Taaltoetsen voor taalzwakke leerlingen in het primair onderwijs

E. van Schooten[1]
E. Smeets[2]
G. Driessen[2]

[1] SCO-Kohnstamm Instituut,
Universiteit van Amsterdam
[2] ITS, Radboud Universiteit Nijmegen

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Schooten, E. van, Smeets, E., Driessen, G.

Taaltoetsen voor taalzwakke leerlingen in het primair onderwijs. E. van Schooten, Amsterdam : SCO-Kohnstamm Instituut van de Faculteit der Maatschappij- en Gedragswetenschappen, Universiteit van Amsterdam (SCO-rapport nr. 771 – projectnummer 40077), E. Smeets & G. Driessen, Nijmegen: ITS van de Radboud Universiteit Nijmegen.

ISBN 978-90-6813-826-9

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of op enige manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher.

Uitgave en verspreiding:
SCO-Kohnstamm Instituut
Nieuwe Prinsengracht 130, Postbus 94208, 1090 GE Amsterdam
tel.: 020-525 1201
http://www.sco-kohnstamminstituut.uva.nl
Copyright © SCO-Kohnstamm Instituut, 2007

Inhoudsopgave

Voorwoord

Managementsamenvatting

1 Aanleiding 1
1.1 Probleemstelling en onderzoeksvragen 1

2 Opzet van het onderzoek 5
2.1 Keuze van de te evalueren taaltoetsen 5
2.2 Afnamedesign 9
2.3 Steekproefkader 10
2.4 Afnameprocedures 13

3 Analyses 15
3.1 Kwantitatieve analyses 15
3.2 Kwalitatieve analyses 19

4 Resultaten 21
4.1 Steekproeftrekking en respons 21
4.2 Afname van de toetsen 25
4.3 Aantallen afgenomen toetsen 27
4.4 Achtergrondkenmerken van de leerlingen 29
4.5 De betrouwbaarheid van de (sub)toetsen 32
4.6 Validiteit van de (sub)toetsen 41
4.7 Selectie van leerlingen met de laagste taalvaardigheid 78
4.8 Verschil in afnamekwaliteit testleiders en leerkrachten 142
4.9 Afnamegemak 147
4.10 Oordelen schoolleiders over toetsen van jonge leerlingen 158
4.11 Praktische aspecten van de toetsing van taalvaardigheid 163

5 Eindconclusies **171**

6 Samenvatting **181**

Literatuur **195**

Appendix

Voorwoord

In het voorliggende rapport wordt verslag gedaan van een onderzoek naar de kwaliteit van taaltoetsen voor het meten van taalvaardigheid bij leerlingen in groep één en twee van het basisonderwijs. Het onderzoek is aangevraagd en gefinancierd door het Ministerie van OCW en gezamenlijk uitgevoerd door twee onderzoeksinstituten, het ITS van de Radboud Universiteit en het SCO-Kohnstamm Instituut van de Universiteit van Amsterdam. Hoewel gedurende het onderzoek regelmatig is overlegd tussen de auteurs over aanpak, interpretatie van resultaten en rapportage, is wel aan te geven welk instituut hoofdverantwoordelijk was voor onderscheiden delen van de studie en het rapport. De data zijn verzameld door het ITS en de analyses zijn verricht door het SCO-Kohnstamm Instituut, behalve de analyses gerapporteerd in paragraaf 4.1 tot en met 4.4 en de analyses gerapporteerd in tabel 77. De dataverzameling is gecoördineerd door Hans Versteegen van het ITS en voor de individuele toetsafnamen op de scholen zijn testleiders van het ITS ingezet. De paragrafen 4.1 tot en met 4.4, een deel van 4.10 en paragraaf 4.11 zijn geschreven door het ITS, de overige delen van het rapport zijn geschreven door het SCO-Kohnstamm Instituut.

Taaltoetsen voor taalzwakke leerlingen in het primair onderwijs

Managementsamenvatting

Aanleiding
Het SCO-Kohnstamm Instituut en het ITS zijn door het ministerie van OCW verzocht om een onderzoek te doen naar de geschiktheid van taaltoetsen voor het identificeren van leerlingen met een taalachterstand bij aanvang van het basisonderwijs. Daarbij werd gevraagd speciale aandacht te besteden aan de geschiktheid voor het toetsen van de meest taalzwakke leerlingen.

Onderzoeksvragen
Het onderzoek moet antwoord geven op de vraag hoe de toetsing ter bepaling van taalachterstanden bij jonge kinderen, dat wil zeggen kinderen in groep één en twee van de basisschool, het best kan worden uitgevoerd. Na overleg met de opdrachtgever zijn voor dit onderzoek de volgende onderzoeksvragen geformuleerd:
1) Zijn er deugdelijke taaltoetsen voor het meten van de taalvaardigheid van leerlingen in groep één en twee van het basisonderwijs? Welke taaltoetsen of onderdelen daarvan zijn het meest geschikt om de in het Nederlands minst taalvaardige leerlingen te oormerken? Welk type leerlingen kan met behulp van deze toetsen worden geselecteerd?
2) Indien er voor het hierboven genoemde doel geschikte toetsen zijn, hoe kan de taaltoetsing dan in de praktijk het best worden georganiseerd, wat zijn de kosten van grootschalige invoering in groep één en twee van het basisonderwijs en wat is de administratieve en organisatorische belasting voor de school bij het gebruik van deze toetsen?
3) Zijn deze taaltoetsen geschikt bij gebruik voor diagnostische doeleinden door leerkrachten?

Geëvalueerde toetsen
Op grond van verschillende criteria is gekozen voor het evalueren van (delen van) vier verschillende taaltoetsen: van de toets Taal voor Kleuters het onderdeel passieve woordenschat (TvK), de Peabody, die eveneens passieve woordenschat meet, van de Taaltoets Alle Kinderen (TAK) de onderdelen passieve woordenschat, klankarticulatie en klankonderscheiding en de gehele OBIS (Onderbouw Informatiesysteem) die naast passieve woordenschat, klankarticulatie en klankon-

derscheiding nog dertien andere subtoetsen kent, waarvan zes bedoeld als taaltoetsen en zeven als rekentoetsen.

Conclusies

De belangrijkste conclusies van het onderzoek zijn:

- De verschillende (sub)toetsen blijken voor leerlingen van vier tot acht jaar oud over het algemeen voldoende systematische variatie te bevatten of anders gezegd, een niet te grote proportie ruis te vertonen. Elk van de (sub)toetsen blijkt op zich in voldoende mate één vaardigheid te meten om somscores over de toetsen te rechtvaardigen.Hiermee is nog niet zeker dat de (sub)toetsen de vaardigheden meten waarvoor ze bedoeld zijn, maar wel dat de vragen behorend bij één (sub)toets in voldoende mate dezelfde vaardigheid meten. In andere woorden, de mate van willekeurige variatie in de scores overschrijdt voor vrijwel alle (sub)toetsen niet de algemeen geldende methodologische normen.
- De stabiliteit van de (sub)toetsen is minder goed. Hiermee bedoelen we de mate waarin de (sub)toetsen bij herhaalde afname gelijke resultaten per getoetste leerling opleveren. Acht van de 21 (sub)toetsen scoren onvoldoende volgens de algemeen geldende methodologische normen voor toetsing op individueel niveau. Taal voor Kleuters en de OBIS subtoetsen doen het hier iets beter dan de subtoetsen van de Taaltoets Alle Kinderen en de Peabody. De stabiliteit van de (sub)toetsen in het oormerken van de 25% minst taalvaardige leerlingen is alleen voldoende voor de TvK en voor de OBIS-subtoetsen passieve woordenschat, klankarticulatie, kennis van letters en kennis van getallen.
- Als we nagaan in hoeverre verschillende (sub)toetsen overeenstemmen in het oormerken van de 25% slechtst presterende leerlingen, dan blijken de (sub)toetsen hierin flink te verschillen. Het maakt dus veel uit welke (sub)toets men kiest om de zwakst taalvaardige leerlingen te selecteren.
- Uit de samenhang tussen (sub)toetsscores en de oude indeling naar leerlinggewicht[1], blijkt dat de (sub)toetsen niet erg hoog samenhangen met

[1] De gewichtenregeling is de basis voor de financiering van basisscholen in het kader van het onderwijsachterstandenbeleid. De hoogte van het leerlinggewicht onder de oude gewichtenregeling was o.a. afhankelijk van het opleidingsniveau en de etnische afkomst van ouders. Onder de huidige regeling is het leerlinggewicht alleen afhankelijk van het opleidingsniveau van ouders.

deze indeling. Ook de indeling van leerlingen naar of ze een vreemde taal spreken thuis dan wel Nederlands hangt niet erg hoog met de toetsscores samen. Duidelijk is dat een belangrijk deel van de leerlingen die voorheen het hoogste gewicht kregen (allochtoon van lage sociaal economische status) niet de minst taalvaardige leerlingen zijn;

- Er is een forse partijdigheid geconstateerd in de (sub)toetsscores, wat wijst op een probleem dat nader onderzocht moet worden om zicht te krijgen op de vraag wat de consequenties zijn voor de inzetbaarheid van de (sub)toetsen bij jonge leerlingen, bij leerlingen die het Nederlands niet als moedertaal hebben en bij leerlingen uit verschillende sociaaleconomische milieus. Duidelijk is dat de (sub)toetsen in de onderscheiden groepen niet alleen op een verschillende schaal (thuistaal en leeftijd), maar zelfs ongelijke vaardigheden meten (thuistaal, leerlinggewicht en leeftijd). Schaalverschil betekent dat leerlingen van gelijke vaardigheid, maar behorend bij verschillende groepen, ongelijke scores krijgen. Deze schaalverschillen belopen een halve tot een hele standaarddeviatie, wat veel is. Het meten van verschillende vaardigheden is overigens nog veel vervelender. Bij gevonden verschil in schaal zou men kunnen kiezen om (sub)toetsen te normeren per groep. Daar er echter in de voornoemde groepen verschillende vaardigheden gemeten worden, is gebruik van de (sub)toetsen over deze groepen heen niet te verdedigen. Partijdigheid kan getoetst worden aan de hand van een strenge en een coulante vorm van toetsing. De hierboven beschreven partijdigheid is aangetoond met de strenge toets. Bij gebruik van de coulante vorm van toetsing, vinden we dat alleen de partijdigheid naar leeftijd geschonden blijft. De resultaten laten zien dat ook bij coulante toetsing de in het onderzoek onderzochte (sub)toetsen bij leerlingen respectievelijk jonger en ouder dan zes jaar ongelijke aspecten van taalvaardigheid meten. Om meer duidelijkheid te krijgen over de mogelijke oorzaak van de gevonden partijdigheid naar leeftijd, zijn nadere analyses verricht. Het blijkt dat de gehanteerde meetmethode (papier versus p.c.) niet verantwoordelijk is voor de gevonden partijdigheid. Wel blijkt dat het model dat weergeeft welke aspecten van taalvaardigheid door de verschillende (sub)toetsen gemeten wordt, zeer goed past bij leerlingen ouder dan zes en relatief slecht bij de leerlingen jonger dan zes. De conclusie luidt dat, zelfs als we uitgaan van de coulante vorm van toetsing, is gebleken dat de (sub)toetsen niet bij

leerlingen van vier tot zes jaar afgenomen moeten worden als de scores consequenties hebben op individueel niveau.

- Het vooraf selecteren van leerlingen op grond van leerlinggewicht, sekse, thuistaal en het oordeel van de leerkracht over de taalvaardigheid Nederlands van de leerling, om zo bij slechts een deel van de leerlingen taaltoetsen af te nemen en op een efficiëntere wijze de 25% slechtst presterende leerlingen te detecteren, is niet aan te raden. De voorspellende kracht van de genoemde variabelen blijkt relatief gering.
- Om na te gaan hoe goed de voorspellende waarde van de toetsen is voor later presteren, is een meting van het taalvaardigheidniveau van dezelfde leerlingen in groep drie en vier nodig. Zou men besluiten tot deze longitudinale dataverzameling, dan kan tevens nagegaan worden in hoeverre de taalscores schools presteren voorspellen en of de gevonden partijdigheid bij oudere leerlingen vermindert. Gegeven de uitstekende modelfit bij leerlingen van zes tot acht jaar lijkt dit voor partijdigheid naar leeftijd zeer waarschijnlijk. Of de andere gevonden partijdigheid leeftijdafhankelijk is, is een open vraag. Overigens bleek deze andere partijdigheid bij coulante toetsing geen probleem.
- De toetsen blijken makkelijk af te nemen bij de jonge leerlingen en leraren verrichten de afnamen niet minder goed dan getrainde testleiders.
- De selecte steekproef van schoolleiders die deelnam aan het onderzoek staat over het algemeen positief tegenover het toetsen van leerlingen in groep één en twee van het basisonderwijs.
- Afname van de vier in het onderzoek gebruikte toetsen levert in de praktijk geen belemmeringen op. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400, naast personele kosten gemaakt voor het verwerken van de gegevens. Alleen bij de OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Bij de Peabody kost de verwerking van de gegevens weinig tijd, maar is de verkregen informatie ook summier;
- De gemiddelde afnameduur van de toetsen zoals in dit onderzoek afgenomen (TAK en TvK dus deels) varieert van 15 tot 20 minuten. Ook bleek een grote spreiding in de benodigde tijd per leerling. Bij de TvK kan worden bespaard op begeleidingstijd bij de afname, door groepsgewijze afname of door afname via de computer, al weten we niet of de

betrouwbaarheid en validiteit van de TvK zoals hier gerapporteerd, ook geldt voor groepsgewijze afname;

- Drie van de vier toetsen bieden aanwijzingen voor diagnostisch gebruik door leerkrachten. De TAK en de OBIS bieden gedetailleerde informatie op subtoetsniveau. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht, het observatie- en hulpprogramma. Bij de Peabody is het inzicht in de mate van taalontwikkeling beperkt en ontbreken suggesties voor verbetering.

Nadere toelichting op de conclusies

Steekproef

Om de onderzoeksvragen te beantwoorden zijn bij een steekproef van ruim 800 leerlingen uit groep één en twee taaltoetsen afgenomen. De deelnemende leerlingen zaten op scholen met relatief veel achterstandleerlingen. Een deel van de leerlingen maakte dezelfde toets twee maal met een tussenpoos van ongeveer twee weken, om na te kunnen gaan of de taalscores stabiel blijven over de tijd. Een deel van de toetsen is door de groepsleerkrachten zelf afgenomen, een ander deel door externe, getrainde testleiders. De TvK, TAK en Peabody zijn schriftelijk afgenomen, de OBIS werd door de kinderen op de p.c. gemaakt.

Homogeniteit (mate waarin vragen van één toets onderling samenhangen ofwel hetzelfde meten)

Uit het onderzoek blijkt dat de taaltoetsen over het algemeen voldoende homogeen zijn (niet teveel ruis bevatten) in de totale steekproef (leerlingen van vier tot acht jaar oud). Als we als criterium een minimale homogeniteit van .80 nemen (een algemeen geaccepteerd minimum voor het geven van oordelen op individueel niveau), dan blijkt alleen de homogeniteit van enkele kleine OBIS-subtoetsen lager (leesbegrip, kennis van rekenjargon, kennis van namen van vormen, hoofdrekenen en sommen B).

Omdat de homogeniteit niet een kenmerk is van een toets, maar van een toets afgenomen in een specifieke populatie, is de homogeniteit nogmaals bepaald, maar nu alleen voor de jongste leerlingen. Het onderzoek moet immers nagaan in hoeverre de toetsen voor jonge leerlingen geschikt zijn en het is denkbaar dat toetsen voor de jongste leerlingen in de steekproef minder homogeen zijn dan voor de

oudere leerlingen. De homogeniteit voor leerlingen van vier tot vijf jaar en voor de leerlingen van vier tot zes jaar blijkt echter eveneens goed.

Stabiliteit over de gehele scorerange
Ook de stabiliteit van de toetsscores blijkt over het algemeen goed, al is deze voor enkele subtoetsen laag. Met name de subtoetsen van de TAK, en in mindere mate ook de Peabody, scoren niet hoog qua stabiliteit. Na twee weken bleken scores teveel veranderd. Daarnaast is de stabiliteit van enkele kleinere subtoetsen uit de OBIS laag. Ook de stabiliteit is nogmaals bepaald voor alleen de jongere leerlingen. In een substeekproef van vier tot zes jaar zien we dat de stabiliteit van zes (sub)toetsen te laag uitvalt. Het betreft de Peabody, de TAK-toetsen klankarticulatie, klankonderscheiding en passieve woordenschat en de OBIS-toetsen aanvankelijk schrijven, leesbegrip en rekenjargon. Blijkbaar zijn de scores op deze toetsen bij de leerlingen jonger dan zes jaar onvoldoende stabiel. Ook blijkt dat de OBIS subtoetsen over het algemeen juist wat stabieler zijn voor de jongere leerlingen.

Stabiliteit in het selecteren van het kwart minst taalvaardige leerlingen
De hierboven gepresenteerde resultaten betreffen de homogeniteit en stabiliteit van de toetsen over de gehele scorerange van de toetsen. Daar we in dit onderzoek moeten kijken in welke mate de taaltoetsen voldoen als het gaat om het detecteren van de meest taalzwakke leerlingen, is ook nagegaan hoe goed de (sub)toetsen de 25% laagst taalvaardige leerlingen kunnen onderscheiden. Uit deze analyses blijkt dat de stabiliteit van dit oordeel veel lager is. Als we berekenen in welke mate de toetsen bij de eerste afname en bij de hertest dezelfde leerlingen aanwijzen als de 25% meest taalzwakke leerlingen, dan blijkt deze overeenstemming 12 van de 17 keer onder de norm die algemeen gesteld wordt bij het geven van individuele oordelen. De enige toetsen die bij deze analyses wel voldoende betrouwbaar blijken, zijn de TvK en de OBIS-toetsen passieve woordenschat, klankarticulatie, kennis van letters en kennis van getallen.

Wat meten de verschillende (sub)toetsen?
Nagegaan is in welke mate *verschillende* subtoetsen dezelfde dan wel verschillende vaardigheden meten. Het blijkt dat wanneer de verschillende subtoetsen in één analyse worden onderzocht, zij een zeer goed inhoudelijk te verdedigen structuur vertonen. We vinden in de toetsen een woordkennisfactor, een fonologische

factor, een technisch lezenfactor en een rekenfactor. De woordkennisfactor wordt gemeten door de TvK (onderdeel passieve woordenschat), de Peabody, het onderdeel passieve woordenschat van de TAK en de onderdelen passieve woordenschat, rekenjargon en kennis van vormen van de OBIS. De fonologische factor wordt gemeten door de toetsen klankarticulatie en klankonderscheiding van zowel de OBIS als de TAK (dus vier toetsen in totaal). Op de technisch lezenfactor laden de OBIS-toetsen aanvankelijk schrijven, leesbegrip, kennis van letters, leesvaardigheid woorden en kennis van getallen. Op de rekenfactor tenslotte laden de OBIS-toetsen geheugen, sommen A, hoofdrekenen en sommen B. Duidelijk is dat enkele OBIS-subtoetsen die bedoeld zijn om rekenvaardigheid te meten in feite taalvaardigheid meten (rekenjargon en kennis van vormen).

Partijdigheid

Er zijn verschillende analyses verricht om na te gaan of de toetsen partijdig zijn ten aanzien van specifieke groepen leerlingen in onze steekproef. Bij het verifiëren van de partijdigheid van de toetsen is eerst nagegaan of de subtoetsen in de onderscheiden groepen op een gelijke schaal meten. De te beantwoorden vraag is hier of leerlingen uit verschillende groepen die even vaardig zijn, gelijke scores krijgen. Anders gezegd, deze analyses gaan na of leerlingen die behoren tot verschillende groepen en die gelijke scores op de toets hebben, ook even vaardig zijn op de gemeten trek. Daarna is nagegaan of de toetsen dezelfde vaardigheid meten in de onderscheiden groepen leerlingen.

De partijdigheidsanalyses zijn verricht voor groepen gebaseerd op sekse (jongens versus meisjes), thuistaal (spreekt men thuis Nederlands of een vreemde taal), de oude leerlinggewichten (de .00-leerlingen versus de overigen en de .90-leerlingen versus de overigen) en leeftijd (leerlingen van vier tot zes versus leerlingen van zes tot acht). Verder zijn deze analyses niet per toets verricht, maar voor de gehele toetsbatterij tezamen, maar dan uitsluitend voor de taaltoetsen. De rekentoetsen zijn om technische redenen uit deze analyses verwijderd.

Uit de analyses blijkt dat de toetsen niet partijdig zijn voor jongens of meisjes, maar wel ten aanzien van alle andere onderscheiden groepsindelingen (thuistaal, leerlinggewicht en leeftijd). Deze partijdigheid betreft zowel de schaal waarop gemeten wordt (thuistaal en leeftijd) als de gemeten vaardigheid (thuistaal, leerlinggewicht en leeftijd). Binnen de door deze variabelen onderscheiden groepen meten de toetsen taalvaardigheid dus niet op dezelfde schaal en, wat problematischer is, ze meten ook niet hetzelfde aspect van taalvaardigheid. Dit is een

probleem als men de toetsen wil gebruiken om individuele leerlingen uit de verschillende groepen (dus b.v. leerlingen die thuis wel of niet Nederlands spreken) op een eerlijke wijze te identificeren als zijnde zwak of juist niet zwak taalvaardig.
Om een indicatie te krijgen van de ernst van de gevonden partijdigheid, is voor de gevonden schaalverschilpartijdigheid nagegaan hoe groot de onterechte verschillen tussen scores van leerlingen in verschillende groepen zijn als we ze uitdrukken in standaarddeviaties. De onterechte verschillen in de scores variëren van een derde tot één en een kwart standaarddeviatie, wat veel is. Het werkelijke invaliderende effect van de gevonden partijdigheid is echter nog veel groter dan in deze effectmaat wordt uitgedrukt. Immers, we meten niet alleen op een verschillende schaal, we meten ook verschillende vaardigheden in de onderscheiden groepen (naar thuistaal, leerlinggewicht en leeftijd). De mate waarin toetsen verschillende vaardigheden meten, zijn niet in standaarddeviaties uit te drukken, daar we dan verschillende eenheden zouden vergelijken.
Met name de partijdigheid naar leeftijd blijkt groot. Daarom is apart nagegaan hoe het gehanteerde factormodel past bij respectievelijk leerlingen jonger en leerlingen ouder dan zes jaar. Uit deze analyses blijkt dat het model voor de leerlingen ouder dan zes jaar heel erg goed past, terwijl voor leerlingen jonger dan zes het model relatief slecht blijkt te passen. Dit wijst op een probleem bij het toetsen van leerlingen jonger dan zes. Ook is nagegaan of verschil in gehanteerde afnamemethode (papier versus p.c.) verantwoordelijk zou kunnen zijn voor (een deel van) de gevonden partijdigheid, maar dit bleek niet het geval te zijn.

Welke leerlingen komen als taalzwak uit de bus met de gebruikte toetsen?
Hoewel de resultaten van het partijdigheidonderzoek dusdanig zijn dat men zich terecht af kan vragen of een toetsing van de taalvaardigheid bij de jongste helft van de leerlingen in onze steekproef op een deugdelijke wijze kan geschieden aan de hand van deze toetsen, is toch nog verder gekeken naar hoe de toetsen zich gedragen. Nagegaan is in hoeverre de scores op de subtoetsen samenhangen met de oude gewichtenregeling die onderscheid maakt tussen .00-, .25- en .90-leerlingen. Daarnaast is nagegaan in hoeverre de scores op de toetsen samenhangen met de taal die thuis gesproken wordt. Deze variabele onderscheidt leerlingen die thuis Nederlands spreken, leerlingen die thuis een Nederlands dialect of Fries spreken en leerlingen die thuis een vreemde taal spreken. Tegelijk is nagegaan of de toetsen voldoende spreiding vertonen bij de onderzochte groepen leerlingen.

Uit de analyses blijkt dat de scores van de meeste subtoetsen redelijk tot goed spreiden. Ook blijkt dat de toetsscores niet erg hoog samenhangen met de groepsindelingen naar leerlinggewicht of thuistaal. Hieruit volgt dat als we leerlingen op grond van de toetsscores selecteren, we andere groepen zullen vinden dan wanneer we selecteren op thuistaal of leerlinggewicht.

