7     | -.333     | 24                              | -.027  |
|                             | 2    | 7     | -.263     | 25                              | -.010  |
|                             | 3    | 5     | -.266     | 26                              | .007   |
|                             | 4    | 5     | -.225     | 27                              | .023   |
|                             | 5    | 5     | -.291     | 28                              | .039   |
|                             | 6    | 4     | -.090     | 29                              | .056   |
|                             | 7    | 4     | -.264     | 30                              | .072   |
|                             | 8    | 5     | .160      | 31                              | .088   |
|                             | 9    | 6     | .203      | 32                              | .105   |
|                             | 10   | 6     | .147      | 33                              | .121   |
|                             | 11   | 3     | .102      | 34                              | .138   |
|                             | 12   | 4     | -.171     | 35                              | .155   |
|                             | 13   | 6     | -.235     | 36                              | .173   |
|                             | 14   | 4     | -.212     | 37                              | .190   |
|                             | 15   | 8     | .154      | 38                              | .209   |
|                             | 16   | 7     | .201      | 39                              | .228   |
|                             | 17   | 8     | .092      | 40                              | .248   |
|                             | 18   | 6     | .064      | 41                              | .269   |
|                             | 19   | 10    | .064      | 42                              | .291   |
|                             | 20   | 5     | -.296     | 43                              | .314   |
|                             | 21   | 5     | .047      | 44                              | .339   |
|                             | 22   | 6     | .282      | 45                              | .367   |
|                             | 23   | 5     | .281      | 46                              | .397   |
|                             | 24   | 5     | -.036     | 47                              | .432   |
|                             | 25   | 8     | .204      | 48                              | .471   |
|                             | 26   | 6     | .056      | 49                              | .518   |
|                             | 27   | 4     | -.015     | 50                              | .576   |
|                             | 28   | 7     | .210      | 51                              | .653   |
|                             | 29   | 4     | .464      | 52                              | .770   |
|                             |      |       |           | 53                              | 1.022  |

(Geometric mean of discrimination indices = 5.463)



## Bijlage 8 Itemkarakteristieke curven van een lesmethode-onzuiver item

Vraagonzuiverheid bij item 23 (A\_23)



Wereld= Wereld in getallen, Operator= Operator rekenen,  
Pluspunt= Pluspunt, en Rekenen= Rekenen en wiskunde