Maakt het uit met welke toets we de taalzwakke leerlingen opsporen?
Een ander belangrijk aspect is of de verschillende subtoetsen overeenstemmen bij het identificeren van de 25% qua taal slechtst presterende leerlingen. Uit analyses blijkt dat het vrij veel uitmaakt met welk van de toetsen men deze selectie maakt. De mate van overeenstemming tussen verschillende taaltoetsen is niet erg hoog. Duidelijk is dat de toetsen niet op gelijke wijze differentiëren tussen het qua prestatie laagste kwart en bovenste driekwart van de leerlingen. Ook is een overzicht gemaakt waarin per toets de leerlingkenmerken leerlinggewicht, sekse en thuistaal van het onderste kwartiel aan scores op de betreffende toets gegeven worden. Ook uit dit overzicht blijkt dat het veel uitmaakt met welke toets men de 25% laagst taalvaardige leerlingen selecteert.

In hoeverre zijn oordelen van leerkrachten over de taalvaardigheid van hun leerlingen gelijk aan de toetsscores?
De leerkrachtoordelen over de luister- en spreekvaardigheid Nederlands van de leerlingen blijken vrij sterk samen te hangen met de toetsscores. De correlaties tussen de leerkrachtoordelen over spreekvaardigheid Nederlands en de vier woordenschattoetsen lopen bijvoorbeeld van .53 (Peabody) tot .59 (TAK-passieve woordkennis). Voor luistervaardigheid lopen de correlaties met de passieve woordkennistoetsen van .51 (TvK) tot .56 (OBIS-passieve woordenschat). De onderlinge correlaties tussen de toetsscores voor woordkennis lopen van .86 (Peabody * TvK) tot .63 (OBIS-passieve woordkennis * Peabody). Als we bedenken dat de spreek- en luistervaardigheid niet identiek zijn aan passieve woordkennis, dan kunnen we concluderen dat de leerkrachtoordelen het zeer goed doen. Zeker als we ook nog bedenken dat elke leerkracht alleen zijn eigen leerlingen als referentie kan gebruiken. Het is immers aannemelijk dat de oordelen die leerkrachten geven over de taalvaardigheid van hun leerlingen gekleurd worden door de gemiddelde taalvaardigheid van de leerlingpopulatie waar de leerkracht mee werkt. Deze gemiddelde taalvaardigheid, ofwel het aantal taalzwakke leerlingen, verschilt per leerkracht en dus kan men verwachten dat leerkrachtoordelen

hierdoor een zwakkere relatie met objectiever toetsscores zouden vertonen dan we nu zien. Blijkbaar komen de referentiekaders van de verschillende leerkrachten vrij aardig overeen.

Is preselectie voorafgaand aan toetsing zinvol?
Om na te gaan of het verantwoord zou zijn om leerlingen voorafgaand aan een eventuele taaltoetsing te selecteren op grond van leerlingkenmerken, om zo alleen die leerlingen te toetsen waarvan men kan aannemen dat ze meer kans hebben om als taalzwak aangemerkt te worden, zijn aparte analyses verricht. Het blijkt dat de door ons gebruikte achtergrondkenmerken van de leerlingen (leerlinggewicht, sekse, thuistaal en de leerkrachtoordelen over de spreek- en luistervaardigheid van de leerling) het al dan niet behoren bij de 25% zwakst presterende leerlingen maar matig voorspellen (5 tot 45% correct, afhankelijk van de gebruikte taaltoets). Zouden we alleen leerlingen toetsen waarvan we op grond van deze predictoren voorspellen dat ze tot de 25% minst taalvaardige leerlingen behoren, dan zouden we dus 55 tot 95% van de taalzwakke leerlingen onterecht niet toetsen. Uiteraard is het mogelijk deze foutenmarge te verkleinen. Men zou b.v. kunnen proberen om niet het meest taalzwakke kwart maar de meest taalzwakke helft van de leerlingen via preselectie te bepalen. Als we deze onderste helft van de leerlingen vervolgens toetsen om hiervan weer de meest taalzwakke helft te selecteren en zo uiteindelijk het minst taalvaardige kwart over te houden, zal het aantal onterecht niet getoetste leerlingen afnemen. De voorspellende kracht van de gebruikte predictoren is echter zo matig, dat dit met deze predictoren niet raadzaam lijkt.

Kunnen leerkrachten de toetsen ook afnemen?
Ook is gekeken of leerkrachten de toetsen even goed kunnen afnemen als externe, getrainde testleiders. Deze twee groepen blijken elkaar in afnamekwaliteit niet veel te ontlopen. De conclusie is dan ook dat leerkrachten goed in staat zijn om deze taaltoetsen af te nemen.
Kijken we naar de commentaren van testleiders op het afnamegemak van de toetsen, dan blijken de toetsafnamen nauwelijks problemen te hebben opgeleverd. Een enquête onder schoolleiders om te peilen hoe zij denken over het toetsen van jonge leerlingen laat zien dat de schoolleiders die bij dit onderzoek betrokken waren (een kleine en niet-random steekproef) positief staan tegenover het toetsen van jonge leerlingen.

Wat kosten afnamen?
Voor het beantwoorden van de laatste twee onderzoeksvragen, zijn kwalitatieve analyses verricht. De afname van de vier toetsen levert in de praktijk geen belemmeringen op. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400. Daarnaast is inzet van personeel nodig om de gegevens te verwerken en de resultaten te berekenen. Bij OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Bij de Peabody kost de verwerking van de gegevens weinig tijd, maar is de verkregen informatie ook summier.
De afnameduur varieert. Gemiddeld kan worden uitgegaan van een afnameduur van 15 à 20 minuten. Daarbij moet worden aangetekend dat van de TAK en de TvK delen zijn geselecteerd voor dit onderzoek. Afname van de gehele toetsen kost (aanzienlijk) meer tijd. Ook bleek dat er een grote spreiding was in de benodigde tijd. De snelste leerlingen konden de aangeboden toetsen of toetsonderdelen in enkele minuten verwerken, terwijl afname bij de langzaamste leerlingen meer dan een half uur in beslag nam.

Voorzien de toetsen diagnostische toepassingen?
Ook is gekeken naar de geschiktheid van de taaltoetsen voor diagnostisch gebruik door leerkrachten. De vier toetsen hebben als doel inzicht te geven in de mate van achterstand die leerlingen hebben op taalgebied. Bij de TAK en de OBIS kan gedetailleerde informatie op subtoetsniveau worden verkregen. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht, het observatie- en hulpprogramma. Bij de Peabody is het geboden inzicht in de mate van taalontwikkeling beperkt en ontbreken suggesties voor verbetering. Deze toets biedt dus de minste mogelijkheden op diagnostisch gebied.

Discussie
In het bovenstaande worden de conclusies gepresenteerd aangaande het onderzoek naar de kwaliteit van taaltoetsen voor het toetsen van leerlingen in groep één en twee van het primair onderwijs. Nu maakt het voor het oordeel over de kwaliteit van de toetsen veel uit of men de toetsscores gebruikt zodanig dat er consequenties uit volgen voor het getoetste individu, of dat we alleen beslissingen op groepsniveau aan de toetsuitslagen willen verbinden. In ons onderzoek zijn we

uitgegaan van gebruik van scores op individueel niveau. Dit is bijvoorbeeld het geval als we op grond van de toetsscores leerlingen extra taalonderwijs gaan geven. Bij het gebruik van de toetsen voor het toekennen van een leerlinggebonden budget, is niet noodzakelijk een individuele consequentie van de toetsuitslag het gevolg. Als scholen alle ontvangen leerlinggebonden gelden in de schoolkas storten, is niet de individuele toekenning van gelden het criterium waar we naar moeten kijken. Dan is de vraag eerder of scholen krijgen waar ze recht op hebben. Fout toegekende leerlinggebonden budgetten kunnen elkaar dan binnen scholen uitmiddelen. Zouden scholen echter leerlingen meer of minder graag accepteren vanwege verschil in leerlinggebonden budget, dan is er bijvoorbeeld wel weer sprake van een individueel gevolg.

Een ander aspect dat van belang is, is om uit te maken wat de argumenten zouden zijn om een taaltoets te gebruiken om het onderwijs te financieren. Er zijn minstens twee soorten leerlingen te onderkennen die laag scoren op een taaltoets. Leerlingen voor wie het Nederlands een vreemde taal is en leerlingen die het Nederlands als moedertaal spreken. Bij de eerste groep is extra taalonderwijs zowel voor de leerling als voor de samenleving in potentie zeer lucratief. Immers, deze leerlingen kunnen hoog intelligent zijn, waarbij het waarschijnlijk is dat hun schoolloopbaan gefrustreerd wordt door een gebrekkige beheersing van de instructietaal, het Nederlands. Bij moedertaalsprekers van het Nederlands die relatief laag scoren op een taaltoets hebben we veel meer kans een leerling te vinden die minder intelligent is, vooral voor wat betreft de talige intelligentie. Extra taalonderwijs voor deze leerlingen is nuttig om ze te helpen te realiseren wat mogelijk is in hun schoolloopbaan, maar deze investering zal waarschijnlijk zowel voor het individu als voor de maatschappij minder opleveren, waarmee we niet bedoelen dat een dergelijke investering niet zou moeten worden gedaan.
Een probleem dat hieraan verwant is, is het bepalen van het juiste aspect van taalvaardigheid dat in een dergelijke taaltoets gemeten moet worden. We hebben laten zien dat er in de taaltoetsen die in dit onderzoek zijn gebruikt, drie verschillende aspecten van taalvaardigheid gemeten worden: woordkennis, fonologische vaardigheid en technisch lezen. Er zijn echter nog vele andere aspecten van taalvaardigheid, al is de vraag of we die bij zeer jonge leerlingen kunnen meten. Het probleem is dat we niet weten of we nu woordkennis dan wel bijvoorbeeld de fonologische vaardigheid zouden moeten kiezen of nog een andere vaardigheid. Om een beredeneerde keuze te kunnen maken, zouden we moeten weten in hoeverre

de verschillende aspecten van taalvaardigheid de schoolvorderingen (of een ander criterium) prediceren en eigenlijk ook welke aspecten van taalvaardigheid in een causale relatie staan tot datgene wat we willen remediëren. Dit laatste heeft ook weer te maken met de keus die in de voorgaande alinea geschetst wordt met betrekking tot het onderscheid tussen vreemde en moedertaalsprekers.

Bij het toetsen van individuen om te bepalen of ze ja dan nee als taalzwak aangemerkt moeten worden, kunnen twee soorten fouten gemaakt worden; men kan onterecht besluiten dat een leerling taalzwak is en men kan onterecht besluiten dat een leerling niet taalzwak is. Deze twee soorten fouten beïnvloeden elkaar. Als men probeert de proportie onterecht niet geselecteerde leerlingen te verkleinen, zal de proportie onterecht wel geselecteerde leerlingen toenemen. Zou men bijvoorbeeld stellen dat alle taalzwakke leerlingen als taalzwak aangemerkt moeten worden, dan is dat alleen te realiseren door alle leerlingen als taalzwak te oormerken. Het mag duidelijk zijn dat je dan net zo goed niet kunt toetsen. Alleen in de hypothetische situatie waarin een gebruikte toets perfect valide is en totaal geen ruis bevat, zullen beide typen fouten niet voorkomen. In de werkelijkheid zijn toetsen niet perfect valide en bevatten toetsscores wel ruis. Uit onze analyses blijkt dat de door ons onderzochte toetsen zelfs meer ruis bevatten (zie stabiliteit) dan men als norm voor toetsscores op individueel niveau hanteert. Ook uit de gebrekkige overeenkomst tussen verschillende toetsen in het oormerken van het kwart minst taalvaardige leerlingen, is duidelijk dat beide voornoemde typen fouten zelfs relatief vaak voorkomen en vaker dan de algemeen aanvaarde methodologische norm voor individueel toetsen toestaat.

De hierboven besproken partijdigheid naar leeftijd betrof simultane analyses op twee groepen leerlingen; leerlingen van vier tot zes jaar oud en leerlingen van zes tot acht jaar oud. Uit deze analyses bleken de toetsen partijdig naar leeftijd, en wel zo partijdig dat we kunnen stellen dat in beide groepen leerlingen niet dezelfde vaardigheid gemeten wordt. Hieruit zou men kunnen concluderen dat de leeftijdsgrens waaronder de taaltoetsen niet deugen exact zes jaar is. Een dergelijke conclusie kan echter niet hard gemaakt worden. Duidelijk is alleen dat als we de steekproef splitsen in een jongste en een oudste helft, we dan partijdigheid vinden. Om na te gaan onder welke leeftijd het mis gaat, zouden aanvullende analyses verricht moeten worden. Hier is gekozen voor een splitsing op ongeveer zes jaar, omdat de steekproef zo in twee even grote delen gesplitst wordt. Of de

leeftijd waaronder de toetsen niet meer meten wat ze zouden moeten meten op zes jaar ligt of bijvoorbeeld op vijf jaar en zes maanden, of zes jaar en zes maanden, weten we (nog) niet. Overigens is het waarschijnlijk dat de leeftijd waarop de toetsen het goed doen per kind verschilt en ook dat de mate waarin de toetsen partijdig zijn niet opeens, maar gradueel afneemt met de leeftijd.

Mocht men besluiten om taaltoetsen te gebruiken voor het nemen van beslissingen op individueel niveau, dan moeten we naast alle problemen die we hierboven hebben gemeld ook nog wijzen op het aangetoonde oefeneffect. Uit de herhaalde afnamen (zelfde toets voor de tweede keer bij dezelfde leerlingen) die werden gehouden om de stabiliteit van de toetsscores te kunnen bepalen, is gebleken dat leerlingen op de herhaalde afname voor 12 van de 21 (sub)toetsen significant hoger scoorden dan op de eerste afname. Dit is waarschijnlijk een leereffect dat volgt op afname van de toetsen. Hieruit volgt dus dat als we de toetsen herhaald zouden afnemen of als leerlingen de gebruikte toetsen, die commercieel op de markt zijn, eerder zouden hebben gemaakt, we ook hierdoor taalzwakke leerlingen als niet-taalzwak zouden oormerken.

1 Aanleiding

Het SCO-Kohnstamm Instituut en het ITS zijn door het ministerie van OCW verzocht om een onderzoek te doen naar de kwaliteit van taaltoetsen voor het identificeren van leerlingen met een taalachterstand bij aanvang van het basisonderwijs. Daarbij dient speciale aandacht te worden besteed aan de geschiktheid voor het toetsen van de meest taalzwakke leerlingen. Naast psychometrische kwaliteiten als betrouwbaarheid en validiteit, is ook de praktische bruikbaarheid van de toetsen van belang. Hierbij kan gedacht worden aan bijvoorbeeld afnameduur en kosten per afname. Ook van belang is hoe de organisatie van de taaltoetsing in de praktijk het best kan plaatsvinden. Tevens dient het onderzoek informatie op te leveren over de mate waarin de geselecteerde toetsen aanvullende informatie geven boven in het kader van de huidige gewichtenregeling gehanteerde criteria (SES en etniciteit).

1.1 Probleemstelling en onderzoeksvragen

Het onderhavige onderzoek moet antwoord geven op de vraag hoe de toetsing ter bepaling van taalachterstanden bij jonge kinderen, dat wil zeggen kinderen in groep één en twee van de basisschool, het best kan worden uitgevoerd. Na overleg met de opdrachtgever zijn voor dit onderzoek verschillende onderzoeksvragen vastgesteld:

1) Zijn er betrouwbare en valide taaltoetsen voor het meten van de taalvaardigheid van jonge kinderen? Welke taaltoetsen of onderdelen daarvan zijn het meest geschikt om de leerlingen te identificeren met de laagste taalvaardigheid Nederlands? Welk type leerlingen kan met behulp van deze toetsen worden geselecteerd?
2) Indien er voor het hierboven genoemde doel geschikte toetsen zijn, hoe kan de taaltoetsing dan in de praktijk het best worden georganiseerd, wat zijn de kosten van grootschalige invoering in groep één en twee van het basisonderwijs en wat is de administratieve en organisatorische belasting voor de school bij het gebruik van deze toetsen?

3) Zijn deze taaltoetsen geschikt bij gebruik voor diagnostische doeleinden door leerkrachten?

Het gaat erom dat op basis van afnamen van toetsen bij leerlingen en analyse van de resultaten daarvan, wordt nagegaan welke toetsen het meest geschikt zijn om taalachterstanden in kaart te brengen. Daarbij speelt bijvoorbeeld de vraag of de toetsen voor uiteenlopende groepen leerlingen geschikt zijn. Wat zijn eventuele redenen voor het niet slagen van de toetsafnamen? Deze vraag is van belang omdat het bij toetsafnamen bij jonge kinderen, en zeker bij leerlingen met een geringe taalvaardigheid Nederlands, voorstelbaar is dat afname van toetsen niet in alle gevallen slaagt[1]. In het onderzoek moet worden vastgesteld bij welke (groepen) leerlingen dat het geval is. Bij redenen voor het niet slagen van afnamen kan gedacht worden aan gebrek aan motivatie, verkeerd begrip van opdrachten, te moeilijke of te makkelijke opgaven, etc.

In hoeverre maken de toetsen het mogelijk om leerlingen uit groep één en twee met de laagste taalvaardigheid Nederlands te identificeren? Bij het beantwoorden van deze vraag moet nagegaan worden of de toetsen voldoende discrimineren op het lage taalvaardigheidniveau. Door de toetsresultaten te relateren aan een aantal achtergrondkenmerken van de leerlingen, kan worden nagegaan welk type leerling met behulp van de desbetreffende toets kan worden geselecteerd. Daarbij gaat het tevens om de vraag wat de toets toevoegt aan achtergrondkenmerken zoals sociaaleconomische status en etniciteit[2].

Bij de organisatie horen aspecten als afnamecondities, kostenbeheersing, registratie van resultaten, fraudegevoeligheid, en dergelijke. Aan welke eisen moet de afname van de toets voldoen? Moeten toetsen afgenomen worden door een getrainde toetsleider of is het ook mogelijk toetsen door leerkrachten af te laten nemen? Wat zijn de voor- en nadelen verbonden aan beide opties? Het antwoord op deze vraag is van belang om een inschatting van de toekomstige kosten te kunnen maken bij grootschalige afname. Ook is van belang hoe de registratie het

---

[1] Zie bijvoorbeeld de ervaringen met het toetsen van jonge kinderen in het onderzoek naar effecten van Kaleidoscoop en Piramide (Veen et al., 2000).

[2] Uit onderzoek blijkt dat de correlatie tussen leerprestaties en genoemde achtergrondkenmerken tussen .30 en .35 ligt (Sirin, 2005). Van een goede begintoets mag een hogere correlatie met latere leerprestaties worden verwacht.

best kan plaatsvinden en op welk moment de taaltoetsen het best kunnen worden afgenomen. Een belangrijk aspect van de uiteindelijke taaltoetsing is de omvang van de groep leerlingen die getoetst moet worden. Uit overwegingen van efficiency zou het wenselijk kunnen zijn om voor de toetsing op voorhand een selectie te maken van leerlingen met een extra grote kans op omvangrijke taalachterstand, op basis van bepaalde kenmerken[3]. De vraag naar deze mogelijkheid, ofwel de ideale vorm van preselectie zal ook in de studie worden meegenomen.

Ook de geschiktheid van de toetsen voor diagnostische doeleinden is een belangrijk punt, omdat daarop een verbeterprogramma kan worden gebaseerd, waarbij een individueel handelingsplan het uitgangspunt kan vormen. De toets moet dan concrete aanwijzingen geven over de gewenste inhoud van het verbeterprogramma.

[3] Gedacht kan worden aan criteria als aanspreekbaarheid in het Nederlands, inschatting van taalvaardigheid door de leerkracht, thuistaal waaronder ook dialect, etc.

2 Opzet van het onderzoek

2.1 Keuze van de te evalueren taaltoetsen

In een publicatie van Emmelot & Van Schooten (2005) wordt een schets gegeven van mogelijkheden, problemen en oplossingen aangaande het meten van de taalvaardigheid bij de laagst taalvaardige achterstandsleerlingen in groep één tot en met drie van het primair onderwijs. In deze publicatie worden, gebaseerd op interviews met deskundigen en de beoordelingen in de COTAN (Evers et al., 2002), een aantal taaltoetsen genoemd die bedoeld zijn voor leerlingen van vier tot zeven jaar en geschikt zouden kunnen zijn voor het selecteren van leerlingen met de grootste taalachterstand. Sommige van deze toetsen meten schrijf- of leesvaardigheid en zijn dus niet geschikt voor kinderen in groep één en twee. De overblijvende potentieel geschikte toetsen volgens Emmelot & Van Schooten (2005) zijn:

1. Taaltests voor Kinderen (Bon, 1982)
2. Reynell Test voor Taalbegrip (RTT) (Eldik, Schlichting, Lutje Spelberg, Meulen, & Meulen, 1995)
3. OBIS, Onderbouwinformatiesysteem (Hoeven, 2004-05)
4. PRAVOO (Koning & Westra, 2000)
5. Luisteren 1 (Krom, Ouborg & Kamphuis, 2001) (vanaf groep 3)
6. Taal voor Kleuters (TvK) (Kuyk, 1996)
7. TARSP (Schlichting, 1993; 2003)
8. Schlichting Test voor Taalproduktie (Schlichting, Eldik, Lutje Spelberg, Meulen, & Meulen, 1995)
9. Taaltoets Alle Kinderen (TAK) (Verhoeven & Vermeer, 2001)
10. Begrippentest-plaatjes, begrippentest woorden en woordenschattest (Aarnoutse 1988/1996) (groep twee en drie)

Selectiecriteria toetsen

Niet alle toetsen lijken even geschikt op grond van de COTAN beoordelingen (Evers et al., 2002) en de beschrijvingen van de toetsen. Zo wordt gesteld dat de RTT en de Schlichting Test voor Taalproduktie expliciet alleen voor autochtone leerlingen bedoeld zijn, wat overigens niet noodzakelijk impliceert dat ze onge-

schikt zijn voor anderstalige leerlingen, of, beter gezegd, minder geschikt zijn voor anderstalige leerlingen dan de andere toetsen die een dergelijk voorbehoud niet maken. Daarnaast wordt gesteld dat deze beide toetsen afgenomen moeten worden door een psycholoog, pedagoog of logopedist, wat de afnamen aanzienlijk compliceert. Onzeker is of deze twee toetsen ook door getrainde toetsleiders afgenomen zouden kunnen worden. Luisteren 1 is bedoeld voor leerlingen vanaf groep drie en de Begrippentesten zijn bedoeld voor groep twee en drie. Als we bedenken dat de toetsen nu gebruikt moeten worden om de minst taalvaardige leerlingen in groep één en twee te detecteren, wordt duidelijk dat de toetsen met name op de laagste taalvaardigheidniveaus goed moeten differentiëren. Toetsen bedoeld voor reguliere leerlingen in groep twee of drie zijn wellicht daartoe minder geschikt dan toetsen bedoeld voor leerlingen in groep één.
De keuze voor opname van toetsen in het onderhavige onderzoek is ook afhankelijk van de afnameduur van de toetsen. De geschatte afnameduur, zoals beschreven in de COTAN (Evers et al., 2002) dan wel de eigen toetsverantwoording, loopt van 12 minuten (Taaltests voor Kinderen) tot 25 à 35 minuten (Begrippentest). Van twee toetsen wordt de afnameduur 'variabel' genoemd (RTT en Schlichting). Van de PRAVOO, Luisteren 1, de TARSP en de Taaltoets Alle Kinderen (TAK), is de afnameduur niet aangegeven. Een andere overweging bij de keuze van toetsen betreft de gemeten aspecten van taalvaardigheid. Sommige toetsen zijn wat dit betreft smal en meten één aspect, zoals de Peabody (Dunn & Dunn, 2005) en Taal voor Kleuters (TvK), die alleen woordenschat meten. Andere toetsen zijn breder en meten meerdere aspecten, zoals de TAK en de OBIS.

In de eerste fase van het onderzoek is op grond van enerzijds praktische overwegingen aangaande kwaliteiten van toetsen als de afnameduur, kosten bij afname, vereiste kennis en vaardigheden bij toetsleiders, doelpopulatie e.d., en anderzijds inhoudelijke afwegingen (wat meten de toetsen en hoe doen ze dat?) een keuze gemaakt voor welke toetsen in de huidige studie zijn geëvalueerd. Bij het maken van deze keuze gold de overweging dat we zoveel mogelijk potentieel geschikte toetsen wilden meenemen, waarbij wat 'mogelijk' is, vooral werd bepaald door de moeilijkheidsgraad van de toetsen en de afnameduur die binnen het onderzoek aanvaardbaar was.

Beschrijving van de geselecteerde toetsen

Op grond van de toetsbeschrijvingen, met name op grond van gegevens over de afnameduur en de moeilijkheidsgraad, is na overleg met de opdrachtgever besloten in het onderzoek de volgende delen van toetsen mee te nemen:

- Taal voor kleuters (de versie voor jongste kleuters) (TvK)
- Taaltoets alle Kinderen (TAK), de onderdelen passieve woordenschat, klankonderscheiding en klankarticulatie.
- De Peabody
- De OBIS

De Peabody en de OBIS zijn geheel afgenomen en van de TAK alleen de drie subtoetsen 'passieve woordenschat', 'klankonderscheiding' en 'klankarticulatie'. Van de TvK is alleen het onderdeel 'passieve woordenschat' afgenomen. Selectie was nodig, omdat anders de geschatte afnameduur langer werd dan het half uur dat we per toets wilden reserveren. Van de vier bovengenoemde toetsen zijn er drie afgenomen aan de hand van plaatjes op papier en door toetsleiders gesproken teksten. De OBIS is afgenomen met behulp van een p.c. Elk van de vier bovengenoemde toetsen meet receptieve woordenschat (in de toetsen 'passieve woordenschat' genoemd). Het deel van de TvK dat wij afnamen en de Peabody meten uitsluitend passieve of receptieve woordenschat. Van de TAK is één van de drie geselecteerde toetsen een receptieve woordenschattoets en de OBIS bevat ook een subtoets voor receptieve woordenschat. Verder zijn van de TAK de subtoetsen 'klankonderscheiding' en 'klankarticulatie' gekozen voor het onderzoek, omdat deze toetsen horen tot de makkelijker subtoetsen van de TAK en de geschatte afnameduur van de drie gekozen TAK-toetsen volgens de handleiding ongeveer een half uur bedraagt. In de OBIS zijn eveneens twee subtoetsen voor respectievelijk 'klankonderscheiding' en 'klankarticulatie' opgenomen. De OBIS bevat verder nog subtoetsen voor 'aanvankelijk schrijven', 'leesbegrip', 'kennis van de letters van het alfabet', 'leesvaardigheid woorden', 'leesvaardigheid zinnen', 'leesvaardigheid teksten', 'rekenjargon', 'geheugen voor aantallen', 'sommen A', 'kennis van getallen', 'kennis van namen van vormen', 'hoofdrekenen' en 'sommen B'.

De toetsen voor receptieve woordenschat vragen de leerling om bij een door de testleider gesproken woord een passend plaatje aan te wijzen uit drie of vier plaatjes (TvK, TAK en Peabody), of tonen een getekend tafereel op een p.c.-

scherm waarbij de leerling met de muis moet klikken op de afbeelding die past bij een gesproken woord (OBIS).
De toets klankarticulatie van de TAK vraagt leerlingen om woorden na te zeggen die de testleider uitspreekt. De testleider beoordeelt of het woord correct wordt uitgesproken. De testleiders krijgen de instructie om het toevoegen van een schwa ('mellek' i.p.v. melk) en het stemloos maken van medeklinkers 'z' en 'v' ('s' i.p.v. 'z' en 'f' i.p.v. 'v') niet fout te rekenen. De toets klankarticulatie van de OBIS maakt gebruik van onzinwoorden en vraagt de leerling die na te zeggen.
De toets klankonderscheiding van de TAK vraagt de leerlingen om van twee door de testleider uitgesproken woorden aan te geven of deze woorden gelijk zijn of niet. De toets 'klankonderscheiding' van de OBIS vraagt leerlingen om rijmwoorden te herkennen.
De OBIS meet nog een aantal vaardigheden dat niet door de andere toetsen gemeten wordt, voor een deel taalvaardigheden, voor een deel rekenvaardigheden. De toets leesbegrip vraagt leerlingen aan de hand van plaatjes aan te geven waar iemand op het plaatje leest of schrijft en waar een woord of letter te zien is, waar een punt op hoofdletter te zien zijn, enz. De toets kennis van letters vraagt leerlingen de op een p.c.-scherm aangeboden letters te benoemen of verklanken. Leesvaardigheid woorden vraagt de leerling korte woordjes van het scherm te lezen, leesvaardigheid zinnen doet hetzelfde met korte zinnen. Leesvaardigheid teksten vraagt om een verhaaltje te lezen en opengelaten plekken zelf correct in te vullen (als een close toets). De laatste twee toetsen zijn uiteraard erg moeilijk voor leerlingen in groep één en twee en eigenlijk ook bedoeld voor oudere en vaardiger leerlingen. De OBIS kent een afnameprocedure waarbij deze toetsen alleen worden aangeboden als de leerling op de daarvóór afgenomen delen goed genoeg presteert. Om deze reden zijn deze moeilijke subtoetsen niet verwijderd uit de OBIS, ook al omdat afname van de OBIS dan volgens de gewone procedure kon plaatsvinden.
De eerste rekentoets is de toets rekenjargon. Deze toets vraagt naar de betekenis van woorden als 'groter' en 'kleiner', 'meer' en 'minder' enz. Deze toets is dus ook op te vatten als een woordkennistoets. Bij de geheugentoets wordt de leerlingen enkele objecten op het scherm geboden (bijvoorbeeld vier appels). Nadat het plaatje van het scherm is verdwenen, vraagt de toetsleider de leerling hoeveel appels er te zien waren. De toets sommen A betreft het maken van simpele rekensommetjes. In de toets getallenkennis vraagt men de leerling de namen van

gegeven getallen uit te spreken. In de toets vormen wordt de leerlingen gevraagd naar de namen van gegeven vormen (driehoek, vierkant enz.) en deze toets is dus ook als vocabulairetoets op te vatten. In de toets hoofdrekenen vraagt men de leerling bij een gegeven aantal stippen hoeveel het er worden als er bijvoorbeeld twee bij komen. 'sommen B' tenslotte betreft ook sommetjes, maar nu wat moeilijker dan sommen A. Ook worden er bij sommen B wiskundige symbolen gebruikt (+, -, =).

2.2 Afnamedesign

Om na te kunnen gaan in hoeverre verschillende taaltoetsen gelijke, dan wel verschillende vaardigheden meten, zijn factoriële analyses nodig. Om deze analyses te kunnen uitvoeren, moeten verschillende toetsen gemaakt worden door dezelfde leerlingen. Ook voor de analyses gericht op het bepalen in hoeverre verschillende toetsen dezelfde leerlingen oormerken als zeer taalzwak, moeten meerdere toetsen door dezelfde leerlingen gemaakt worden. In een psychometrisch gezien ideaal design maken alle kinderen in de steekproef alle toetsen. Als we echter gemiddeld een half uur afnametijd per toets rekenen, zou dat bij afname van vier toetsen neerkomen op twee uur per leerling, wat volgens ons redelijkerwijs niet van een school of leerling als inspanning gevraagd kon worden. Toch moeten we garanderen dat de data zo verzameld worden, dat alle combinaties van geselecteerde toetsen op samenhang en factoriële structuur onderzocht kunnen worden. Om deze reden gebruikten wij een onvolledig, gebalanceerd blokontwerp (Meerling, 1981) als design voor de dataverzameling. In dit design komen alle combinaties even vaak voor (zie Figuur 1).

Figuur 1 Onvolledig gebalanceerd blokontwerp met vier toetspakketten

	Toetspakketten			
afnamecondities	1	2	3	4
A	X	X		
B	X		X	
C	X			X
D		X	X	
E		X		X
F			X	X

Zoals hierboven al is gesteld, is in overleg met de opdrachtgever besloten vier toetsen af te nemen met een half uur als bovengrens voor de afnameduur per toets. Met inachtneming van de bovenstaande overwegingen, is de grootte van de steekproef bepaald. Uitgangspunten hierbij waren; (a) voldoende leerlingen om de noodzakelijke analyses uit te kunnen voeren, (b) rekening houden met uitval van leerlingen die niet toetsbaar blijken, en (c) kostenbeheersing. Uitgaande van vier toetspakketten die elk een half uur afnametijd vragen en het afnemen van niet meer dan twee toetspakketten per leerling, levert dit een afnamedesign op met zes condities. Het gebalanceerd blokontwerp is weergegeven in Figuur 1 en 2. De toewijzing van leerlingen aan afnamecondities is aselect geschied, zodat bij het analyseren van de data uitgegaan kan worden van ontbrekende waarnemingen die 'missing completely at random' zijn.

2.3 Steekproefkader

De te onderzoeken groep zijn leerlingen in groep één en twee van de basisschool. Om verschillende redenen is het wenselijk af te zien van beperking van de afname tot leerlingen waarvan vooraf wordt verwacht dat zij tot de taalzwakke groep behoren. In de eerste plaats is het wenselijk dat de geschiktheid van de toetsen – mede gezien de vraag naar mogelijke diagnostische toepassingen – wordt bepaald aan de hand van afname bij een bredere groep dan alleen (zeer) taalzwakke leerlingen. Immers, als alleen taalzwakke leerlingen getoetst worden, kan niet meer nagegaan worden in hoeverre de toetsen in staat zijn deze taalzwakke groep uit een bredere groep te onderscheiden. Daarnaast kan dan ook niet nagegaan worden of er achtergrondvariabelen zijn die gebruikt kunnen worden voor een preselectie. In de tweede plaats zou selectie ertoe leiden dat er per school slechts een kleine groep leerlingen getoetst wordt. Daarmee neemt het aantal te bezoeken scholen – en daarmee de tijdsinvestering door de testleiders – onevenredig toe.
Om praktische redenen (kosten, organisatie) is het wenselijk om de leerlingen voor het onderzoek vooral te zoeken op scholen die relatief veel achterstandsleerlingen hebben. Voor de selectie van deze scholen is gebruik gemaakt van het PRIMA-bestand 2005. Het PRIMA-bestand bevat een representatieve steekproef van scholen voor primair onderwijs in Nederland. Daarnaast kent het PRIMA-bestand een aanvullende steekproef van scholen met relatief veel leerlingen uit

achterstandsituaties. Bij de steekproeftrekking is ook gebruik gemaakt van deze aanvullende steekproef. Ook is bij de steekproeftrekking gezorgd voor voldoende spreiding over variabelen als urbanisatiegraad, verblijfsduur in Nederland, thuistaal dan wel -dialect, etnische achtergrond, enz. In de PRIMA-data zijn veel gegevens opgenomen over de voornoemde variabelen.

Voor de geplande analyses naar de factoriële structuur van de toetsen, zijn minimaal 200 maar liever 300 leerlingen per toetsende factoranalyse gewenst (Comrey & Lee, 1992). Als de correlaties hoog zijn en het aantal factoren gering, zou eventueel met een iets kleinere steekproef gewerkt kunnen worden. Gegeven de verwachte 'restriction of range', we toetsen immers vooral laag taalvaardige leerlingen, hebben we toch 200 cases per factoranalyse als absolute ondergrens aangehouden. De restriction of range impliceert immers lagere correlaties.
Voor een toetsende factoranalyse per toets, impliceert dit 200 leerlingen per kolom, dus een minimum van 67 leerlingen per X in Figuur 1. Er blijven dan 134 proefpersonen per correlatie tussen toetsen over, wat voldoende is voor het schatten van een correlatie maar te weinig voor het doen van factoriële analyses over toetsen heen. Het totale aantal afnamen komt dan op ongeveer 800 leerlingen.
Daar we ook willen nagaan in hoeverre twee verschillende toetsen dezelfde vaardigheid meten (tevens een voorwaarde om toetsscores op een zinnige wijze in elkaar om te kunnen zetten), moeten we minimaal 200 leerlingen per rij rekenen. Dat betekent dat elke X in Figuur 1 overeenkomt met 100 leerlingen. Voor Figuur 1 zouden we dan op een steekproef van 1200 leerlingen uitkomen. Daarnaast moet elke toets door een groep proefpersonen tweemaal gemaakt worden om de test-hertestbetrouwbaarheid te bepalen. Duidelijk is dat we zo op zeer grote aantallen individuele afnamen uit zouden komen.
Om de kosten binnen de perken te houden is er daarom voor een andere oplossing gekozen. Door ervoor te zorgen dat leerlingen volkomen aselect worden toegewezen aan afnamecondities in het blokdesign, kunnen we de missende waarden in het blokdesign beschouwen als 'missing completely at random'. Zodoende kunnen we 'full information maximum likelihood' analyses uitvoeren alsof elke leerling in de steekproef elk toetspakket gemaakt heeft. Uitgaande van de eerder genoemde 200 proefpersonen als ondergrens voor factoranalyse, en uitgaande van het blokdesign in Figuur 1, moeten we 800 informatiepunten realiseren (4 toetspakketten, elk door 200 leerlingen gemaakt). Delen we deze 800 informatiepunten door de

zes afnamecondities, dan zien we dat elke conditie 800/6=133 informatiepunten moet opleveren. Elke afnameconditie betreft twee toetspakketten, dus elke X in Figuur 1 moet dan ongeveer 70 afnamen weerspiegelen. In totaal komen we dan op 840 individuele afnamen (maar twee per individu, dus levert dit 420 proefpersonen), zonder de test-hertestdata.
Voor de test-hertest-analyses moeten per toetspakket minimaal 80 leerlingen het pakket twee keer maken. Dit komt neer op 640 extra afnamen (320 extra leerlingen). De goedkopere optie om een deel van de steekproef een derde half uur te toetsen voor het bepalen van de hertestbetrouwbaarheid, zou betekenen dat dit deel van de leerlingen anderhalf uur getoetst wordt, wat ons te lang leek. Zo komen we dus uit op een steekproef van 740 leerlingen en 1480 afnamen van een half uur elk. Voordeel van de aparte steekproef voor de test-hertest is dat we de ondergrens van 200 leerlingen per factoranalyse in dit design met test-hertest leerlingen ophogen naar 280 leerlingen per toets (zie Figuur 2). Een ander voordeel hiervan is dat we ons enigszins indekten tegen uitval van leerlingen gaandeweg het onderzoek (ziekte, verhuizen, etc.). Daarnaast houdt de ondergrens van 200 leerlingen geen rekening met de vele condities die we eigenlijk moeten onderscheiden. Als we willen nagaan in hoeverre resultaten verschillen naar leeftijd, leerlinggewicht, thuistaal, enz. is een ondergrens van 280 zelfs nog zuinig.

Op grond van het bovenstaande is besloten om een steekproef van 740 leerlingen te trekken. Om voldoende spreiding te krijgen in de achtergrondvariabelen zijn we uitgegaan van een steekproef van minimaal 34 scholen waarbij we minstens één groep (gemiddeld 22 leerlingen) per school toetsen (34x22=748). Toewijzing van leerlingen aan toetspakketten heeft, zoals gezegd, volkomen aselect plaatsgevonden. Daar we voor deze studie scholen benaderen die al vaak aan onderzoek meedoen vanwege de proportie achterstandsleerlingen op school en die ook al aan het PRIMA-onderzoek meedoen, en omdat we ook nog eens veel vragen, hebben wij de scholen voor deelname een financiële vergoeding gegeven.

Figuur 2 Aantallen geplande afnamen (cursief) met vier toetspakketten en aantallen proefpersonen (N= aantal leerlingen; n= aantal afnames)

		Toetspakketten								
afnamecondities	N	1	2	3	4	1	2	3	4	n
A	70	70	70							140
B	70	70		70						140
C	70	70			70					140
D	70		70	70						140
E	70		70		70					140
F	70			70	70					140
test-hertest pakket 1	80	80				80				160
test-hertest pakket 2	80		80				80			160
test-hertest pakket 3	80			80				80		160
test-hertest pakket 4	80				80				80	160
Totaal	740	290	290	290	290	80	80	80	80	1480

2.4 Afnameprocedures

De afnamen zijn gehouden bij leerlingen in de groepen één en twee van reguliere basisscholen. De hertest is bij de leerlingen die hiervoor in aanmerking komen (zie Figuur 2) na ongeveer twee weken gehouden[5]. Een aantal van 80 leerlingen per toets is voldoende om de test-hertestbetrouwbaarheid te bepalen[6].

[5] De overweging hiervoor is dat een hertest niet zo lang na de eerste afname moet plaatsvinden dat er een substantiële groei in de gemeten vaardigheid kan zijn opgetreden, maar ook weer niet zo snel dat herinneringseffecten de resultaten op de hertest aanzienlijk kunnen beïnvloeden.

[6] Als we b.v. 100 proefpersonen een hertest laten maken en deze komt uit op .35, dan ligt het 95%-betrouwbaarheidsinterval van deze test-hertest tussen .17 en .51. Van belang is echter dat we een test-hertest vinden van ongeveer .80. Berekenen we dan met 100 proefpersonen het 95%-interval, dan komen we op een interval van .72 tot .86. Nemen we genoegen met 50 proefpersonen, dan zijn beide intervallen respectievelijk .078 - .572 en .672 - .882 (zie Hays, 1981; p. 467 en 660). Problematisch is dat dus bij een gevonden hertest van .80 en een steekproef van 50 leerlingen, we net niets kunnen beslissen aangaande test-hertest-betrouwbaarheid (vooral als we niet .80 maar bijvoorbeeld .70 vinden). Om deze reden ne-

Op voorhand voorzagen we dat een deel van de te toetsen leerlingen onvoldoende taalvaardig in het Nederlands zou kunnen zijn om de toetsen volledig te kunnen maken. De toetsen voorzien echter dit probleem en werken met instap- en afbreekregels. Zo kent de OBIS instapregels voor moeilijke toetsen: alleen leerlingen die voldoende hoog scoren op voorgaande subtoetsen krijgen deze moeilijke toetsen aangeboden. De Peabody stelt dat zodra een leerling negen fouten maakt in een set van 12 items, de toets moet worden afgebroken. In voorkomende gevallen krijgen leerlingen op alle niet gemaakte items dan een score die aangeeft dat ze die items fout gemaakt hebben. Zo kunnen deze leerlingen wel meegenomen worden in de factoranalyses. Alleen leerlingen die geen enkele vraag van een toets beantwoorden, kunnen niet meegenomen worden in de factoranalyses.

Er zijn voor het verrichten van de toetsafnamen 13 toetsleiders geselecteerd uit het daarvoor aangelegde toetsleiderbestand van PRIMA. Een bijkomend voordeel hiervan is, dat we toetsleiders konden selecteren op gebleken kwaliteit. Elke toetsleider is geïnstrueerd voor het afnemen van elk toetspakket, daar de toewijzing van leerlingen aan toetspakketten aselect is geschied en dus iedere toetsleider alle vier de toetspakketten moest afnemen binnen elke klas die werd getoetst. Daarnaast zijn op enkele scholen ook leerkrachten geïnstrueerd. Deze leerkrachten hebben een deel van de toetsen in de test-hertestconditie afgenomen om na te gaan of de data verzameld via afnamen door leerkrachten andere eigenschappen (gemiddelden, betrouwbaarheid, e.d.) vertonen dan die welke zijn verzameld via testleiders. Dit om na te gaan of de toetsen eventueel ook door leerkrachten zouden kunnen worden afgenomen. De leerkrachten hebben voor deze werkzaamheden een gelijke beloning ontvangen als de toetsleiders.

men wij 80 leerlingen voor de test-hertest (intervallen van bovenstaande voorbeelden zijn dan .14 - .53 en .71 - .87).

3 Analyses

3.1 Kwantitatieve analyses

Om de geschiktheid van de toetsen te evalueren voor het selecteren van de minst taalvaardige leerlingen, zijn verschillende kwantitatieve analyses verricht. Deze analyses moeten onder meer zicht geven op de betrouwbaarheid en validiteit van de verschillende toetsen.
De betrouwbaarheid van de toetsen is op verschillende manieren bepaald. Ten eerste is gekeken naar de homogeniteit (Cronbach's alfa) van elke (sub)toets. Deze dient voldoende hoog te zijn om somscores over de items te rechtvaardigen. De homogeniteitanalyses geven ook inzicht in de deugdelijkheid van de verschillende items in de toets. Verder zijn exploratieve en toetsende factoranalyses verricht[7]. Op basis van deze analyses zou ook besloten kunnen worden bepaalde items van de toetsen te verwijderen of om meer dan één somscore per (sub)toets te genereren.
Op grond van de exploratieve en toetsende factoranalyses per (sub)toets kan bepaald worden of de onderscheiden toetsen één of meerdere vaardigheden meten[8]. Deze gegevens staan ook in verband met de validiteit van de toetsen. Als toetsen meerdimensioneel blijken, moet de betrouwbaarheid eigenlijk per dimensie worden bepaald en zal ook een keuze gemaakt moeten worden in welke mate dimensies de totaalscore mogen beïnvloeden. Immers, als blijkt dat de toetsen meerdere dimensies meten, moet een beredeneerde keuze gemaakt worden voor het relatieve gewicht van elk der dimensies voor het bepalen van de toekenning van het predicaat 'zeer taalzwak'. Stel dat bijvoorbeeld twee dimensies worden gevonden die respectievelijk te benoemen zijn als 'woordenschat' en 'grammatica'. De groep die zeer laag scoort qua woordenschat hoeft dan niet gelijk te zijn aan de groep die zeer zwak scoort qua grammaticale vaardigheid (cf. Bishop, 2004).

---

[7] De factoranalyses per toets geven overigens vergelijkbare informatie op itemniveau als homogeniteitsanalyses. Ook geven ze een schatter van de betrouwbaarheid van de toets (de proportie verklaarde variantie per factor c.q. de eigenwaarde).

[8] Doordat scores op toetsitems alleen de waarden 'goed' en 'fout' kennen, kunnen we de factoranalyses alleen uitvoeren op tetrachorische correlaties, waardoor het aantal in een factoranalyse gevonden factoren iets te hoog uit zal vallen (Muthén & Muthén, 1999). Bij het interpreteren van de factoranalyses zullen we hiermee rekening houden.

De dimensionaliteit van de toetsen is van belang voor de validiteitkwestie, voor het bepalen van grensscores en voor een eventuele equivalering van de scores op verschillende toetsen die hetzelfde construct meten.
Naast de homogeniteit en de factoriële structuur is de test-hertestbetrouwbaarheid bepaald. Deze index is van belang om na te gaan of de zeer jonge leerlingen constant presteren op de toetsen. Naast instabiliteit inherent aan de gebruikte toets is ook de stabiliteit van het toetsgedrag van leerlingen hier in het geding. Het is niet ondenkbaar dat de zeer jonge leerlingen bij verschillende toetsafnamen niet even goed hun best doen, even gemotiveerd of geconcentreerd zijn, etc. Ook worden de test-hertest data gebruikt om uit te maken hoe betrouwbaar of stabiel per gebruikte toets de beslissing is waarmee we de leerlingen willen oormerken die het laagst taalvaardig zijn. Aan de hand van kruistabellen laten we zien in hoeverre toetsen bij herhaalde afname dezelfde leerlingen aanmerken als taalzwak. Bij deze analyses wordt kappa berekend als overeenstemmingmaat[9].
Omdat we na willen gaan of het mogelijk is om jonge, taalzwakke leerlingen te identificeren met de toetsen, zijn de betrouwbaarheidsanalyses ook nog apart verricht voor leerlingen in de steekproef jonger dan vijf en jonger dan zes. Omdat in een qua leeftijd homogenere groep de varianties zouden kunnen afnemen, is het denkbaar dat de betrouwbaarheid van de metingen voor jonge leerlingen lager uitvalt dan voor de gehele steekproef van vier tot acht jaar. Ook als jongere leerlingen instabieler zijn in hun toetsprestaties door bijvoorbeeld een gebrek aan concentratie of een wisselende motivatie, zal dit blijken uit lagere betrouwbaarheden voor beide voornoemde substeekproeven van jongste leerlingen.

Zoals hierboven al werd aangehaald, moet om na te gaan of de verschillende toetsen hetzelfde meten, gekeken worden naar de factoriële structuur van de toetsen. Deze analyses betreffen de validiteit van de toetsen. Per (sub)toets is onderzocht hoeveel factoren de (sub)toets meet. Daarnaast wordt gekeken in welke mate de verschillende toetsen hetzelfde meten, ofwel of de somscores van de verschillende subtoetsen te passen zijn op eenzelfde factor of op combinaties van meerdere factoren. Op deze wijze wordt duidelijk in welke mate toetsen gelijke dan wel

[9] Probleem is wel dat de te kiezen grensscore gericht op het benoemen van respondenten als taalzwak in feite arbitrair is. Er is geen objectief criterium om tot het predikaat 'taalzwak' te komen. We zullen dus een percentage moeten kiezen (norm- i.p.v. criterion referenced), b.v. de 25% meest taalzwakke kinderen in het reguliere onderwijs.

verschillende aspecten van taalvaardigheid meten. Zo kan een overzicht gemaakt worden waarin getoond wordt in welke mate verschillende toetsen vergelijkbare vaardigheden meten en in hoeverre ze unieke variantie bezitten. Deze gegevens zijn van belang bij het kiezen van een toets voor groep één of twee. Als blijkt dat de toetsen verschillende aspecten van taalvaardigheid meten, moet een beredeneerde keuze gemaakt worden voor welke vaardigheden men wil meten. Een leerling die laag scoort op bijvoorbeeld vocabulaire hoeft niet laag te scoren op fonologische vaardigheid.

Voor de toetsende factoranalyses zijn naast de beschrijvingen van de toetsen door de constructeurs, de uitkomsten van de exploratieve factoranalyses en de ideeën van de onderzoekers als leidraad genomen. Zoals in Emmelot en Van Schooten (2005) is te lezen, worden bij sommige toetsen door de constructeurs al meerdere dimensies onderscheiden. Indien we meerdere dimensies vinden in de toetsen[10], moet een beredeneerde keuze gemaakt worden over het relatieve gewicht dat elk der dimensies krijgt bij het bepalen van de taalachterstand. Om de gevonden dimensies te kunnen interpreteren, zal op grond van de toetsitems en de toetsbeschrijvingen een poging gedaan worden de dimensies inhoudelijk te benoemen.

Verder is met behulp van de factoranalyses nagegaan of verschillende toetsen gelijkwaardig zijn en of er vergelijkbare grensscores op verschillende toetsen gemaakt kunnen worden ('equilibreren' of 'equivaleren'). Op deze wijze is na te gaan of er wellicht verschillende gelijkwaardige toetsen te gebruiken zijn voor de bepaling van taalachterstanden. Een voorwaarde hiervoor is uiteraard wel dat de verschillende toetsen dezelfde vaardigheid meten, wat moet blijken uit de factoranalyses[11].

---

[10] Als fit-index bij de toetsende factoranalyses zal naast de Chi-kwadraat, CFI, TLI en SRMR ook de Root mean square error of approximation (RMSEA) gebruikt worden, daar de Chi-kwadraat in sociaal-wetenschappelijk onderzoek te streng is (MacCallum, Browne & Sugawara, 1996). Wel zal indien mogelijk de passing van geneste modellen met behulp van de verschillen in Chi-kwadraat en vrijheidsgraden worden getoetst. Bij de toetsende factoranalyses per subtoets op itemniveau (de meetmodellen) wordt vanwege de dichotome onafhankelijke variabelen (de items) de WLSMV schattingsmethode gebruikt en is vergelijking van geneste modellen alleen mogelijk op grond van RMSEA, CFI en TLI.

[11] Met behulp van Rasch-schalingen zijn in dat geval de scores op verschillende toetsen op één schaal te zetten (Hambleton & Swaminathan, 1985).

Nadat op de bovenbeschreven wijze is bepaald welk model de somscores op de (sub)toetsen adequaat beschrijft, zal nagegaan worden of toetsen partijdig zijn tegen specifieke subgroepen van leerlingen. Deze analyses tonen of de toetsen in verschillende deelpopulaties gelijke dan wel verschillende vaardigheden meten en als ze gelijke vaardigheden meten, of ze dat op dezelfde schaal doen. Een ongelijke schaal betekent dat leerlingen die tot verschillende groepen behoren (b.v. jongens en meisjes) en die een gelijke vaardigheid hebben toch een verschillende score krijgen.

De partijdigheidanalyses worden verricht aan de hand van tweegroepenmodellen. Eerst wordt een tweegroepenmodel gefit met in beide groepen ongelijke residuen, gelijke intercepten en gelijke factorladingen. In dit model wordt er dus van uitgegaan dat in beide groepen dezelfde vaardigheid gemeten wordt op dezelfde schaal, maar mogelijk wel met een verschillende mate van betrouwbaarheid. Dan wordt ditzelfde model nogmaals gefit, maar nu worden naast de residuele varianties ook de intercepten in beide groepen apart geschat. Als dit model significant beter past dan het model met alleen ongelijke residuele varianties, dan is aangetoond dat in beide groepen op verschillende schalen gemeten wordt. Wel gaat dit model nog uit van het meten van eenzelfde construct in beide groepen. Vervolgens wordt hetzelfde model gepast, maar nu met in beide groepen apart geschatte residuele varianties, intercepten en factorladingen. Als dit model significant beter past dan het voorgaande model, dan is aangetoond dat in beide onderscheiden groepen verschillende constructen gemeten worden. Als dit laatste het geval is, is het gebruik van de toetsen in de onderscheiden deelpopulaties niet goed te verdedigen. Men meet dan immers verschillende constructen, al komen de scores van hetzelfde instrument[12].

Een ander aspect dat zowel de betrouwbaarheid als de validiteit betreft, is hoe de verdeling van toetsscores eruit ziet binnen de doelpopulatie. Toetsen die sterke plafond- of bodemeffecten vertonen lijken minder geschikt voor het identificeren van taalzwakke leerlingen. Uiteraard zijn plafondeffecten minder erg dan bodemeffecten, gegeven het doel van het onderzoek. We zoeken immers toetsen die

[12] Ter verduidelijking; stel een toets meet de vaardigheid in het maken van sommen. En stel dat een deel van de getoetste leerlingen sommige opgaven niet kan maken omdat er voor hen onbekende woorden in de vragen staan. In dat geval meet de toets bij deze laatste groep leerlingen (ook) woordkennis. De scores zijn dan voor de beide groepen getoetste leerlingen niet meer op één schaal te zetten, er zijn verschillende constructen gemeten.

taalzwakke leerlingen kunnen identificeren. We gebruiken boxplots om plafond- en bodemeffecten te detecteren.
Verder is aan de hand van kruistabellen voor alle combinaties van twee (sub)toetsen nagegaan in hoeverre ze overeenstemmen in het oormerken van de 25% minst taalvaardige leerlingen. Weer is kappa als overeenstemmingmaat berekend.
Bij de implementatie van de taaltoetsen ten bate van het selecteren van zeer taalzwakke leerlingen, kan een kostenbesparing worden verkregen als er een voorselectie gemaakt kan worden op grond van enige eenvoudig te scoren objectieve leerlingvariabelen. Om dit te kunnen realiseren, is met logistische regressie per (sub)toets nagegaan in hoeverre leerlingkenmerken de ja/nee-beslissing (of er sprake is van extreme taalachterstand) voorspelt (zie ook Bosker & Guldemond, 2004). Als predictoren gebruikten we thuistaal, sekse, leerlinggewicht en de door de eigen leerkracht ingeschatte spreek- en luistervaardigheid Nederlands van de leerling. De verdeling van de toetsscores over de groepen gevormd door thuistaal en leerlinggewicht worden getoond met behulp van boxplots. Het onderzoek geeft zo informatie over de toegevoegde waarde van de toetsuitkomst ten opzichte van achtergrondkenmerken van de leerling.
Een andere kwantitatieve analyse betreft de kwaliteit van de ingezette toetsleiders. Door middel van verschillende analyses is nagegaan in hoeverre toetsleiders invloed hebben op de verkregen toetsscores. Aan de hand van t-toetsen is nagegaan of de leerkrachten en toetsleiders verschillen in strengheid. Daarnaast is ook gekeken of er verschillen zijn in de betrouwbaarheid van de scores indien verzameld via leerkrachten dan wel via toetsleiders.

3.2 Kwalitatieve analyses

Het kwalitatieve deel van de studie moet antwoord geven op de vragen naar de praktische bruikbaarheid van de toetsen en naar de vormgeving van de praktische organisatie van een te implementeren toetssysteem. Vragen die hierbij aan de orde komen, zijn welke knelpunten blijken tijdens toetsafnamen, hoe de registratie van de gegevens het best kan plaatsvinden en hoe toetsleidereffecten geminimaliseerd kunnen worden. Ook moet een inschatting gemaakt worden van de kosten die verschillende opties met zich meebrengen.

Om na te gaan welke praktische problemen er optreden bij het afnemen van de toetsen, hebben toetsleiders en leerkrachten een evaluatieformulier ingevuld voor de door hen verrichte toetsafnamen. Ook is informatie verzameld waaruit kan worden afgeleid of toetsleiders en leerkrachten even geschikt zijn om toetsen af te nemen. In de evaluatieformulieren worden vragen gesteld over de getoetste leerling, het afnameproces, de duur van de afname, over de mate waarin de invuller van het evaluatieformulier zichzelf als competent ziet om de toetsen af te nemen, welke oplossingen voor gesignaleerde problemen men zelf ziet, hoe de motivatie en concentratie en dergelijke van de leerlingen is, etc. Ook is gevraagd wat naar de indruk van de toetsleider de oorzaak is van een eventueel niet geslaagde afname. Per toetspakket schreven toetsleiders na afloop van de afnamen een kort verslag over de problemen die zij per toets tegenkwamen.
Om de kosten in te kunnen schatten van de verschillende opties die volgen uit het onderzoek, is op grond van de verzamelde gegevens een schatting gemaakt van de verschillende kosten die aan de betreffende optie verbonden zijn. Vragen die hierbij aan bod komen, zijn bijvoorbeeld: Hoe moeten de gegevens verwerkt en geregistreerd worden (bijvoorbeeld afname per p.c.)? Door wie moeten de gegevens verwerkt worden en wat kost dat ongeveer?
Ook is bij schoolleiders een vragenlijstje afgenomen om te vragen hoe men tegen het toetsen van jonge leerlingen aankijkt. Op deze wijze hopen we enig zicht te krijgen op de acceptatie in het veld van de verschillende mogelijke wijzen om de taaltoets voor het vroegtijdig signaleren van taalachterstand in te voeren.

4 Resultaten

4.1 Steekproeftrekking en respons

De te onderzoeken groep betreft, zoals gezegd, leerlingen in groep één en twee van de basisschool. Om verschillende redenen is het wenselijk af te zien van beperking van de afname tot leerlingen waarvan vooraf wordt verwacht dat zij tot de taalzwakke groep behoren. In de eerste plaats is het wenselijk dat de geschiktheid van de toetsen – mede gezien de vraag naar mogelijke diagnostische toepassingen – wordt bepaald aan de hand van afname bij een bredere groep dan alleen (zeer) taalzwakke leerlingen (zie hierboven).
Om praktische redenen (kosten, organisatie) was het wenselijk om de leerlingen voor het onderzoek vooral te zoeken op scholen die relatief veel achterstandsleerlingen hebben. Er is gekozen voor benadering van scholen die deel hebben genomen aan de laatste meting van het PRIMA-cohortonderzoek. Het voordeel daarvan is dat we daardoor konden beschikken over een aantal achtergrondkenmerken van de scholen en van de samenstelling van hun leerlingenpopulatie. Bovendien omvat het PRIMA-bestand naast een representatieve groep scholen ook een aanvullende steekproef van scholen met relatief veel leerlingen uit achterstandsgroepen. Bij de steekproeftrekking is ook gebruik gemaakt van deze aanvullende steekproef.
Om dubbele afname en mogelijke vertekening van resultaten te voorkomen, zijn scholen die de OBIS-toets zelf al afnemen, uit het bestand verwijderd. Uit het resterende bestand is vervolgens een random selectie gemaakt van 150 scholen. Deze groep is aangeschreven via een brief aan de directie. In de brief wordt het doel van het onderzoek beschreven en gevraagd om medewerking aan het onderzoek. Bij deelname is een tegemoetkoming in de vorm van een cadeaubon in het vooruitzicht is gesteld. Tevens is gevraagd of scholen eventueel zelf (een deel van de) toetsen zouden willen afnemen of dit wilden overlaten aan toetsleiders.
Van de 150 aangeschreven scholen hebben er 74 gereageerd. Daarvan hebben 51 scholen (34 procent van de groep die is benaderd) aangegeven aan het onderzoek te willen deelnemen. Het gaat om 127 groepen met in totaal 2368 leerlingen (van 16 tot 135 leerlingen per school). Op basis van de aanmeldingen van scholen voor deelname en op basis van de aantallen leerlingen in groep één en twee van deze

scholen, is besloten een selectie van 41 scholen te maken voor het onderzoek. Bij deze selectie zijn drie criteria gehanteerd:

- er is gestreefd naar voldoende vertegenwoordiging van scholen met veel allochtone achterstandleerlingen;
- er is gestreefd naar een evenredige verdeling van groepen één, gemengd één en twee en twee;
- scholen met een gering aantal leerlingen in de genoemde groepen zijn uit kostenoverwegingen buiten het onderzoek gelaten[13].

Kenmerken responderende scholen

Hoewel er niet gestreefd is naar een aselecte steekproef, is wel nagegaan in hoeverre de scholen die weigerden aan het onderzoek deel te nemen afwijken van de scholen die wel deelnamen. Ook zijn enkele kenmerken van de deelnemende scholen in kaart gebracht. De tabellen 1 tot en met 4 geven een overzicht van een aantal kenmerken van de scholen in de getrokken steekproef en van de scholen die daadwerkelijk aan het onderzoek hebben deelgenomen. De schoolcompositie, dat wil zeggen de samenstelling van de leerlingenpopulatie binnen de scholen, verschilt niet veel tussen beide groepen (tabel 1). Het belangrijkste verschil is dat scholen met een leerlingenpopulatie die voor minimaal de helft bestaat uit allochtone achterstandsleerlingen, iets minder zijn vertegenwoordigd in de groep die deelneemt en scholen met veel autochtone achterstandsleerlingen iets sterker zijn vertegenwoordigd.

[13] Het aantal getoetste leerlingen in de geselecteerde scholen is daardoor minimaal 16.

Tabel 1 Schoolcompositie

	Steekproef (n=150)		Deelname (n=40[14])	
	aantal	%	aantal	%
≥ 50% ouders laag opgeleid Turks/Marokkaans	9	6.0	2	5.0
≥ 50% ouders laag opgeleid allochtoon	19	12.7	4	10.0
≥ 50% ouders laag opgeleid autochtoon	9	6.0	5	12.5
heterogeen, max. 33% achterstandsgroepen	58	38.7	13	32.5
zeer heterogeen, 33-50% achterstandsgroepen	27	18.0	9	22.5
overwegend kansarm, ≥50% achterstandsgroepen	28	18.7	7	17.5

De wat sterkere vertegenwoordiging van scholen met meer autochtone achterstandsleerlingen hangt samen met een ondervertegenwoordiging van scholen die tot deelname bereid zijn in zeer sterk stedelijke gebieden en een oververtegenwoordiging van scholen in gebieden die niet stedelijk zijn (tabel 2).

Tabel 2 Mate van stedelijkheid

	Steekproef (n=150)		Deelname (n=41)	
	aantal	%	aantal	%
zeer sterk stedelijk	28	18.7	3	7.3
sterk stedelijk	43	28.7	13	31.7
matig stedelijk	33	22.0	9	22.0
weinig stedelijk	31	20.7	9	22.0
niet stedelijk	15	10.0	7	17.1

[14] Eén school gaf hierover geen gegevens.

De taal die thuis wordt gesproken, is – volgens de informatie van de leraar – in de steekproef gemiddeld in 57 procent van de gevallen Nederlands. Bij de deelnemers aan het onderzoek geldt dat gemiddeld voor 63 procent (zie tabel 3). De groep leerlingen die volgens de leraar thuis dialect spreekt, al dan niet in combinatie met Nederlands, maakt in beide groepen gemiddeld 12 procent van het totaal uit. Leerlingen die thuis een buitenlandse taal spreken, eventueel in combinatie met Nederlands, zijn iets ondervertegenwoordigd in de groep die deelneemt. Het betreft respectievelijk gemiddeld 31 procent in de steekproef en 25 procent bij de deelnemers. Ook dit kan worden verklaard doordat in verhouding meer bereidheid tot deelname aan het onderzoek bestond bij scholen op het platteland dan bij scholen in de grote steden.

Tabel 3 Taal die – volgens de leraar – bij de leerlingen thuis wordt gesproken; percentages

	Steekproef (150 scholen)	Deelname (36 scholen)
thuistaal Nederlands	57	63
thuistaal dialect of Fries	12	12
thuistaal allochtoon	31	25

Zoals al bleek uit tabel 1, is in vergelijking met de steekproef het aandeel allochtone achterstandsleerlingen iets ondervertegenwoordigd en het aandeel autochtone achterstandsleerlingen iets oververtegenwoordigd in de groep die aan het onderzoek deelneemt. Dit is ook te zien in de verdeling van de verschillende groepen gewichtenleerlingen volgens het onderwijsachterstandenbeleid (tabel 4). Het totale aandeel gewichtenleerlingen is ongeveer de helft, zowel in de steekproef als bij de deelnemers.

Tabel 4 Over scholen gemiddelde percentages gewichtenleerlingen

	Steekproef taaltoets (n=150)	Deelname (n=40)
autochtone achterstandsleerlingen (0,25)	17.3%	21.1%
allochtone achterstandsleerlingen (0,90)	34.2%	27.3%
totaal gewichtenleerlingen (0,25 t/m 0,90)	51.9%	49.0%

Wat de verdeling naar provincie betreft, valt op dat er vooral veel scholen uit Noord-Brabant onder de deelnemers zijn. Het betreft acht van de 41 scholen. Zuid-Holland volgt met zeven scholen en Overijssel met vijf. Vooral Utrecht is zwak vertegenwoordigd, met maar één school.

Conclusie

Op grond van de vergelijking op een aantal achtergrondkenmerken tussen de groep scholen die is aangeschreven en de groep die aan het onderzoek heeft deelgenomen, kan het volgende worden geconcludeerd:

- er is ten opzichte van de steekproef sprake van een ondervertegenwoordiging van scholen uit zeer sterk stedelijke gebieden en een oververtegenwoordiging van scholen uit niet stedelijke gebieden;
- de scholen die aan het onderzoek hebben deelgenomen, tellen relatief minder allochtone en thuis een vreemde taal sprekende leerlingen dan de scholen in de steekproef;
- het aandeel leerlingen uit achterstandsgroepen ('gewichtenleerlingen') wijkt in de groep deelnemende scholen niet noemenswaardig af van de scholen in de steekproef.

4.2 Afname van de toetsen

Bij nagenoeg alle geselecteerde scholen is gekozen voor afname van toetsen in één groep. Bij twee (grote) scholen is gekozen voor toetsafname in twee groepen. De te toetsen groepen zijn at random geselecteerd. Er zijn twee groepen afnamecondities, reguliere condities en test-hertestcondities. In de reguliere groep

afnamecondities zijn per leerling twee verschillende toetsen afgenomen. Het betreft de volgende zes condities (zie ook figuur 1 en 2):
A – OBIS en TvK;
B – OBIS en TAK;
C – OBIS en Peabody;
D – TvK en TAK;
E – TvK en Peabody;
F – TAK en Peabody.

In de test-hertestgroepen is per school bij iedere geselecteerde leerling één en dezelfde toets twee maal afgenomen met enkele weken tussen beide afnamen. In de test-hertestgroepen is per klas de helft van de leerlingen (test en hertest) getoetst door een toetsleider en de andere helft door de leerkracht. Uit praktische overwegingen zijn voor de test-hertestgroep andere scholen geselecteerd dan voor de reguliere toetscondities. Eveneens uit praktische overwegingen is ervoor gekozen alleen leerkrachten zelf toetsen te laten afnemen in de test-hertestconditie. Door in de scholen in deze conditie alle leerlingen dezelfde toets te laten maken, is voorkomen dat de leerkracht zich de afname van vier toetsen eigen moest maken. Om zicht te houden op de eventuele invloed van de leerkracht op de toetsresultaten, heeft de leerkracht in de test-hertestconditie bij een aselect gekozen helft van de leerlingen zowel de test als de hertest afgenomen en de toetsleider heeft de test en hertest afgenomen bij de andere helft van de groep.
Aan de test-hertestconditie zijn 18 scholen at random toegewezen, uit de groep scholen die bereid was zelf toetsen af te nemen[15]. In deze condities zijn vervolgens at random steeds vier of vijf scholen toegewezen aan de vier toetsen[16]. De overige 23 scholen zijn toegewezen aan de reguliere toetscondities (waar de combinaties A t/m F zijn afgenomen).

Voor de random toewijzing van de toetscombinaties aan leerlingen (in de reguliere toetscondities) en van afname door een toetsleider of door de leerkracht (in de

[15] Driekwart van de deelnemende scholen was bereid om zelf toetsen af te nemen.
[16] Of er 4 of 5 scholen aan een bepaalde toets zijn toegewezen, is afhankelijk van het aantal leerlingen dat in de desbetreffende scholen – volgens opgave van de school – kon worden getoetst. Bleek het totale aantal na random toewijzing van 4 scholen te klein, dan werd er een extra school toegewezen.

test-hertestcondities) is de volgende procedure gevolgd: Per school is een lijst gemaakt met leerlingnummers, waarbij de door de scholen aangegeven aantallen het uitgangspunt vormden. Vervolgens zijn de verschillende mogelijkheden at random verdeeld over leerlingnummers. Daarbij is gebruik gemaakt van een generator van random-getallen[17].
Voor de afname zijn dertien ervaren toetsleiders ingeschakeld, die ook werden ingezet om taal- en rekentoetsen af te nemen in het kader van het PRIMA-onderzoek. Voor de afname van de toetsen zijn handleidingen geschreven. Bovendien is een instructiedag georganiseerd waarop alle toetsleiders door de onderzoekers en de coördinator van het veldwerk zijn geïnformeerd over het onderzoek en zijn geïnstrueerd in de afname van de toetsen. De toetsen TvK, TAK en Peabody zijn op papier beschikbaar, terwijl de OBIS-toets met behulp van een CD-ROM en p.c. is afgenomen.

4.3 Aantallen afgenomen toetsen

Zoals aangegeven, zijn 41 scholen geselecteerd voor deelname. Er zijn toetsen afgenomen bij 891 leerlingen. Daarvan zaten 515 leerlingen in de reguliere groep afnamecondities (A t/m F) en elk van deze leerlingen heeft dus twee verschillende toetsen gemaakt. In de test-hertestconditie zaten 376 leerlingen (zie tabel 5).

[17] http://calculators.stat.ucla.edu/perm.php

Tabel 5 Toetscondities en aantallen afnamen

	aantal leerlingen
A) OBIS + TvK	86
B) OBIS + TAK	92
C) OBIS + Peabody	83
D) TvK + TAK	84
E) TvK + Peabody	88
F) TAK + Peabody	82
L) Test-hertest door leerkracht	184
T) Test-hertest door testleider	192
Totaal	891

Tabel 6 geeft een overzicht van de aantallen afnamen per toets. Deze liggen voor elke toets rond 350, de tweede afname in de hertestconditie niet meegerekend. Het totaal komt uit op 1406 gemaakte toetsen. Bij dit aantal is de tweede afname in de hertestconditie niet meegeteld. Doen we dat wel, dan komen er 376 (tweede) afnamen bij, wat in totaal neerkomt op 1782 toetsafnamen. Dit is aanzienlijk meer dan de geplande 1480 afnamen (zie figuur 2).

Tabel 6 Aantallen afgenomen toetsen (zonder tweede afnamen voor de hertest)

	aantal conditie A t/m F	aantal test-hertest (1e afname)	aantal totaal
OBIS	261	89	350
TvK	258	97	355
TAK	258	94	352
Peabody	253	96	349
Totaal	1030	376	1406

Van de 43 geselecteerde groepen is bijna de helft, 21 groepen, een combinatiegroep (groep één en 2). Daarnaast zijn er negengroepen één en 13 groepen twee (zie tabel 7).

Tabel 7 Groepen waarin toetsen zijn afgenomen

	aantal groepen	aantal leerlingen
groep 1	9	186
groep 1/2	21	469
groep 2	13	236

4.4 Achtergrondkenmerken van de leerlingen

Voor elke deelnemende groep heeft de toetsleider gezorgd dat een overzichtsformulier is ingevuld door of samen met de leerkracht. Daarin zijn per leerling een aantal gegevens opgenomen die relevant kunnen zijn voor het taalvaardigheidniveau. Ook is de leerkracht gevraagd een inschatting te maken van het taalvaardigheidniveau van elke leerling.

Een klein deel van de leerlingen, drie procent, was net in groep één gestart en werd ingedeeld in de categorie 'groep 0'. Bijna de helft, 49 procent, zat tijdens de toetsafname in groep één. Een bijna even grote groep, 46 procent, zat in groep twee en twee procent zat in de verlengde kleuterperiode ('groep 2+'). De verhouding jongens-meisjes was bijna 50/50. Het geboortejaar was overwegend 2000 (40%) of 2001 (39%). De groep die ouder was, vormde 13 procent van het totaal. Deze leerlingen zijn bijna allemaal in 1999 geboren (op drie na). De jongste leerlingen, zeven procent, zijn in 2002 geboren.

Het geboorteland is in 97 procent van de gevallen Nederland. Het geboorteland van de ouders laat meer variatie zien (zie tabel 8). Ruim twee derde is in Nederland geboren. Daarna worden Marokko (rond 10%) en Turkije (rond 7%) het meest genoemd.

Tabel 8 Geboorteland van vader en moeder van de leerlingen (percentages); 891 leerlingen

	vader	moeder
Nederland	66	69
Marokko	10	9
Turkije	7	6
Afrika	2	3
Suriname	2	1
Antillen	1	1
Azië	2	2
Oost-Europa	1	1
Midden-Oosten	2	2
overig	3	4
onbekend / niet ingevuld	5	1

Van 825 van de 891 leerlingen is door de school de wegingsfactor doorgegeven zoals die destijds werd gehanteerd in het kader van het onderwijsachterstandenbeleid. Van de deelnemende leerlingen behoort 16 procent tot de autochtone achterstandsleerlingen (.25-leerlingen) en 31 procent tot de allochtone achterstandsleerlingen (.90-leerlingen). Iets meer dan de helft (53%) behoort niet tot een achterstandsgroep die in aanmerking komt voor een wegingsfactor (.00-leerlingen).
Het opleidingsniveau van de ouders is bij bijna de helft van de vaders en bij de helft van de moeders niet hoger dan lager beroepsonderwijs (tabel 9). Eén op de zeven ouders heeft een HBO- of universitaire opleiding.

Tabel 9 Opleiding van vader en moeder van de leerlingen (percentages); 891 leerlingen

	vader	moeder
maximaal lager onderwijs	14	19
maximaal LBO / VBO	31	30
maximaal MBO	33	34
HBO / universiteit	14	15
onbekend / niet ingevuld	8	2

Leerlingen spreken volgens hun leerkrachten thuis en met vriendjes vooral Nederlands. Het aantal dialectsprekers is beperkt. Ruim een vijfde van de deelnemende leerlingen spreekt volgens de leerkracht thuis een vreemde taal. Met vriendjes en vriendinnetjes wordt aanzienlijk vaker Nederlands gesproken dan thuis (tabel 10).

Tabel 10 Taal die de leerlingen overwegend thuis en met vriendjes/vriendinnetjes spreken (percentages); 891 leerlingen

	thuis	met vriendjes / vriendinnetjes
vooral Nederlands	60	74
vooral dialect / Fries	8	6
vooral een buitenlandse taal	21	5
combinatie Nederlands / dialect	4	3
combinatie Nederlands / buitenlands	2	2
onbekend / niet ingevuld	6	11

De leerkrachten hebben de taalvaardigheid Nederlands van hun leerlingen, apart voor verstaan/begrijpen en voor spreken, beoordeeld op een tienpuntsschaal (1 = heel laag; 10 = heel hoog). Voor verstaan/begrijpen is het gemiddelde 7.2 en de standaarddeviatie 1.4 (zie tabel 11). Deze inschatting is gemaakt voor 93 procent van de deelnemende leerlingen. De beoordelingen variëren van één tot en met

tien, 11% van de leerlingen krijgt een score lager dan zes en 60% krijgt een zeven of een acht.
Op het gebied van spreken liggen de inschattingen van het taalvaardigheidniveau iets lager (zie eveneens tabel 11). Het gemiddelde is 6.8 (met een standaarddeviatie van 1.6), 16% scoort een onvoldoende en de helft krijgt een zeven of acht.

Tabel 11 Inschatting van het taalvaardigheidniveau in het Nederlands; 826 leerlingen

score	verstaan / begrijpen		spreken	
	aantal	percentage	aantal	percentage
1	1	0	5	1
2	4	1	14	2
3	9	1	11	1
4	22	3	38	5
5	54	7	64	8
6	101	12	154	19
7	282	34	279	34
8	213	26	151	18
9	104	13	82	10
10	36	4	28	3

4.5 De betrouwbaarheid van de (sub)toetsen

Om de homogeniteit van de items te bepalen die samen één somscore op moeten leveren en dus bedoeld zijn om één vaardigheid te meten (ofwel bedoeld zijn als unidimensionele toets), zijn alfa's berekend. De resultaten van deze analyses staan in tabel 12 en in de Appendix (tabel A2 t/m A21). Bij het berekenen van de alfa's is ook gekeken naar de item-testcorrelaties. Indien bleek dat items een negatieve item-testcorrelatie hadden of geen variantie vertoonden door een plafond- of bodemeffect, zijn de alfa's ook een keer zonder deze items berekend. Daarnaast zijn alfa's berekend op zowel de gewone als op de hertestdata.

Uit de tabel blijkt dat de alfa's voor de verschillende (sub)toetsen variëren van .71 (rekenen OBIS sommen B, 1e afname) tot .98 (TAK passieve woordenschat, hertest). Voor toetsen op grond waarvan beslissingen worden genomen die individuele consequenties hebben, geldt volgens de COTAN (Evers et al., 2002) een minimale norm voor de betrouwbaarheid van .80. Een waarde van .90 wordt in de COTAN 'goed' genoemd. Als de toetsen dus gebruikt gaan worden om beslissingen te nemen die directe consequenties hebben voor de betrokken leerling, voldoen vrijwel alle subtoetsen aan de minimumnorm die de COTAN stelt. Alleen enkele subtoetsen van de OBIS vertonen een lagere alfa, maar veel subtoetsen van de OBIS hebben alfa's ruim boven .90. Verder zijn de OBIS-subtoetsen bedoeld om in combinatie met elkaar gebruikt te worden en niet als afzonderlijke toets.
Om beslissingen te nemen op groepsniveau (bijvoorbeeld schoolniveau) gelden lagere normen voor de betrouwbaarheid (.70). Duidelijk is dat voor beslissingen op groepsniveau de toetsen dus voldoende betrouwbaar zijn.
Een interessante vraag voor het onderhavige onderzoek is of de betrouwbaarheid van de toetsen verandert als we alleen de zeer jonge leerlingen beschouwen. Als de toetsen minder variantie vertonen bij jonge leerlingen of als de opgaven voor de jonge en minder vaardige leerlingen minder geschikt zijn, is het denkbaar dat daardoor de betrouwbaarheid van de toetsen voor de jongste leerlingen in negatieve zin afwijkt van de in tabel 12 gerapporteerde indices. Om deze reden wordt in tabel 13 de betrouwbaarheid (alfa) nogmaals gepresenteerd, maar nu apart voor de leerlingen van vier tot vijf jaar en de groep van vier tot zes jaar.

Uit de homogeniteitanalyses blijkt dat vrijwel alle (sub)toetsen voldoende betrouwbaar zijn. Zowel over de gehele steekproef (4-8 jaar) als voor de jongere leerlingen (4-6 jaar) vinden we te lage homogeniteiten bij de OBIS-toetsen rekenjargon, kennis van vormen, hoofdrekenen en sommen B. Voor de leerlingen van vier tot acht jaar blijkt ook nog leesbegrip onvoldoende homogeen. Bij leerlingen van vier tot vijf jaar oud blijken meer toetsen onvoldoende homogeen: de TVK en de OBIS-toetsen leesbegrip, kennis van letters, leesvaardigheid woorden, geheugen, sommen A, kennis van vormen, hoofdrekenen en sommen B.

Tabel 12 Homogeniteiten van de verschillende schalen. Variabelen, Cronbach's alfa per schaal, aantal proefpersonen (N), aantal items (n) en het aantal items dat verwijderd is vanwege negatieve item-testcorrelatie of ontbrekende variantie (exit)

Variabelen	alfa	N	n	exit
TvK, 1e afname	.84	341	31	1
TvK, hertest	.87	95	31	1
Peabody, 1e afname	.95	349	101	7
Peabody, hertest	.95	96	80	28
TAK - klankonderscheiding, 1e afname	.93	352	50	
TAK - klankonderscheiding, hertest	.95	94	50	
TAK - klankarticulatie, 1e afname	.86	352	45	
TAK - klankarticulatie, hertest	.87	94	45	
TAK - passieve woordenschat, 1e afname	.97	352	96	
TAK - passieve woordenschat, hertest	.98	94	96	
OBIS - passieve woordenschat, 1e afname	.95	344	31	
OBIS - passieve woordenschat, hertest	.96	82	31	
OBIS – leesbegrip, 1e afname	.73	344	10	
OBIS – leesbegrip, hertest	.81	82	10	
OBIS – klankarticulatie, 1e afname	.83	344	8	
OBIS – klankarticulatie, hertest	.89	82	8	
OBIS – klankonderscheiding, 1e afname	.92	344	9	
OBIS – klankonderscheiding, hertest	.95	82	9	
OBIS - kennis van letters, 1e afname	.97	344	27	
OBIS - kennis van letters, hertest	.97	82	27	
OBIS - leesvaardigheid woorden, 1e afname	.93	344	10	
OBIS - leesvaardigheid woorden, hertest	.94	82	10	
OBIS – leesvaardigheid zinnen, 1e afname	.95	344	10	
OBIS - leesvaardigheid zinnen, hertest	.95	82	10	
OBIS - leesvaardigheid teksten, 1e afname	.98	344	17	1
OBIS – leesvaardigheid teksten, hertest	.99	82	14	4
OBIS – rekenjargon, 1e afname	.72	344	7	
OBIS - rekenjargon, hertest	.78	82	7	

Vervolg tabel 12

Variabelen	alfa	N	n	exit
OBIS - geheugen voor aantallen, 1e afname	.84	344	4	
OBIS - geheugen voor aantallen, hertest	.90	82	4	
OBIS - sommen A, 1e afname	.86	344	8	
OBIS - sommen A, hertest	.92	82	8	
OBIS – kennis van getallen, 1e afname	.93	344	21	
OBIS – kennis van getallen, hertest	.93	82	21	
OBIS - kennis namen van vormen, 1e afname	.73	344	5	
OBIS - kennis namen van vormen, hertest	.79	82	5	
OBIS - hoofdrekenen, 1e afname	.75	344	8	
OBIS - hoofdrekenen, hertest	.82	82	7	1
OBIS - sommen B (met wiskundige symbolen), 1e afname	.71	344	12	4
OBIS - sommen B (met wiskundige symbolen), hertest	.81	82	11	5

Tabel 13 Homogeniteiten (Alfa) van de verschillende schalen, apart voor leerlingen van 4.4 tot 5 jaar en van 4.4 tot 6 jaar. (tussen haken het aantal proefpersonen). n = aantal items

Variabelen	Alfa 4.4 - 5 jr	Alfa 4.4 - 6 jr	n
TvK	.78 (21)	.81 (148)	31
Peabody	.94 (38)	.95 (180)	108
TAK - klankarticulatie	.88 (24)	.88 (161)	45
TAK - klankonderscheiding	.95 (24)	.93 (161)	50
TAK - passieve woordenschat	.97 (24)	.96 (161)	96
OBIS - passieve woordenschat	.97 (24)	.96 (152)	31
OBIS - leesbegrip	.67 (24)	.69 (152)	10
OBIS – klankarticulatie	.89 (24)	.87 (152)	8
OBIS – klankonderscheiding	.89 (24)	.92 (152)	9
OBIS - kennis van letters	.60 (24)	.96 (152)	27
OBIS - leesvaardigheid woorden	.67 (24)	.84 (152)	10
OBIS – leesvaardigheid zinnen	* (24)	.98 (152)	10
OBIS – leesvaardigheid teksten	* (24)	* (152)	18
OBIS – rekenjargon	.89 (24)	.78 (152)	7
OBIS - geheugen voor aantallen	.59 (24)	.82 (152)	4
OBIS - sommen A	.79 (24)	.84 (152)	8
OBIS – kennis van getallen	.91 (24)	.92 (152)	21
OBIS - kennis namen van vormen	.79 (24)	.74 (152)	5
OBIS - hoofdrekenen	.45 (24)	.62 (152)	8
OBIS - sommen B	.44 (24)	.63 (152)	16

* schaal heeft geen variantie, alle items in de schaal zijn door alle leerlingen in de substeek-proef fout gemaakt

Nu betreft de homogeniteit de interne consistentie van de toetsen. De items behorend bij één (sub)toets meten voor veel van de (sub)toetsen bij leerlingen van vier tot acht jaar in ruime mate dezelfde trek ofwel hebben genoeg gemeenschappelijke variantie. Voor jongere leerlingen blijken de toetsen dus veel minder betrouwbaar of homogeen.
Of de scores ook stabiel blijven over de tijd, is een andere vraag. Het zou bijvoorbeeld kunnen dat leerlingen heel wisselend presteren op verschillende tijdstippen. Om na te gaan of de scores stabiel blijven over tijd, is de test-hertest betrouwbaarheid bepaald. Zoals beschreven in het bovenstaande zijn de toetsen bij een deel van de respondenten na één of twee weken nogmaals afgenomen. De test-hertestbetrouwbaarheid is per (sub)toets berekend als de correlatie tussen de toetsscores op de eerste en de tweede afname (zie tabel 14).
Uit de tabel blijkt dat de test-hertest betrouwbaarheid voor enkele (sub)toetsen niet erg hoog is. De waarden variëren van .54 (TAK passieve woordenschat) tot .99 (OBIS-leesvaardigheid zinnen). Nemen we als minimale test-hertestbetrouwbaarheid .80 (cf. COTAN), dan zien we acht van de 21 (sub) toetsen met lagere test-hertest betrouwbaarheden. Het betreft de Peabody, de TAK-subtoetsen klankarticulatie, klankonderscheiding en passieve woordenschat en de OBIS-subtoetsen leesbegrip, rekenjargon, geheugen en kennis van vormen. Voor de Peabody en de drie TAK-subtoetsen wijzen deze waarden op een te geringe betrouwbaarheid. Voor de OBIS past een nuancering bij deze resultaten. De subtoets leesbegrip heeft tien items, rekenjargon heeft zeven items, geheugen maar vier items en vormen slechts vijf. De lage aantallen items kunnen zowel de alfa's als de test-hertest negatief beïnvloeden. Daarnaast blijkt uit de resultaten van de factoranalyses die in het onderstaande gepresenteerd worden (zie paragraaf 4.6), dat deze subtoetsen met een lage test-hertest heel goed gecombineerd kunnen worden met andere, wel betrouwbare delen van de OBIS, omdat ze eigenlijk hetzelfde meten. Zo blijken de subtoetsen rekenjargon en kennis van vormen hetzelfde te meten als passieve woordenschat, geheugen meet rekenvaardigheid en leesbegrip valt samen met de toetsen getallenkennis, letterkennis, leesvaardigheid woorden en schrijven in een factor die we de 'technisch lezen factor' noemen. De onbetrouwbare OBIS-subtoetsen zouden dus zonder verlies aan informatie weggelaten kunnen worden.

Om na te gaan of er een leereffect optreedt bij herhaald afnemen van de toetsen, zijn t-toetsen uitgevoerd op de somscores van de test- en de hertestdata. Uit tabel 14 en tabel A1 in de Appendix blijkt dat de leerlingen op de hertest in alle gevallen hoger scoren dan bij de eerste afname. Van de 21 (sub)toetsen blijken van 12 de test- en de hertestgemiddelden significant te verschillen op 5%-niveau. In alle gevallen is de gemiddelde score op de hertest hoger dan de gemiddelde score op de eerste afname van de toets. Dit betekent dat er inderdaad een leereffect optreedt voor deze subtoetsen. Indien men deze toetsen dus zou willen gebruiken om bijvoorbeeld achterstandsleerlingen te selecteren, dan zouden leerlingen die de toetsen meermaals gemaakt hebben meer kans hebben onterecht niet als achterstandsleerling aangemerkt te worden.

In de laatste kolom van tabel 14 staan de kappa's die aangeven in welke mate de toetsen overeenstemmen op respectievelijk de test en de hertest bij het oormerken van de 25% laagst scorende leerlingen. Bij toetsen die erg scheef verdeeld zijn omdat ze te moeilijk zijn, is deze analyse niet verricht. De kappa's kunnen opgevat worden als de betrouwbaarheid van de toetsen in het oormerken van de 25% slechtst presterende leerlingen. De kruistabellen waarop deze kappa's zijn gebaseerd, staan in de Appendix. Dit resultaat nuanceert de resultaten van de hiervoor gepresenteerde betrouwbaarheidsanalyses. Hoewel de toetsen over de gehele vaardigheidsrange veelal voldoende betrouwbaar zijn, geldt dit niet voor het oormerken van de 25% slechtst presterende leerlingen. Alleen de TvK en de OBIS-toetsen passieve woordenschat, klankarticulatie, kennis van letters en kennis van getallen zijn op individueel niveau qua stabiliteit (test-hertest) voldoende betrouwbaar (>.80) in het aanwijzen van de 25% minst taalvaardige leerlingen.

Ook voor de test-hertestbetrouwbaarheid is nagegaan wat de betrouwbaarheid is als we alleen de jongste leerlingen uit de steekproef in de analyses betrekken. De resultaten van deze analyses staan in tabel 15. Omdat het aantal leerlingen van vier tot vijf jaar in de test-hertestconditie te klein was, zijn deze analyses alleen verricht voor de leerlingen van vier tot zes jaar.

*Tabel 14 Gegeven worden de gemiddelde score (M), de standaarddeviatie (s.d.), de steekproefgrootte (N), de test-hertestbetrouwbaarheid (t-th) met bijbehorende steekproefgrootte (N). Een * onder 'sig t/ht' geeft aan dat het verschil in gemiddelde tussen test en hertest significant is en onder K staat de kappa van de kruistabel van beide dichotome scores (test en hertest) die aangeven of leerlingen tot het onderste kwartiel behoren op respectievelijk test en hertest*

Subtest	M	s.d.	N	t-ht (N)	sig t/ht	K
TvK, 1e afname (zonder item 15)	24.33	5.07	341	.90**		.81
TvK, hertest (zonder item 15)	25.20	5.22	95	(92)		
Peabody, 1e afname	71.05	15.65	349	.74**	*	.63
Peabody, hertest	76.85	14.32	96	(96)		
TAK - klankonderscheiding, 1e afname	41.76	8.52	352	.74**	*	.58
TAK - klankonderscheiding, hertest	42.14	9.73	94	(94)		
TAK - klankarticulatie, 1e afname	42.14	3.95	352	.65**	*	.65
TAK - klankarticulatie, hertest	42.06	4.15	94	(94)		
TAK – passieve woordenschat, 1e afname	52.66	18.93	352	.54**	*	.59
TAK - passieve woordenschat, hertest	56.64	24.24	94	(94)		
OBIS - aanvankelijk schrijven, 1e afname	2.83	1.76	344	.83**		.75
OBIS - aanvankelijk schrijven, hertest	3.07	1.71	82	(82)		
OBIS - passieve woordenschat, 1e afname	24.46	6.99	344	.91**	*	.80
OBIS - passieve woordenschat, hertest	26.28	6.76	82	(82)		
OBIS - leesbegrip, 1e afname	5.35	2.23	344	.70**	*	.35
OBIS - leesbegrip, hertest	6.77	2.53	82	(82)		
OBIS - klankarticulatie, 1e afname	6.49	2.06	344	.82**	*	.87
OBIS - klankarticulatie, hertest	6.96	1.99	82	(82)		
OBIS - klankonderscheiding, 1e afname	6.42	3.12	344	.91**	*	.76
OBIS - klankonderscheiding, hertest	7.44	2.86	82	(82)		
OBIS - kennis van letters, 1e afname	6.61	7.94	344	.96**		.83
OBIS - kennis van letters, hertest	7.84	8.39	82	(82)		
OBIS - leesvaardigheid woorden, 1e afname	2.33	3.25	344	.85**		-
OBIS - leesvaardigheid woorden, hertest	3.21	3.68	82	(82)		
OBIS - leesvaardigheid zinnen, 1e afname	2.78	11.71	344	.99	*	-
OBIS - leesvaardigheid zinnen, hertest	5.23	16.89	82	(82)		
OBIS – leesvaardigheid teksten, 1e afname	.13	1.27	344	.96**		-
OBIS - leesvaardigheid teksten, hertest	.28	1.80	82	(82)		

Vervolg tabel 14

Subtest	M	s.d.	N	t-ht (N)	sig t/ht	K
OBIS – rekenjargon, 1e afname	6.58	.98	344	.77**	*	.55
OBIS - rekenjargon, hertest	6.74	.86	82	(82)		
OBIS - geheugen, 1e afname	3.31	1.21	344	.77**		.69
OBIS - geheugen, hertest	3.34	1.28	82	(82)		
OBIS - sommen A, 1e afname	5.55	2.43	344	.87**	*	.72
OBIS - sommen A, hertest	6.32	2.56	82	(82)		
OBIS – kennis van getallen, 1e afname	9.07	5.03	344	.95**		.85
OBIS – kennis van getallen, hertest	9.27	5.11	82	(82)		
OBIS – kennis van vormen, 1e afname	4.21	1.23	344	.65**		.56
OBIS – kennis van vormen, hertest	4.39	1.19	82	(82)		
OBIS - hoofdrekenen, 1e afname	1.76	1.73	344	.83**		-
OBIS – hoofdrekenen, hertest	2.41	2.02	82	(82)		
OBIS - sommen B, 1e afname	2.11	1.78	344	.81**	*	.73
OBIS - sommen B, hertest	2.79	2.28	82	(82)		

Uit de tabel blijkt dat voor enkele OBIS-(sub)toetsen de test-hertestbetrouwbaarheid hoger uitvalt voor de jongere leerlingen. Stellen we weer een ondergrens van .80 als norm, dan voldoen nu de TvK en de OBIS-toetsen passieve woordenschat, klankarticulatie, klankonderscheiding, kennis van letters, leesvaardigheid woorden, leesvaardigheid zinnen, geheugen, sommen A, kennis van getallen, kennis van vormen, hoofdrekenen en sommen B. De OBIS blijkt dus qua test-hertest betrouwbaarder naarmate de leerlingen jonger zijn. Dit is opvallend daar de homogeniteit (zie hierboven) voor de jongere leerlingen over het algemeen niet beter bleek dan voor de gehele steekproef.

Tabel 15 De test-hertestbetrouwbaarheid van de verschillende schalen (T-ht), apart voor leerlingen van 4.4 tot 6 jaar. (tussen haken het aantal proefpersonen). n = aantal items

Variabelen	T-ht 4.4 - 6 jr	N
TvK	.85 (37)	31
Peabody	.48 (51)	108
TAK - klankarticulatie	.56 (49)	45
TAK - klankonderscheiding	.60 (49)	50
TAK - passieve woordenschat	.43 (49)	96
OBIS - aanvankelijk schrijven	.71 (33)	1
OBIS - passieve woordenschat	.95 (33)	31
OBIS - leesbegrip	.73 (33)	10
OBIS – klankarticulatie	.92 (33)	8
OBIS – klankonderscheiding	.91 (33)	9
OBIS - kennis van letters	.99 (33)	27
OBIS - leesvaardigheid woorden	.89 (33)	10
OBIS – leesvaardigheid zinnen	1.00 (33)	10
OBIS – leesvaardigheid teksten	*	18
OBIS – rekenjargon	.77 (33)	7
OBIS - geheugen voor aantallen	.85 (33)	4
OBIS - sommen A	.91 (33)	8
OBIS – kennis van getallen	.93 (33)	21
OBIS - kennis namen van vormen	.81 (33)	5
OBIS - hoofdrekenen	.80 (33)	8
OBIS - sommen B	.87 (33)	16

* correlatie niet te berekenen door te kleine N of gebrek aan variantie

Conclusie
Ter bepaling van de betrouwbaarheid van de voor het onderzoek afgenomen (sub)toetsen is een aantal verschillende statistieken berekend: de homogeniteit (alfa) over alle leerlingen, over de leerlingen van vier tot vijf en over de leerlingen van vier tot zes, de test-hertest voor alle leerlingen en voor leerlingen van vier tot zes en de kappa over de overeenstemming tussen test en hertest in het oormerken van de 25% minst taalvaardige leerlingen. De (sub)toetsen die op al deze maten voldoende scoren (>.80) zijn alleen de OBIS-toetsen passieve woordenschat, klankarticulatie, klankonderscheiding en kennis van getallen. Verder blijkt de OBIS op de test-hertest betrouwbaarder voor de leerlingen van vier tot zes jaar dan voor de gehele steekproef van vier tot acht jaar oud.

4.6 Validiteit van de (sub)toetsen

In het bovenstaande hebben we de homogeniteit en de test-hertest-betrouwbaarheid van de toetsen beschreven. In deze paragraaf beschrijven we de resultaten van exploratieve en toetsende factoranalyses. De resultaten van deze analyses geven eveneens een indruk van de betrouwbaarheid van de toetsen en een indicatie van de validiteit. De principale componentenanalyses per (sub)toets geven, net als de alfa's, inzicht in de homogeniteit van de toets, maar geven daarnaast ook inzicht in het aantal dimensies dat er per (sub)toets wordt gemeten. De exploratieve en toetsende factoranalyses op somscores per (sub)toets geven inzicht in de mate waarin verschillende (sub)toetsen gelijke dan wel verschillende vaardigheden meten en dus informatie over de validiteit van de (sub)toetsen. Daarnaast wordt middels het toetsen van verschillende meergroepenmodellen ook nagegaan of de toetsen meetinvariant ofwel onpartijdig zijn voor onderscheiden deelpopulaties in de steekproef.

Exploratieve en toetsende factoranalyses per (sub)toets op itemniveau
Met behulp van exploratieve factoranalyses is bepaald in hoeverre de (sub)toetsen unidimensioneel zijn. Alfa's geven een ondergrens van de betrouwbaarheid van een itemverzameling, maar zijn zeer gevoelig voor de aantallen items. Indien items positief samenhangen, zal een hoger aantal items vanzelf tot een hogere alfa leiden. Een hoge alfa is dus geen bewijs van unidimensionaliteit. Voor de explo-

ratieve factoranalyses is besloten om principale componentenanalyse (PCA) te gebruiken, daar we met dichotome items geen (echte) factoranalyse kunnen doen. In tabel 16 staan de resultaten van de exploratieve PCA's. Vooraf is aldoor eerst de Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) bepaald. Deze moet groter dan .8 zijn, anders zijn er te weinig proefpersonen om de analyse te doen. Daarnaast is per analyse Bartlett's test of sphericity (BTS) berekend. Deze moet significant zijn en geeft aan dat er voldoende samenhang is tussen de items voor het doen van PCA. De PCA's zijn alleen verricht op de items met spreiding. Items zonder spreiding zijn vooraf verwijderd.

Tabel 16 Resultaten PCA'S. Gegeven worden per (sub-)toets de Kaiser-Meyer-Olkin measure of sampling adequacy (KMO), de Chi-kwadraat (χ2), het aantal vrijheidsgraden (df) en de overschrijdingskans van Bartlett's test of sphericity (BTS), het aantal componenten met een eigenwaarde groter dan 1 (E.W. > 1), het zinvol aantal te onderscheiden componenten uitgaande van de scree plot (scree) en de percentages verklaarde variantie in de scores door de componenten die volgen uit de scree plot, zowel ongeroteerd als geroteerd (laatste tussen haken)

Toets/subtest	KMO	BTS (Chi^2, df en p)	E.W. > 1	scree	% v.v.
TvK	.850	χ2=1996.577 df=465 p=.000	9	1	1e comp. 19%
Peabody	*	*	28	2	1e comp. 19% (15%) 2e comp. 8% (13%)
TAK - klankonderscheiding	.906	χ2=6794.901 df=1225 p=.000	13	2	1e comp. 26% (24%) 2e comp. 5% (7%)
TAK - klankarticulatie	.774	χ2=4633.545 df=990 p=.000	15	1 a 2	1e comp. 17% (12%) 2e comp. 5% (9%)
TAK - passieve woordenschat	.933	χ2=15855 df=4560 p=.000	22	3	1e comp. 25% (16%) 2e comp. 7% (11%) 3e comp. 3% (9%)
OBIS - passieve woordenschat	.943	χ2=8019.659 df=465 p=.000	5	2 a 3	1e comp. 42% (30%) 2e comp. 11% (18%) 3e comp. 5% (11%)
OBIS - leesbegrip	.778	χ2=729.602 df=45 p=.000	2	2	1e comp. 31% (29%) 2e comp. 15% (17%)
OBIS – klankarticulatie	.889	χ2=824.121 df=28 p=.000	1	1	1e comp. 47%
OBIS - klankonderscheiding	.934	χ2=2100.488 df=36 p=.000	1	1	1e comp. 62%

Vervolg tabel 16

Toets/subtest	KMO	BTS (Chi^2, df en p)	E.W. > 1	scree	% v.v.
OBIS - kennis van letters	.955	$\chi2$=10201.020 df=351 p=.000	3	2	1e comp. 58% (42%) 2e comp. 11% (28%)
OBIS - leesvaardigheid woorden	.919	$\chi2$=2534.085 df=45 p=.000	2	1	1e comp. 62%
OBIS - leesvaardigheid zinnen	.835	$\chi2$=10247.431 df=45 p=.000	2	2	1e comp. 82% (55%) 2e comp. 14% (14%)
OBIS – leesvaardigheid teksten	*	*	3	1	1e comp. 72%
OBIS – rekenjargon	.781	$\chi2$=833.547 df=21 p=.000	2	1	1e comp. 47%
OBIS - geheugen	.540	$\chi2$=810.561 df=6 p=.000	1	1	1e comp. 69%
OBIS - sommen A	.873	$\chi2$=1120.521 df=28 p=.000	1	1	1e comp. 51%
OBIS – kennis van getallen	*	*	4	4	1e comp. 41% (23%) 2e comp. 19% (19%) 3e comp. 10% (18%) 4e comp. 5% (16%)
OBIS – kennis van vormen	.730	$\chi2$=423.805 df=10 p=.000	1	1	1e comp. 49%
OBIS - hoofdrekenen	.797	$\chi2$=577.435 df=28 p=.000	2	1	1e comp. 38%
OBIS - sommen B	.725	$\chi2$=1327.922 df=66 p=.000	3	3	1e comp. 32% (22%) 2e comp. 16% (22%) 3e comp. 11% (14%)

* matrijs niet positief definiet (determinant=0)

Omdat de communaliteiten niet steeds groter zijn dan .6 en de steekproef per PCA telkens groter is dan 200, kan het best gekeken worden naar de scree-plot om te bepalen of er één dan wel meerdere factoren of dimensies in de scores van één (sub)toets te vinden zijn (Field, 2005: p. 633). De scree-plots zijn opgenomen in de Appendix (onder tabellen A2 t/m A21).

Uit de resultaten van de PCA's blijkt dat de subtoetsen die zijn opgenomen in het onderzoek zijn op te vatten als unidimensionele toetsen. Weliswaar zijn de Chi-kwadraat-waarden significant, maar dat is met een steekproefgrootte van rond de 350 leerlingen per toets niet direct een indicatie van slechte passing. Wat wijst op unidimensionaliteit is dat de eerste component voor rotatie altijd minstens twee keer zoveel variantie verklaart als de tweede component en ook uit de scree-plots blijkt dat de knik in de curve die het verloop van de eigenwaarden per factor aangeeft voor de meeste toetsen na de eerste factor ligt en dus op een éénfactormodel wijst.

De toetsen die een scree-plot vertonen waaruit meer dan één factor zou kunnen worden afgeleid, zijn de passieve woordkennistoetsen van de Peabody, de TAK en de OBIS en de OBIS-subtoetsen leesvaardigheid zinnen, leesvaardigheid teksten, letterkennis, getallenkennis en sommen B. Voor de toetsen leesvaardigheid zinnen en leesvaardigheid teksten is het resultaat niet belangrijk, daar deze toetsen een dusdanig bodemeffect vertonen dat we ze verder niet in de analyses meenemen (zie boxplots in de Appendix). Dat de toetsen voor passieve woordenschat meerdere factoren lijken te vertonen, is niet verrassend. Woordkennistoetsen vertonen vaak een wat lagere homogeniteit, doordat kennis van woorden gerelateerd is aan de mate van bekendheid met bepaalde semantische velden (betekenisgebieden) die niet voor alle kinderen gelijk zijn. Zo zal een bakkerskind meer termen kennen gerelateerd aan de bakkerij, enz. Dat letterkennis en getallenkennis meer dan één factor lijken te bevatten, is moeilijker te duiden. Hetzelfde geldt voor de klankonderscheiding van de TAK en voor sommen B. Overigens worden bij principale componentenanalyses op dichotome data vaak meer factoren gevonden dan er in werkelijkheid zijn. Deze extra factoren betreffen zogenaamde 'moeilijkheidsfactoren' en zijn niet direct op te vatten als bewijs voor meerdimensionaliteit. Onze conclusie luidt dat de resultaten van de exploratieve factoranalyses niet wijzen op meerdimensionele (sub)toetsen.

Wellicht ten overvloede zijn ook nog toetsende factoranalyses per (sub)toets verricht om voor elke (sub)toets na te gaan of een éénfactormodel past bij de data. In deze analyses zijn items zonder variantie (door iedereen goed dan wel door iedereen fout gemaakt) verwijderd. Als fit-indices gebruiken we de Chi-kwadraat, de Comparative Fit Index (GFI), de Tucker-Lewis Index (TLI), de Root Mean Square Error of Approximation (RMSEA) en de Standardized Root Mean Square Residual (SRMR). De interpretatie van deze fit-indices is als volgt. Chi-kwadraat is de strengste fit index en toetst 'exacte fit'. Een probleem met de Chi-kwadraat is dat de power erg groot wordt naarmate de steekproef groter wordt en dat deze maat voor sociaal wetenschappelijk onderzoek wel erg streng is (MacCallum, Browne & Sugawara, 1996). Bij grote steekproeven wordt daarom ook wel het criterium gehanteerd dat de ratio van Chi-kwadraat gedeeld door het bijbehorend aantal vrijheidsgraden niet veel groter dan twee mag zijn. Omdat dit een wat los criterium is, zijn er nog een aantal andere fit-indices opgenomen in de tabel. CFI kan waarden aannemen tussen 0 en 1. TLI ligt meestal ook tussen 0 en 1 maar kan ook buiten deze range terechtkomen. Als CFI en TLI in de buurt van .95 liggen, wijst dat op een goede passing (Hu & Bentler, 1999). Bij een goede passing mag SRMR niet groter worden dan ongeveer .08 (Hu & Bentler, 1999). RMSEA (Steiger & Lind, 1980) geeft net als de andere fit-indices aan in welke mate de data overeenkomen met het gehanteerde model. MacCallum, Browne & Sugawara (1996) laten aan de hand van simulaties zien dat een RMSEA kleiner dan .05 wijst op zeer goede fit, een waarde tussen .05 en .08 wijst op redelijke fit, waarden tussen .08 en .10 wijzen op matige fit en waarden boven .10 op misfit.

De strengste fit-index uit tabel 17 is de Chi-kwadraat. We zien dat deze in een aantal gevallen niet significant is. Voor de TvK is de overschrijdingskans .014. Gezien de power van de Chi-kwadraat (N=341) wijst dit op een goede passing. De OBIS-toetsen 'beginnende geletterdheid', 'rekenjargon', 'sommen A', 'namen van vormen' en 'hoofdrekenen' is de Chi-kwadraat zelfs niet significant op 5%-niveau. Ook voor deze subtoetsen van de OBIS past een éénfactormodel dus goed. Ook de andere fit indices wijzen voor deze toetsen op een goede passing. CFI en TLI liggen rond of boven .95 en RMSEA blijft voor al deze toetsen onder .08 met uitzondering van de toets 'hoofdrekenen'. Kijken we naar de fit van de modellen voor de andere (sub)toetsen, dan zien we dat deze over het algemeen redelijk tot goed passen. CFI en TLI wijzen voor alle modellen op een goede pas-

sing. Alleen de CFI's voor 'leesbegrip' van de OBIS en voor 'klankarticulatie' van de TAK zijn aan de lage kant (respectievelijk .86 en .89). RMSEA wijst ook voor alle modellen op goede passing, behalve voor de OBIS-subtoetsen 'passieve woordenschat', 'leesbegrip', 'sommen A', 'getallenkennis' en 'sommen B'. Opvallend is dat SRMR voor enkele toetsen hoog is, terwijl de andere fit-indices op goede passing voor deze toetsen wijzen. Het betreft de TvK, Peabody, alle drie de TAK-subtoetsen en de OBIS-subtoetsen 'passieve woordenschat', 'leesbegrip', 'kennis van letters', 'geheugen', 'kennis van getallen', 'hoofdrekenen' en 'sommen B'. De SRMR wijst op grootte van de gemiddelde residuen oftewel op enkele minder betrouwbare items. Gezien de andere passingsmaten lijkt misfit echter niet problematisch.

Tabel 17 Resultaat toetsende factoranalyses per toets. Alle toetsen betreffen éénfactormodellen. Gegeven worden de steekproefgrootte (N), het aantal items (n), de Chi-kwadraat (χ^2) met aantal vrijheidsgraden (df) en overschrijdingskans (p), de Comparative Fit Index (CFI), de Tucker-Lewis Index (TLI) de Root mean square error of approximation (RMSEA), de Standardized Root Mean Square Residual (SRMR), en het aantal items dat een niet significante factorlading vertoont (n.s.) (C.R. < 2.58)

Analyse	N	n	χ^2, df, p	CFI	TLI	RMSEA	SRMR	n.s.
TvK	341	31	139.618; 105; .014	.961	.970	.031	.10	2
Pea	349	81	297.408; 180; .000	.959	.977	.043	.10	-
TAK – klank-onderscheiding	352	50	126.800; 72; .000	.962	.975	.047	.13	7
TAK – klank-articulatie	352	45	70.919; 40; .002	.893	.917	.047	.16	3
TAK - passieve woordenschat	352	96	384.305; 141; .000	.930	.961	.070	.13	-
OBIS - passieve woordenschat	344	31	222.292; 47; .000	.979	.992	.104	.11	-
OBIS – leesbegrip	344	10	154.698; 22; .000	.863	.913	.132	.14	-
OBIS – klank-articulatie	344	8	15.853; 12; .198	.994	.997	.031	.04	-
OBIS – klank-onderscheiding	344	9	35.671; 18; .008	.997	.999	.053	.04	-
OBIS - kennis van letters	344	26	189.132; 43; .000	.996	.997	.099	.07	-

Vervolg tabel 17

Analyse	N	n	χ^2, df, p	CFI	TLI	RMSEA	SRMR	n.s.
OBIS - leesvaardigheid woorden	344	10	33.258; 14; .003	.998	.998	.063	.04	-
OBIS – rekenjargon	344	7	10.639; 7; .155	.995	.995	.039	.08	-
OBIS - geheugen	344	4	30.109; 2; .000	.997	.997	.202	.09	-
OBIS - sommen A	344	8	17.968; 14; .208	.998	.999	.029	.04	-
OBIS – kennis van getallen	344	21	346.747; 14; .000	.959	.974	.263	.26	-
OBIS – kennis van vormen	344	4	6.790; 2; .033	.960	.940	.083	.05	-
OBIS – hoofdrekenen	344	8	20.571; 12; .057	.989	.989	.046	.11	-
OBIS - sommen B	344	12	42.937; 9; .000	.929	.937	.105	.18	1

Op grond van de resultaten van de PCA's en de toetsende factoranalyses is besloten om alle subtoetsen ongesplitst te sommeren. We hebben immers aannemelijk gemaakt dat alle subtoetsen voldoende unidimensioneel zijn om per subtoets de sommering tot één score te rechtvaardigen.

Exploratieve en toetsende factoranalyses op de somscores van de (sub)toetsen

De volgende vraag die we moeten beantwoorden is welke onderliggende dimensies of vaardigheden er gemeten worden door de verzameling van verschillende subtoetsen. Om dit te bepalen zijn exploratieve en toetsende factoranalyses verricht op de somscores van de subtoetsen. In het onderstaande behandelen we eerst de uitkomsten van de exploratieve factoranalyses. Deze uitkomsten bepalen deels de keuze voor het model waarvan de passing met behulp van toetsende factoranalyses zal worden getoetst.

De exploratieve factoranalyse op somscores zijn verricht met behulp van het programma Mplus (Muthèn & Muthèn, 1999) onder de aanname dat alle missende data aselect over de leerlingen zijn verspreid. Daar we leerlingen aselect hebben toegewezen aan de toetscondities A tot en met F (zie figuur 2), is deze aanname gerechtvaardigd en kunnen de modellen worden gefit met 'full information maximum likelihood'.

Zoals eerder gezegd bleek dat de OBIS-close toets (leesvaardigheid verhaaltjes) en de OBIS-toets 'leesvaardigheid zinnen' te scheef zijn verdeeld om mee te nemen. Overblijven zijn dus de somscores voor TvK (passieve woordenschat[18]), de som van de Peabody (passieve woordenschat), drie somscores van de TAK (klankonderscheiding, klankarticulatie en passieve woordenschat) en de OBIS-toetsen (aanvankelijk schrijven, passieve woordenschat, leesbegrip, klankarticulatie, klankonderscheiding, kennis van letters, leesvaardigheid woorden, rekenjargon, geheugen, sommen A, getallen, vormen, hoofdrekenen en sommen B).
De exploratieve factoranalyses zijn verricht op oplossingen met één, twee, drie, vier en vijf factoren. Omdat we bij vergelijkbare passing de voorkeur moeten geven aan het meest spaarzame model (vgl. Ockham's razor), kijken we eerst naar de passing van een éénfactormodel. Immers, al zouden we a priori uitgaan van meerdere factoren, als een éénfactormodel goed bij de data past en een tweefactormodel niet significant beter bij de data past, zijn we gedwongen te erkennen dat alle toetsen min of meer hetzelfde onderliggende construct meten. Om deze reden presenteren we eerst een éénfactormodel en vervolgens de andere modellen met meer factoren. Telkens toetsen we of een model met een factor meer dan het voorgaande model significant beter past dan dat voorgaande model. Zodra toevoeging van een extra factor geen significante fitverbetering meer oplevert, moet gekozen worden voor het model dat nog wel een fitverbetering liet zien ten opzichte van het voorgaande model.

In tabel 18 staan de resultaten van de eerste exploratieve factoranalyse uitgaand van één factor. Dit model past redelijk. De Chi-kwadraat is significant, maar gegeven de steekproefgrootte is dat nog geen bewijs van misfit. De waarden van RMSEA en SRMR wijzen op een redelijke passing (Hu & Bentler, 1999).
Vervolgens is een exploratieve factoranalyse verricht met twee factoren. De resultaten van deze analyse staan in tabel 19. Daar het éénfactormodel en het tweefactormodel genest zijn, kunnen we ook nagaan of het tweefactormodel significant beter past dan het éénfactormodel. Dit blijkt het geval te zijn. Het verschil in Chi-kwadraat bedraagt 677.737 - 340.314 = 337.423 en het verschil in vrijheidsgraden bedraagt 152 - 134 = 18 ($p < .001$).

[18] Item 15 is niet in de somscore opgenomen vanwege een negatieve item-testcorrelatie.

Nu we een tweefactoroplossing hebben, kunnen we ook proberen de gevonden factoren inhoudelijk te duiden. In tabel 19 staan zowel de orthogonale als de oblique[19] oplossing gegeven. Kijken we naar de factorladingen, dan zien we in de oblique oplossing dat de tweede factor een taalfactor is waar de woordkennistoetsen van de TvK, de TAK en de OBIS hoog op laden. De andere twee subtoetsen van de TAK laden ook voornamelijk op deze factor, evenals leesbegrip, klankarticulatie en klankonderscheiding van de OBIS.

Tabel 18 Exploratieve factoranalyse: 1-factoroplossing (ML)

χ^2 = 677.737; df = 152; p = .000; RMSEA = .063; 90% C.I. RMSEA = .058 - .067; SRMR = .082; N = 883	factorladingen op 1e factor
TvK	.842
Peabody	.807
TAK – klankonderscheiding	.707
TAK – klankarticulatie	.557
TAK – passieve woordenschat	.829
OBIS – aanvankelijk schrijven	.663
OBIS - passieve woordenschat	.755
OBIS - leesbegrip	.672
OBIS – klankarticulatie	.602
OBIS – klankonderscheiding	.739
OBIS - kennis van letters	.650
OBIS - leesvaardigheid woorden	.617
OBIS - rekenjargon	.410
OBIS - geheugen	.630
OBIS - sommen A	.840
OBIS – kennis van getallen	.779
OBIS – kennis van vormen	.597
OBIS - hoofdrekenen	.765
OBIS - sommen B	.753

De OBIS toetsen aanvankelijk schrijven, kennis van letters en leesvaardigheid woorden laden echter op de eerste factor. Kijken we naar de rekentoetsen van de OBIS, dan zien we als eerste rekenjargon op de taalfactor laden. Dit is te verklaren als we naar de items kijken. Gevraagd wordt naar kennis van begrippen als 'grootste' en 'kleinste', 'meer' en 'minder' en 'langste' en kortste'. Deze subtoets is

[19] Een orthogonale oplossing houdt in dat de factoren onderling niet gecorreleerd zijn, in een oblique oplossing mogen factoren onderling wel samenhangen.

dus heel goed als woordkennistoets op te vatten. De geheugentoets (tellen en onthouden) laadt onverwacht ook op de taalfactor, evenals sommen A en kennis van de namen van vormen (ook woordkennis?). Alleen de getallentoets, hoofdrekenen en sommen B laden duidelijk op de eerste factor. De factorladingen van de orthogonale oplossing laten eenzelfde patroon zien (zie vetgedrukte factorladingen).

In tabel 20 staat het resultaat van de exploratieve factoranalyse met drie factoren. De drie-factoroplossing past significant beter dan de tweefactoroplossing. Het verschil in Chi-kwadraat bedraagt 340.314 – 223.661 = 116.653. Het verschil in vrijheidsgraden bedraagt 134 – 117 = 17, wat een significante fitverbetering impliceert ($p < .001$).

Tabel 19 Exploratieve factoranalyse: 2-factoroplossing (ML); factorladingen van de geroteerde oplossing, oblique (Promax) en orthogonaal (Varimax) - De duidelijk hoogste ladingen die de factorinhoud bepalen, zijn vetgedrukt

χ^2 = 340.314; df = 134; p = .000; RMSEA = .042; 90% C.I. RMSEA = .036 - .047; SRMR = .052; N = 883	ladingen op 1e factor Oblique	ladingen op 2e factor Oblique	ladingen op 1e factor Orthogonaal	ladingen op 2efactor Orthogonaal
TvK	-.090	**.951**	.254	**.859**
Peabody	-.050	**.862**	.260	**.789**
TAK – klankonderscheiding	.127	**.625**	.342	**.627**
TAK – klankarticulatie	.119	**.480**	.283	**.488**
TAK - passieve woordenschat	.059	**.803**	.342	**.770**
OBIS – aanvankelijk schrijven	.375	.348	.478	.450
OBIS - passieve woordenschat	-.026	**.831**	.272	**.768**
OBIS - leesbegrip	.221	**.507**	.389	**.548**
OBIS – klankarticulatie	.010	**.625**	.232	**.588**
OBIS – klankonderscheiding	.159	**.633**	.375	**.645**
OBIS - kennis van letters	**.964**	-.174	**.847**	.159
OBIS - leesvaardigheid woorden	**.836**	-.100	**.753**	.186
OBIS - rekenjargon	-.080	**.508**	.106	**.448**
OBIS - geheugen	.010	**.650**	.241	**.611**
OBIS - sommen A	.204	**.689**	.438	**.712**
OBIS – kennis van getallen	**.612**	.262	**.670**	.449
OBIS – kennis van vormen	.063	**.571**	.263	**.554**
OBIS - hoofdrekenen	**.596**	.256	**.653**	.438
OBIS - sommen B	**.649**	.198	**.682**	.402

Promax factorcorrelatie = .648.

Als we de factoren inhoudelijk duiden, zien we dat de eerste factor ladingen krijgt van klankarticulatie van de TAK, en van de OBIS aanvankelijk schrijven, geheugen, sommen A, kennis van getallen, hoofdrekenen en sommen B. Behalve het onderdeel aanvankelijk schrijven (1 item met waarden van 1 t/m 5) en klankarticulatie lijkt dit dus een rekenfactor. De tweede factor is een vocabulairefactor met ladingen van de TvK, de Peabody, passieve woordenschattoetsen van de TAK en de OBIS, en dan ook nog klankonderscheiding van de TAK als ook van de OBIS, en leesbegrip, rekenjargon en vormen van de OBIS. Leesbegrip, rekenjargon en vormen vragen ieder naar de betekenis van woorden, dus zijn de ladingen van deze subtoetsen op de vocabulairefactor goed te verklaren. De ladingen op deze factor van beide toetsen voor klankonderscheiding zijn echter minder makkelijk te duiden. De derde factor bevat in de oblique oplossing hoge ladingen van kennis van letters en leesvaardigheid woorden. Deze twee subtoetsen doen denken aan een technisch lezen factor. In de orthogonale factoren zien we echter ook nog hoge ladingen van kennis van getallen, hoofdrekenen en sommen B. De verklaring hiervan is ons niet duidelijk, al zou kennis van getallen qua gemeten vaardigheid wellicht vergeleken kunnen worden met kennis van letters.

In tabel 21 staan de uitkomsten van de vierfactoroplossing. De vierfactoroplossing past significant beter dan de driefactoroplossing. Het verschil in Chi-kwadraat bedraagt 223.661 – 154.398 = 69.263. Het verschil in vrijheidsgraden bedraagt 117 – 101 = 16 ($p < .001$).
Per (sub)toets zijn de hoogste factorladingen weer vetgedrukt. Bij (sub)toetsen met ongeveer even hoge factorladingen op meer dan één factor, zijn soms meer dan één factorlading vetgedrukt. Dit laatste geldt voor de receptieve woordenschattoets van de TAK en voor enkele OBIS-subtoetsen (sommen A, kennis van getallen, hoofdrekenen en sommen B).

Tabel 20 3-factoroplossing (ML); factorladingen van de geroteerde oplossing, oblique (Promax) en orthogonaal (Varimax). De duidelijk hoogste ladingen die de factorinhoud bepalen, zijn vetgedrukt

$\chi 2 = 223.661$; df = 117; p = .000; RMSEA = .032; 90% C.I. RMSEA = .026 - .038; SRMR = .046; N = 883	1e factor Oblique	2e factor Oblique	3e factor Oblique	1e factor Ortho-gonaal	2e factor Ortho-gonaal	3e factor Ortho-gonaal
TvK	-.191	**1.113**	-.027	.142	.**935**	.187
Peabody	-.050	.**887**	.017	.202	.**803**	.224
TAK – klankonderscheiding	.198	.**507**	.076	.309	.**584**	.278
TAK – klankarticulatie	.**351**	.250	.026	.349	.**408**	.223
TAK - passieve woordenschat	.136	.**694**	.073	.310	.**727**	.299
OBIS – aanvankelijk schrijven	.**452**	.050	.226	.421	.319	.399
OBIS - passieve woordenschat	.157	.**700**	-.010	.310	.**724**	.231
OBIS - leesbegrip	.096	.**451**	.219	.248	.**518**	.360
OBIS – klankarticulatie	.204	.453	-.002	.282	.**518**	.193
OBIS – klankonderscheiding	.279	.**433**	.105	.360	.**562**	.318
OBIS - kennis van letters	-.082	-.030	.**960**	.149	.164	.**865**
OBIS - leesvaardigheid woorden	-.177	.127	**.846**	.090	.237	**.762**

Vervolg tabel 20

	1e factor Oblique	2e factor Oblique	3e factor Oblique	1e factor Ortho-gonaal	2e factor Ortho-gonaal	3e factor Ortho-gonaal
OBIS - rekenjargon	.019	.**468**	-.037	.132	.**435**	.093
OBIS - geheugen	.**534**	.242	-.119	.458	.454	.155
OBIS - sommen A	.**957**	.044	-.105	.**741**	.477	.280
OBIS – kennis van getallen	.**484**	-.037	.436	.**472**	.304	.**588**
OBIS – kennis van vormen	.199	.**418**	.040	.278	.**494**	.222
OBIS - hoofdrekenen	.**473**	-.016	.406	.462	.312	.**561**
OBIS - sommen B	.**495**	-.089	.452	.470	.266	.**593**

Promax correlaties: factor 1*2 = .758; factor 1*3 = .660; factor 2*3 = .528.

Tabel 21 4-factoroplossing (ML); factorladingen van de geroteerde oplossing, oblique (Promax) en orthogonaal (Varimax). De duidelijk hoogste ladingen die de factorinhoud bepalen, zijn vetgedrukt

χ^2 = 154.398; df = 101; p = .001; RMSEA = .024; 90% C.I. RMSEA = .016 - .032; SRMR = .039; N = 883	1e factor Oblique	2e factor Oblique	3e factor Oblique	4e factor Oblique	1e factor Ortho-gonaal	2e factor Ortho-gonaal	3e factor Ortho-gonaal	4e factor Ortho-gonaal
TvK	-.108	**.839**	.408	-.021	.079	**.902**	.474	.167
Peabody	.154	**.550**	.308	.012	.259	**.668**	.403	.227
TAK – klankonderscheiding	-.050	.152	**.660**	.099	.148	.345	**.620**	.268
TAK – klankarticulatie	.177	.024	**.443**	.018	.265	.204	**.442**	.201
TAK - passieve woorden-schat	**.371**	**.438**	.246	.023	**.422**	**.602**	.381	.287
OBIS – aanvankelijk schrij-ven	.231	-.114	**.398**	.227	.325	.107	**.434**	.388
OBIS - passieve woorden-schat	.080	.259	**.622**	-.030	.238	.451	**.607**	.200
OBIS - leesbegrip	.063	.203	**.381**	.193	.199	.359	**.427**	.334
OBIS – klankarticulatie	.046	.070	**.580**	-.007	.185	.250	**.535**	.171
OBIS – klankonderscheiding	.150	.129	**.508**	.094	.281	.332	**.526**	.295
OBIS - kennis van letters	-.071	-.069	.051	**.932**	.110	.088	.211	**.860**

Vervolg tabel 21

	1e factor Oblique	2e factor Oblique	3e factor Oblique	4e factor Oblique	1e factor Ortho-gonaal	2e factor Ortho-gonaal	3e factor Ortho-gonaal	4e factor Ortho-gonaal
OBIS - leesvaardigheid woorden	-.089	.133	-.014	.**813**	.076	.240	.158	.**752**
OBIS - rekenjargon	-.102	.068	.**535**	-.036	.042	.190	.**455**	.076
OBIS - geheugen	.303	-.052	.**552**	-.116	.369	.173	.**526**	.138
OBIS - sommen A	.**687**	-.035	.**425**	-.101	.**672**	.259	.**506**	.264
OBIS – kennis van getallen	.263	-.205	.**388**	.**437**	.378	.061	.**460**	.**582**
OBIS – kennis van vormen	-.014	.000	.**639**	.045	.151	.192	.**573**	.203
OBIS - hoofdrekenen	.**484**	.123	-.011	.**377**	.**491**	.311	.199	.**552**
OBIS - sommen B	.**487**	.031	-.003	.**424**	.**495**	.233	.201	.**586**

Promax correlaties: factor 1*2 = .431; factor 1*3 = .586; factor 1*4 = .597; factor 2*3 = .493; factor 2*4 = .353; factor 3*4 = .522.

De eerste factor interpreteren we als een rekenfactor. Deze factor heeft hoge ladingen van OBIS-sommen A, OBIS-hoofdrekenen en OBIS-sommen B. De tweede factor blijkt een woordkennisfactor waar de woordkennistoetsen van de TvK, de Peabody en de TAK hoog op laden. De OBIS-woordkennistoets laadt echter niet hoog op deze factor. In de driefactoroplossing deed deze OBIS-subtoets dat echter wel. Het nu niet erg hoog op de woordkennisfactor laden, zou veroorzaakt kunnen worden doordat de OBIS een andere wijze van afnemen kent dan de overige toetsen. De OBIS wordt afgenomen met een p.c. en vraagt, in tegenstelling tot de andere woordkennistoetsen, bij uitgesproken woorden het juiste object op een plaatje aan te wijzen. De andere woordkennistoetsen vragen het juiste plaatje aan te wijzen (meerkeuze met raadkans) bij een gegeven uitgesproken woord. Het kan zijn dat methodespecifieke variantie[20] de woordkennistoets van de OBIS op de derde factor laat laden. Op deze factor laden immers veel OBIS-subtoetsen hoog en al deze subtoetsen delen dezelfde meetmethode. Uiteraard is deze verklaring slechts een ad-hoc verklaring.
De derde factor krijgt hoge ladingen van de subtoetsen klankonderscheiding en klankarticulatie van zowel de TAK als de OBIS. Daarnaast heeft deze factor veel gemeen met de OBIS-subtoetsen schrijven, woordenschat, leesbegrip, rekenjargon (wat als woordkennis is op te vatten), geheugen en tenslotte de subtoets kennis van vormen (weer op te vatten als woordkennis).

Na het vierfactormodel is een vijffactormodel gefit. Dit model geeft voor de TvK echter merkwaardige resultaten (ladingen groter dan 1 en een negatief residu). Het verwijderen van TvK uit de analyse biedt ook geen soelaas daar het vijffactormodel dan niet meer convergeert.

Op grond van de exploratieve factoranalyses en op grond van inhoudelijke overwegingen is een keuze gemaakt voor het ons inziens ideale theoretische model. Dit model bevat een vocabulairefactor, een fonologische factor, een technisch-lezenfactor en een rekenfactor.
De vocabulairefactor wordt gemeten via de toetsen TvK, Peabody, het onderdeel passieve woordenschat van de TAK alsook van de OBIS en twee subtoetsen uit

[20] Methodespecifieke variantie is variantie veroorzaakt door de wijze van toetsen. Deze hoeft niet samen te gaan met het construct dat men beoogt te meten.

het rekendeel van de OBIS, namelijk rekenjargon en kennis van vormen. De reden om deze laatste twee subtoetsen in te delen bij de vocabulairefactor is dat beide toetsen vragen naar de betekenis van woorden. De toets rekenjargon vraagt naar de betekenis van termen als 'groter' en 'kleiner', 'langer' en 'korter', enz. De toets kennis van vormen vraagt naar de namen van vormen als de rechthoek, driehoek, enz. In de factoroplossingen van de exploratieve factoranalyses is te zien dat deze toetsen op één factor (zie tweede factor van de driefactoroplossing) terecht kunnen komen, al blijkt tevens dat beide toetsen op de fonologische factor zouden kunnen laden (zie vierfactoroplossing exploratieve factoranalyse, factor 3).
De fonologische factor in het door ons gekozen model wordt gemeten door de subtoetsen klankonderscheiding en klankarticulatie van zowel de TAK als de OBIS.
De technisch-lezenfactor wordt bepaald door de toetsen kennis van letters en leesvaardigheid woorden van de OBIS. Uit de exploratieve factoranalyses blijkt dat de toets kennis van getallen van de OBIS ook bij deze factor gerekend kan worden. Inhoudelijk is dit te beargumenteren daar de toets kennis van getallen vraagt naar de namen van getallen, ongeveer op vergelijkbare wijze als het vragen van namen van letters in de toets kennis van letters. Daarnaast voegen we ook de éénitemtoets aanvankelijk schrijven van de OBIS aan de technisch lezen-factor toe. Deze toets meet in welke mate het kind beseft wat schrijven is. Inhoudelijk is deze toets verschillend van de andere technisch lezen toetsen, maar er lijkt wel inhoudelijke verwantschap met kennis van letters en leesvaardigheid woorden. Ook de toets leesbegrip van de OBIS rekenen we tot de technisch lezen-factor. Deze toets meet of het kind bij een gegeven plaatje kan aangeven of iemand schrijft dan wel leest, wat het begin of eind van een zin is, en wat verschillende leestekens zijn.
De rekenfactor tenslotte, bevat de OBIS-toetsen geheugen, sommen A, hoofdrekenen en sommen B.

De toetsende factoranalyses zijn verricht ervan uitgaande dat de missende data 'missing completely at random' zijn, wat gegeven het afnamedesign is gerechtvaardigd. Zo is het mogelijk om alle aan het onderzoek deelnemende leerlingen in de analyses te betrekken, ook al hebben deze altijd maar twee van de vier toetsen gemaakt. Hoewel we op grond van theoretische noties en exploratieve factorana-

lyses menen dat een vierfactormodel ideaal is, toetsen we ook modellen die spaarzamer zijn (één-, twee- en driefactormodellen). We willen immers nagaan of de door ons gedachte onderscheiding in vier factoren houdbaar is. We zullen dus moeten laten zien dat ons vierfactormodel significant beter past dan de spaarzamer modellen die denkbaar zijn. Het meest spaarzame model is uiteraard een éénfactormodel. Dit model past redelijk bij de data (N=883; χ^2 =677.737; df=152; p=.000; CFI=.846; TLI=.827; RMSEA=.063; 95% c.i. RMSEA= .058-.067; SRMR=.075). Alle subtests laden significant (kritieke ratio's ofwel C.R. 's > 2.58; p<.001) op de factor.

Vervolgens is een tweefactormodel getoetst met een taalfactor en een rekenfactor. De keuze welke subtests op een factor laden is mede gebaseerd op de uitkomsten van de exploratieve factoranalyses, maar wel altijd inhoudelijk te verdedigen. Op de taalfactor laden de TvK, de Peabody, de drie subtoetsen van de TAK, en van de OBIS de toetsen passieve woordkennis, leesbegrip, kennis namen van vormen, klankarticulatie en klankonderscheiding, kennis van letters, leesvaardigheid woorden, aanvankelijk schrijven, rekenjargon en kennis van getallen. De rekenfactor bevat de geheugentoets, sommen A en sommen B en hoofdrekenen. Dit model past ook redelijk (N=883; χ^2 =654.612; df=151; p=.000; CFI=.852; TLI=.833; RMSEA=.061; 95% c.i. RMSEA= .057-.066; SRMR=.074). Alle subtests laden significant op de betreffende factor (C.R.'s > 2.58; p<.001). De beide factoren vertonen een correlatie van .94. Daar de modellen genest zijn, kunnen we toetsen of het tweefactormodel significant beter fit dan het éénfactormodel. Dat blijkt het geval te zijn ($\Delta\chi^2$ =23.125; Δdf=1; p=.000).

Hierna is gekeken naar de fit van een driefactormodel. In dit model onderscheiden we een vocabulairefactor, een factor voor de overige taaltoetsen en een rekenfactor. De vocabulairefactor in dit model bestaat uit de woordkennistoetsen TvK en Peabody en de toetsen passieve woordkennis van zowel de OBIS als de TAK. Daarnaast zijn ook de toetsen rekenjargon en kennis van namen van vormen, beide van de OBIS, aan deze factor toegevoegd. De taalrestfactor bestaat uit de fonologische toetsen (de klankonderscheiding en de klankarticulatietoetsen van zowel de OBIS als de TAK) en uit toetsen die te maken lijken te hebben met de vaardigheid technisch lezen (de OBIS-toetsen kennis van letters, leesvaardigheid woorden, aanvankelijk schrijven, leesbegrip en kennis van getallen). De derde factor is een rekenfactor en bevat de OBIS-toetsen sommen A, sommen B, hoofdrekenen en geheugen. Ook dit model past redelijk (N=883; χ^2 =576.744; df=149;

p=.000; CFI=.875; TLI=.856; RMSEA=.057; 95% c.i. RMSEA= .052-.062; SRMR=.071). Weer vertonen alle subtests significante ladingen op de factor die ze moeten meten (C.R.'s > 2.58; p<.001). De correlaties tussen de drie factoren in dit model zijn weer hoog. De vocabulairefactor vertoont een correlatie van .86 met de taalrestfactor en een correlatie van .84 met de rekenfactor. De taalrestfactor en de rekenfactor vertonen een correlatie van .95. Ook kunnen we weer nagaan of het driefactormodel significant beter past dan het tweefactormodel. Dit blijkt het geval te zijn ($\Delta\chi^2$ =77.858; Δdf=2; p=.000).

Het laatste getoetste model is het eerder beschreven vierfactormodel. Dit model past het best bij onze theoretische noties. In dit model onderscheiden we een vocabulairefactor, een fonologische factor, een technisch-lezenfactor en een rekenfactor. De vocabulairefactor bevat weer de woordkennistoetsen TvK en Peabody, de toetsen passieve woordkennis van zowel de OBIS als de TAK en de OBIS-toetsen rekenjargon en kennis van vormen. De fonologische factor bestaat uit de klankonderscheiding en de klankarticulatietoetsen van zowel de OBIS als de TAK. De technisch-lezenfactor bevat de OBIS-toetsen kennis van letters, leesvaardigheid woorden, schrijven, leesbegrip en kennis van getallen. De rekenfactor bevat weer de OBIS-toetsen sommen A, sommen B, hoofdrekenen en geheugen. Dit model past redelijk tot goed (N=883; χ^2 =485.825; df=146; p=.000; CFI=.900; TLI=.883; RMSEA=.051; 95% c.i. RMSEA= .046-.056; SRMR=.067) en significant beter dan het driefactormodel ($\Delta\chi^2$ =90.919; Δdf=3; p=.000). Alle factorladingen zijn in dit model weer significant (C.R.'s > 2.58; p<.001). De correlaties tussen de vier factoren zijn weer hoog. De vocabulairefactor correleert .93 met de fonologische factor, .72 met de technisch lezen-factor en .82 met de rekenfactor. De fonologische factor correleert .81 met de technisch lezen-factor en .87 met de rekenfactor. De technisch lezen- en de rekenfactor tenslotte correleren .91. De resultaten van de toetsende factoranalyse met vier factoren staan in tabel 22.

Nu we hebben laten zien dat de (sub)toetsen die we hebben afgenomen voor het onderzoek opgevat kunnen worden als unidimensioneel en we hebben laten zien dat verschillende subtoetsen op één factor passen, kunnen we stellen dat de subtoetsen die op dezelfde factor laden onderling gekalibreerd kunnen worden, bijvoorbeeld met behulp van item-responsmodellen (Hambleton & Swaminathan, 1985).

Het in tabel 22 gepresenteerde vierfactormodel past weliswaar redelijk, vooral als we kijken naar RMSEA en SRMR, maar CFI en TLI zijn aan de lage kant en ook de ratio van Chi-kwadraat en vrijheidsgraden is wat groot (486/146=3.3). Zoals we eerder opmerkten moeten CFI en TLI in de buurt van .95 liggen om van een echt goede passing te kunnen spreken (Hu & Bentler, 1999). Voor het verifiëren van meetinvariantie willen we graag een zeer goed passend model hebben, zodat onze resultaten aangaande meetinvariantie niet verstoord kunnen worden door een teveel aan ruis.

Kijkend naar de modificatie-indices van het in tabel 22 gepresenteerde model, is besloten een nieuw model te passen met drie factoren. De rekenfactor is verwijderd, mede omdat de onderzoeksvragen voor het onderhavige onderzoek zich richten op het meten van taalvaardigheid. Verder zijn enkele onbetrouwbare kleine subtoetsen verwijderd. Het resultaat van deze nieuwe modeltoetsing staat in tabel 23. Het verschil tussen de modellen in tabel 22 en in tabel 23 betreft dus uitsluitend het verwijderen van subtests. Omdat alle rekensubtests zijn verwijderd, is de rekenfactor niet in dit laatste model opgenomen. De meetinvariantie wordt alleen bepaald aan de hand van dit driefactormodel, omdat dit model aanzienlijk beter past dan het vierfactormodel.

Tabel 22 Resultaat toetsende factoranalyse 4-factormodel. Gestandaardiseeerde en ongestandaardiseerde factorladingen (kritieke ratio's tussen haken)

χ^2 = 485.825; df = 146; p = .000; CFI=.900; TLI=.883; RMSEA = .051; 90% C.I. RMSEA = .046 - .056; SRMR=.067; N = 883	vocab. factor ongest.	vocab. factor gest.	fonol. factor ongest.	fonol. factor gest.	techn. lez. factor ongest.	techn. lez. factor gest.	reken factor ongest.	reken factor gest.
TvK	1.000 (.00)	.915						
Peabody	2.836 (16.51)	.853						
TAK - klankonderscheiding			1.000 (.00)	.745				
TAK - klankarticulatie			.360 (9.71)	.580				
TAK - passieve woordenschat	3.425 (15.61)	.858						
OBIS – aanv. schrijven					.201 (12.33)	.683		
OBIS - passieve woordenschat	1.197 (15.89)	.830						

	vocab. factor ongest.	vocab. factor gest.	fonol. factor ongest.	fonol. factor gest.	techn. lez. factor ongest.	techn. lez. factor gest.	reken factor ongest.	reken factor gest.
OBIS – leesbegrip					.233 (11.20)	.624		
OBIS – klankarticulatie			.205 (10.50)	.639				
OBIS - klankonderscheiding			.375 (12.27)	.773				
OBIS - kennis van letters					1.000 (.00)	.753		
OBIS - leesvaardigheid woorden					.379 (13.37)	.697		
OBIS - rekenjargon	.095 (8.52)	.465						
OBIS – geheugen							1.000 (.00)	.614
	vocab.	vocab.	fonol.	fonol.	techn.	techn.	reken	reken

	vocab. factor ongest.	vocab. factor gest.	fonol. factor ongest.	fonol. factor gest.	techn. lez. factor ongest.	techn. lez. factor gest.	reken factor ongest.	reken factor gest.
OBIS - sommen A							2.758 (12.54)	.847
OBIS – kennis van getallen					.704 (15.58)	.838		
OBIS – kennis van vormen	.155 (11.18)	.606						
OBIS - hoofdrekenen							1.848 (11.57)	.794
OBIS - sommen B							1.883 (11.47)	.786

correlaties: vocab. * fonol. = .93; vocab. * techn. lez. = .72; vocab. * rekenen = .82; fonol. * techn. lez. = .81; fonol. * rekenen = .87; techn. lez. * rekenen = .91.

In tegenstelling tot de modeltoetsing in tabel 22, waarbij de eerste variabele per factor een lading van 1.00 kreeg ter identificatie van het model, is in tabel 23 gekozen om de factorvarianties op 1.00 te stellen ter identificatie van het model, zodat aan de gestandaardiseerde factorladingen gezien kan worden welke subtoets de latente trek (of factor) het best meet. Duidelijk is dat de TvK het meeste gemeenschappelijke variantie vertoont met de andere woordkennistoetsen. Op de fonologische factor doet de toets klankonderscheiding van de OBIS het het best. De technisch lezen factor blijkt de hoogste lading te krijgen van de OBIS-toets kennis van letters.

Tabel 23 Resultaat toetsende factoranalyse 3-factormodel. Gestandaardiseeerde en ongestandaardiseerde factorladingen (kritieke ratio's tussen haken)

χ^2 = 104.321; df = 41; p = .000; CFI=.961; TLI=.948; RMSEA = .042; 90% C.I. RMSEA = .032 - .052; SRMR=.058; N = 883	vocab. factor ongest.	vocab. factor gest.	fonol. factor ongest.	fonol. factor gest.	techn. lez. factor ongest.	techn. lez. factor gest.
TvK	4.807 (21.684)	.928				
Peabody	13.845 (19.643)	.873				
TAK - klankonderscheiding			6.370 (14.972)	.745		
TAK - klankarticulatie			2.311 (10.962)	.585		
TAK – passieve woordenschat	16.424 (19.717)	.863				

Vervolg tabel 23

	vocab. factor ongest.	vocab. factor gest.	fonol. factor ongest.	fonol. factor gest.	techn. lez. factor ongest.	techn. lez. factor gest.
OBIS - passieve woordenschat	5.575 (18.469)	.809				
OBIS – klankarticulatie			1.306 (12.707)	.639		
OBIS - klankonderscheiding			2.388 (16.644)	.774		
OBIS - kennis van letters					6.726 (18.396)	.851
OBIS - leesvaardigheid woorden					2.567 (16.828)	.795
OBIS - kennis van getallen					3.805 (15.570)	.761

correlaties: vocab. * fonol. = .915; vocab. * techn. lez. = .586; fonol. * techn. lez. = .688

Meetinvariantie (onpartijdigheid) ten aanzien van specifieke groepen leerlingen uitgaande van het driefactormodel

Via de hierboven beschreven analyses hebben we inzicht gekregen in de mate waarin de verschillende taaltoetsen dezelfde dan wel verschillende vaardigheden meten. Deze analyses zijn verricht op de totale steekproef. De volgende vraag die we willen beantwoorden is of de taaltoetsen partijdig zijn voor specifieke groepen leerlingen. Eigenlijk gaat het hier om de vraag of de taaltoetsen in onderscheiden groepen respondenten gelijke dan wel verschillende vaardigheden meten en of er, als er gelijke vaardigheiden gemeten worden, vergelijkbare scores ofwel scores op eenzelfde schaal gegenereerd worden voor de onderscheiden groepen leerlingen. Om dit te kunnen bepalen zijn meergroepenmodellen getoetst waarbij is nagegaan of er sprake is van meetinvariantie (Meredith, 1993). De mate van meetinvariantie wordt geverifieerd in stappen.

De strengste vorm van meetinvariantie, ofwel 'strict factorial invariance', impliceert dat de taaltoetsen in de verschillende groepen leerlingen dezelfde vaardigheden meten op dezelfde schaal en met een gelijke betrouwbaarheid. Deze

strengste vorm van meetinvariantie geldt indien het meest restrictieve model, met over groepen voor alle taaltoetsen gelijke residuele varianties, gelijke intercepten en gelijke factorladingen, niet significant slechter past dan het meergroepenmodel waarin alleen de residuele varianties over groepen verschillend geschat worden. Deze vorm van meetinvariantie hebben we niet getoetst in het onderhavige onderzoek, omdat het meten met ongelijke betrouwbaarheden in de verschillende subgroepen wel vervelend is, maar niet een zinvol gebruik van de taaltoetsen in de onderscheiden groepen onmogelijk maakt. We meten dan alleen de leerlingen in de ene groep nauwkeuriger dan de leerlingen in de andere groep.
De op één na strengste vorm van meetinvariantie, die we wel toetsen, is 'strong factorial invariance'. Deze vorm van meetinvariantie impliceert metingen met ongelijke betrouwbaarheden ofwel ongelijke residuele varianties over groepen, maar wel met voor alle taaltests identieke intercepten en factorladingen. Als deze vorm van meetinvariantie houdt, meten de toetsen in beide onderscheiden groepen leerlingen hetzelfde construct op dezelfde schaal. Er is echter geen sprake van 'strong factorial invariance' als een meergroepenmodel met ongelijke intercepten (en ongelijke residuen) over groepen significant beter past dan het meergroepenmodel met gelijke intercepten (en ongelijke residuen). Als dit laatste het geval blijkt, meten de taaltoetsen in de onderscheiden groepen op een verschillende schaal. Een leerling met een score van 10 in de ene groep is dan niet even taalvaardig als een leerling met een score van 10 in de andere groep. We kunnen dan van 'weak factorial invariance' spreken. In dat geval zouden de groepen waarvoor dit geldt apart genormeerd moeten worden. De scores kunnen dan niet zonder meer over groepen heen vergeleken worden.
Blijkt nu vervolgens dat een model met ongelijke residuen, ongelijke intercepten en ongelijke factorladingen per taaltest significant beter past dan het meergroepenmodel met alleen ongelijke residuen en ongelijke intercepten, dan is er evenmin sprake van weak factorial invariance en moeten we concluderen dat de taaltoetsen in de onderscheiden groepen zelfs verschillende constructen meten (Meredith, 1993). Duidelijk is dat in dit laatste geval de taaltoetsen helemaal niet gebruikt kunnen worden over groepen heen, omdat we dan met dezelfde taaltoetsen toch verschillende constructen meten in de onderscheiden groepen. Als dit het geval is, dan is het gebruik van de toetsen in slechts één van de onderscheiden groepen niet echt een oplossing van het probleem. Immers, als we ongelijke vaar-

digheden meten over groepen heen, hoe kunnen we dan weten of en zo ja in welke groep we de bedoelde vaardigheid meten?

De voor detectie van meetinvariantie gehanteerde subgroepen zijn gebaseerd op de variabelen sekse (jongen versus meisje), thuistaal (Nederlands, dialect of Fries enerzijds versus een vreemde taal anderzijds), leeftijd (van vier tot zes en van zes tot acht) en leerlinggewicht (één keer .00 versus de rest en één keer .90 versus de rest). Alle meergroepenmodellen voor het bepalen van de mate van meetinvariantie, ofwel onpartijdigheid, zijn dus getoetst voor steeds twee groepen leerlingen. Voor het passen van modellen met meer dan twee groepen bleek de steekproef te klein. Het getoetste model is steeds het in tabel 23 gerapporteerde model, behalve bij de analyses op basis van de variabele sekse. Om technische redenen die bij het beschrijven van de analyse worden vermeld, is bij de verificatie van de meetinvariantie voor sekse naast het driefactormodel uit tabel 23 ook een tweefactormodel gebruikt.
Telkens is, zoals hierboven uiteengezet, eerst per groepsindeling een tweegroepenmodel gepast met in beide groepen voor elke taaltestsomscore ongelijke residuen, gelijke intercepten en gelijke factorladingen op de factor waartoe de taaltest behoort (zie tabel 23). Dit is het restrictieve model dat uitgaat van het in beide groepen meten van gelijke vaardigheden op een gelijke schaal, maar waarbij de betrouwbaarheid van de metingen wel over groepen mag verschillen. De residuele varianties worden dus wel vrij geschat in de onderscheiden groepen. Dit houdt dus in dat de gestandaardiseerde factorladingen in de groepen wel mogen verschillen in het model, maar de ongestandaardiseerde factorladingen niet.
Vervolgens is hetzelfde model nogmaals gepast, maar nu zonder de restrictie dat de intercepten van de variabelen in de beide groepen aan elkaar gelijk zijn. Zoals gezegd in het voorgaande, als dit laatste model significant beter past dan het restrictievere model met gelijke intercepten, dan is aangetoond dat de toetsen in de onderscheiden groepen ongelijke scoreverdelingen maken. Identieke scores in beide groepen komen dan overeen met verschillende vaardigheidsniveaus. Hierna is getoetst of een model met ongelijke residuen, ongelijke intercepten en ongelijke factorladingen significant beter past dan een model met alleen ongelijke residuen en intercepten. Als dit het geval blijkt, meten de toetsen ongelijke vaardigheden in de beide onderscheiden groepen.

In het onderstaande rapporteren we per groepsindeling (sekse, thuistaal, leerlinggewicht en leeftijd) de resultaten van de analyses gericht op het detecteren van meetinvariantie.

Meetinvariantie naar sekse

De eerste meergroepenanalyse is die met de groepsvariabele sekse. In dit model wordt dus onderzocht of de taaltests partijdig zijn voor jongens of meisjes. De resultaten tonen dat het restrictieve model (gelijke intercepten, gelijke factorladingen, ongelijke residuele varianties) goed past (N=883; Nmeisje=437; Njongen=446; χ^2 =195.673; df=98; p=.000; CFI=.941; TLI=.934; RMSEA=.048; 95% c.i. RMSEA= .038-.057; SRMR=.093). Een probleem is echter dat de psi-matrijs[21] in de groep jongens niet positief definiet blijkt. Passen we hetzelfde model met ongelijke intercepten, dan vinden we uiteraard weer een goede fit (N=883; Nmeisje=437; Njongen=446; χ^2 =178.172; df=90; p=.000; CFI=.947; TLI=.935; RMSEA=.047; 95% c.i. RMSEA= .037-.057; SRMR=.087). Weer echter blijkt de psi-matrijs niet positief definiet. Een model waarbij niet alleen de intercepten over groepen verschillen maar ook de ongestandaardiseerde factorladingen ongelijk zijn in beide groepen, past wederom goed (N=883; Nmeisje=437; Njongen=446; χ^2 =158.129; df=82; p=.000; CFI=.954; TLI=.938; RMSEA=.046; 95% c.i. RMSEA= .035-.057; SRMR=.077), en weer is de psi-matrijs in de groep jongens niet positief definiet. Bij alle drie de modeltoetsingen wordt aangegeven dat het probleem wordt veroorzaakt door de derde latente factor, de technisch lezen-factor dus. Mogelijk wordt deze factor door de andere twee factoren samen perfect voorspeld, wat kan leiden tot een niet-definiete matrijs.

In een poging om een meergroepenmodel naar sekse te vinden zonder het probleem van een niet definiete psi-matrijs, is een tweegroepenmodel gefit. De enige verandering ten opzichte van het model in tabel 23 is dat in dit model de derde factor, de technisch lezen-factor, is verwijderd.

Het eerste tweefactormodel, met over groepen ongelijke residuen, gelijke intercepten en factorladingen, past goed (N=883; Nmeisje=437; Njongen=446; χ^2 =94.615; df=50; p=.000; CFI=.958; TLI=.953; RMSEA=.045; 95% c.i. RMSEA= .031-.059; SRMR=.089). Wel blijken beide factoren in de groep jongens erg hoog

[21] De covariantie-matrijs van latente variabelen.

samen te hangen (.998). Bij meisjes is de correlatie tussen beide latente constructen veel lager (.821).

Een tweede modeltoets met ongelijke residuen, ongelijke intercepten en gelijke factorladingen voor de beide groepen past weer goed (N=883; Nmeisje=437; Njongen=446; χ^2 =83.403; df=44; p=.000; CFI=.963; TLI=.952; RMSEA=.045; 95% c.i. RMSEA= .030-.060; SRMR=.083). Nu blijkt echter weer de psi-matrijs in de groep jongens niet positief definiet. De reden is ook helder, beide latente constructen vertonen een correlatie van 1.00. Bij meisjes is deze correlatie slechts .822. Eigenlijk zien we dus nu dat de structuur van het model in beide groepen enigszins verschilt. Als we het derde model toetsen waarin zowel residuen als intercepten en factorladingen ongelijk geschat worden in beide groepen (N=883; Nmeisje=437; Njongen=446; χ^2 =78.065; df=38; p=.000; CFI=.962; TLI=.944; RMSEA=.049; 95% c.i. RMSEA= .033-.064; SRMR=.075) zien we weer een goede passing. Gelukkig is de psi-matrijs in de groep jongens nu wel positief definiet. De correlatie tussen beide latente constructen is nu in de groep jongens .999 en in de groep meisjes .810. We kunnen dus nu wel het eerste model (alleen ongelijke residuen in beide groepen) en het laatste model (ongelijke residuen, intercepten en factorladingen) vergelijken en kijken of het laatste model significant beter fit dan het eerste. Dit blijkt niet het geval te zijn ($\Delta\chi^2$ =16.550; Δdf=12; p>.05).

Hoewel we niet kunnen toetsen of een model met in beide groepen ongelijke residuen, ongelijke intercepten en gelijke factorladingen significant beter past dan een model met ongelijke residuen en gelijke intercepten en factorladingen, dus hoewel we niet kunnen toetsen of we in beide groepen op dezelfde schaal meten, kunnen we wel stellen dat een model met alle drie groepen parameters ongelijk geschat in beide groepen niet significant beter fit dan het restrictievere model met alleen ongelijke residuen. We kunnen dus concluderen dat er geen partijdigheid naar sekse optreedt in de toetsen, al is het wel opvallend dat de latente constructen bij jongens veel hoger samenhangen dan bij meisjes.

Meetinvariantie naar thuistaal

De volgende groepsvariabele voor het toetsen van meetinvariantie is de variabele 'thuistaal'. Deze variabele onderscheidt drie groepen: leerlingen die thuis Nederlands spreken, leerlingen die thuis een dialect of Fries spreken en als derde groep leerlingen die thuis een vreemde taal spreken. Omdat het aantal dialect/ Fries-

sprekende leerlingen in de steekproef relatief klein is, is ervoor gekozen een tweegroepenmodel te draaien met de leerlingen die thuis een vreemde taal spreken in de ene groep en de overige leerlingen in de andere groep. Deze keus is ook inhoudelijk goed te verdedigen. Leerlingen die uit een Nederlands gezin komen zullen over het algemeen meer affiniteit met het Nederlands hebben, al spreken zij thuis Fries of dialect, dan leerlingen die thuis een vreemde taal spreken. Met de keuze die we nu maken verdelen we de steekproef dus in leerlingen voor wie het Nederlands een tweede taal is en de leerlingen die Nederlands (min of meer) als moedertaal spreken.

Het restrictieve tweegroepenmodel met alleen verschillende residuen past redelijk, al zijn CFI en TLI laag en is SRMR te groot (N=829; NNederl.=626; Nvreemd=203; χ^2 =214.867; df=98; p=.000; CFI=.917; TLI=.907; RMSEA=.054; 95% c.i. RMSEA= .044-.063; SRMR=.109). Het model met voor beide groepen ongelijke residuen en ongelijke intercepten past significant beter (N=829; NNederl.=626; Nvreemd=203; χ^2 =197.012; df=90; p=.000; CFI=.924; TLI=.907; RMSEA=.054; 95% c.i. RMSEA= .043-.064; SRMR=.100; $\Delta\chi^2$ =17.855; Δdf=8; p<.05). Het verschil tussen beide modellen is echter niet heel groot, gezien de ratio van Chi-kwadraat en vrijheidsgraden, die ongeveer twee bedraagt. De verschillen tussen de intercepten in beide groepen lopen, uitgedrukt in standaarddeviaties, van .29 (OBIS-leesvaardigheid woorden) tot 1.29 (Peabody). Deze effectgroottes lopen dus van gemiddeld tot groot.

Een model met ook nog verschillende factorladingen in beide groepen past significant beter dan het vorige model (N=829; NNederl.=626; Nvreemd=203; χ^2 =172.864; df=82; p=.000; CFI=.935; TLI=.913; RMSEA=.052; 95% c.i. RMSEA= .041-.062; SRMR=.076; $\Delta\chi^2$ =24.148; Δdf=8; p<.01) en dit keer is de ratio van $\Delta\chi^2$ en Δdf ongeveer gelijk aan drie, wat wijst op aanzienlijk betere passing voor het minst restrictieve model.

De conclusie luidt dat aangetoond is dat de taaltoetsen die zijn opgenomen in tabel 23 bij leerlingen die thuis een vreemde taal spreken iets anders meten (en ook op een andere schaal) dan bij de andere leerlingen en dat de onterechte scoreverschillen die door het schaalverschil optreden, vrij groot zijn.

Meetinvariantie naar leerlinggewicht

Als derde groepsindeling gebruiken we de variabele leerlinggewicht zoals die gehanteerd werd ten tijde van de dataverzameling. Inmiddels zijn de leerling

gewichten in het primair onderwijs veranderd. De betekenis van de door ons gebruikte leerlinggewichten is als volgt: .90-leerlingen zijn allochtone leerlingen van laag opgeleide ouders, .25-leerlingen zijn autochtone leerlingen van laag opgeleide ouders en .00-leerlingen zijn de overige leerlingen[22]. Het ligt voor de hand om hier een driegroepenmodel te draaien. Echter, het aantal .25-leerlingen is hiervoor te klein. Om de groepsgroottes redelijk te houden, is besloten eerst een tweegroepenmodel te passen voor de groep .90-leerlingen versus de overige leerlingen en daarna hetzelfde te doen voor de .00-leerlingen versus de overige leerlingen.

Het tweegroepenmodel voor de .90-leerlingen versus de overige leerlingen met ongelijke residuen, gelijke factorladingen en gelijke intercepten past redelijk (N=817; N.00- en .25-lln=569; N.90-lln=248; χ^2 =208.991; df=98; p=.000; CFI=.920; TLI=.910; RMSEA=.053; 95% c.i. RMSEA= .043-.063; SRMR=.106). Eenzelfde model met vrijgemaakte intercepten past niet significant beter (N=817; N.00- en .25-lln=569; N.90-lln=248; χ^2 =198.327; df=90; p=.000; CFI=.922; TLI=.904; RMSEA=.054; 95% c.i. RMSEA= .044-.065; SRMR=.102; $\Delta\chi^2$ =10.664; Δdf=8; p>.250). Het vrijmaken van de intercepten is dus niet nodig voor de passing. Een model met verschillende intercepten en verschillende factorladingen past wel significant beter (N=817; N.00- en .25-lln=569; N.90-lln=248; χ^2 =179.710; df=82; p=.000; CFI=.929; TLI=.905; RMSEA=.054; 95% c.i. RMSEA= .043-.065; SRMR=.082; $\Delta\chi^2$ =18.617; Δdf=8; p<.05) dan het model met alleen ongelijke residuen en intercepten. Het verschil in passing laat zien dat in beide groepen, .90- versus de overige leerlingen, verschillende constructen gemeten worden.

Het volgende tweegroepenmodel betreft de .00-leerlingen versus de overige leerlingen. Het restrictieve model met ongelijke residuen en identieke intercepten en factorladingen in beide groepen past redelijk (N=817; N.00-lln=437; N.25- en .90-lln=380; χ^2 =179.970; df=98; p=.000; CFI=.940; TLI=.933; RMSEA=.045; 95% c.i. RMSEA= .035-.056; SRMR=.122) en het model met ongelijke intercepten en gelijke factorladingen in beide groepen past niet significant beter (N=817; N.00-lln=437; N.25- en .90-lln=380; χ^2 =164.562; df=90; p=.000; CFI=.946; TLI=.934; RMSEA=.045; 95% c.i. RMSEA= .034-.056; SRMR=.110; $\Delta\chi^2$

[22] In de steekproef komen geen .40-leerlingen voor en maar drie .70-leerlingen.

=15.408; Δdf=8; p>.05). De passing van het model met ongelijke intercepten en ongelijke factorladingen in beide groepen past wel significant beter dan beide voorgaande modellen (N=817; N.00-lln=437; N.25- en .90-lln=380; χ^2 =143.396; df=82; p=.000; CFI=.955; TLI=.940; RMSEA=.043; 95% c.i. RMSEA= .031-.054; SRMR=.078). Het verschil tussen het model met alleen ongelijke residuen is significant ($\Delta\chi^2$ =36.574; Δdf=16; p<.005) en het verschil met het model met ongelijke residuen en ongelijke intercepten ook ($\Delta\chi^2$ =21.166; Δdf=8; p<.005).
We moeten dus concluderen dat de taaltoetsen uit tabel 23 bij .00-leerlingen iets anders meten dan bij de .25- en .90-leerlingen.

Meetinvariantie naar leeftijd

De volgende toetsing betreft een meergroepenmodel gebaseerd op leeftijd. Voor deze analyse is de totale groep leerlingen gesplitst op de mediaan van de variabele leeftijd (6.09 jaar). Het restrictieve tweegroepenmodel past matig (N=883; Nonder mediaan=439; Nboven mediaan =444; χ^2 =198.483; df=98; p=.000; CFI=.913; TLI=.902; RMSEA=.048; 95% c.i. RMSEA= .038-.058; SRMR=.122). Als we nu dit meergroepenmodel toetsen met niet alleen de residuen, maar ook de intercepten apart geschat, dan zien we de passing veel beter worden (N=883; Nonder mediaan=439; Nboven mediaan =444; χ^2 =172.941; df=90; p=.000; CFI=.928; TLI=.912; RMSEA=.046; 95% c.i. RMSEA= .035-.056; SRMR=.118). Het verschil in fit tussen beide voornoemde modellen is significant ($\Delta\chi^2$ =25.542; Δdf=8; p<.005). De onterechte verschillen in de intercepten tussen beide groepen lopen van .51 (TAK-klankarticulatie) tot 1.08 (TvK). Deze effectgroottes zijn dus aanzienlijk.
In het derde getoetste model met per groep apart geschatte residuen, intercepten en factorladingen, zien we de fit weer veel beter worden (N=883; Nonder mediaan=439; Nboven mediaan =444; χ^2 =124.663; df=82; p=.000; CFI=.963; TLI=.950; RMSEA=.034; 95% c.i. RMSEA= .021-.046; SRMR=.082). Het model past nu goed. Het verschil in fit tussen beide voornoemde modellen is dan ook weer significant ($\Delta\chi^2$ =48.278; Δdf=8; p<.000). De taaltoetsen meten ook in de hier onderscheiden groepen dus weer verschillende constructen.
Uit de bovenstaande resultaten blijkt zeer duidelijk dat als we een tweegroepenmodel passen waarbij we splitsen op de mediane leeftijd, het model alleen dan een goede fit vertoont als we residuen, intercepten en factorladingen in beide groepen ongelijk schatten. Dit betekent dat de taaltests bij leerlingen jonger dan

zes jaar iets anders meten dan bij leerlingen ouder dan zes. Duidelijk is dat we de taaltests dus niet kunnen afnemen bij leerlingen jonger dan zes en leerlingen ouder dan zes om vervolgens de scores van leerlingen in beide groepen onderling te vergelijken. Een bijkomend probleem is dat we (nog) niet weten waardoor deze verschillen optreden. We weten dus ook niet of de toetsen in één van beide groepen wel of meer valide zijn dan in de andere groep.

Conclusie

De conclusie van de analyses naar de mate van meetinvariantie luiden dat de taaltoetsen die zijn opgenomen in tabel 23 niet meetinvariant zijn voor thuistaal, leerlinggewicht en leeftijd. De partijdigheid voor groepen leerlingen met een verschillend leerlinggewicht betreft alleen het gemeten construct. De taaltoetsen meten bij de onderscheiden groepen niet hetzelfde. Voor thuistaal blijkt zowel op een andere schaal als een ander construct gemeten te worden in de beide onderscheiden groepen. De partijdigheid naar leeftijd is het ernstigst en betreft ook zowel schaalverschillen als het meten van verschillende constructen.

De in het bovenstaande gerapporteerde resultaten aangaande meetinvariantie gaan uit van het toetsen van de verbetering in de passing van geneste modellen met behulp van de $\chi 2$. Gegeven de grootte van de gebruikte steekproeven is de power van de $\chi 2$ toets echter erg groot. Om die reden zijn, voorzichtigheidshalve, de fitverbeteringen in de bovengerapporteerde modellen nogmaals getoetst, maar nu met een veel minder strenge toets (Dudgeon, 2003). De toets van Dudgeon (2003) gaat uit van het verschil in RMSEA van beide geneste modellen, zoals voorgesteld door Browne en Du Toit (1992). Browne en Du Toit (1992) ontwikkelden de Root Deterioration per Restriction (RDR) statistiek voor het vergelijken van geneste modellen indien er sprake is van één groep en Dudgeon (2003) breidde deze techniek uit naar meergroepenmodellen, zoals wij hier gebruiken.

Als we dit veel minder strenge criterium hanteren om te bepalen of er sprake is van meetinvariantie, dan blijkt dat we alleen de meetinvariantie naar leeftijd moeten verwerpen voor wat betreft het gemeten construct ($p=.0006$). Het verschil in passing tussen een model met in beide groepen ongelijke residuen, ongelijke intercepten en gelijke factorladingen enerzijds en een model met ongelijke residuen, ongelijke intercepten en ongelijke factorladingen anderzijds, is ook volgens deze minder strenge toets significant. Dit betekent dat ook als we een coulant criterium hanteren, we moeten concluderen dat de hier onderzochte taaltoetsen niet meetin-

variant zijn naar leeftijd. Bij leerlingen respectievelijk ouder en jonger dan zes jaar meten de toetsen ongelijke vaardigheden. Ook als we dit coulante criterium hanteren blijken de toetsen dus ongeschikt om leerlingen jonger en ouder dan zes op één schaal op een eerlijke wijze te oormerken als zijnde wel of niet taalzwak. Volgens de RDR-index is de meetinvariantie voor de groepsindeling naar leerlinggewicht of thuistaal dus niet geschonden.

Deze resultaten betekenen dat de toetsen bij leerlingen uit de verschillende voornoemde groepen volgens de strengste norm (χ2-verandering) niet hetzelfde meten en dus niet gebruikt kunnen worden om leerlingen uit deze verschillende groepen te rangschikken naar taalvaardigheid op één en dezelfde schaal. Met name als er verschillende constructen gemeten worden in de onderscheiden groepen moeten we concluderen dat de toetsen ongeschikt zijn om leerlingen uit de verschillende groepen op een eerlijke en vergelijkbare wijze te oormerken als al dan niet taalvaardig. Gaan we uit van de coulante RDR, dan blijkt alleen de meetinvariantie naar leeftijd geschonden. Volgens de RDR-toets meten de taaltoetsen bij leerlingen respectievelijk jonger en ouder dan zes jaar verschillende constructen.

Nu is uiteraard de vraag hoe ernstig dit gebrek aan meetinvariantie is in de praktijk. In welke mate gaan scores nu onterecht verschillen? Naast de hierboven al gegeven toetsresultaten, is voor thuistaal en leeftijd voor alle toetsen apart de effectmaat in standaarddeviaties berekend (zie tabel 24). Het betreft hier alleen gestandaardiseerde verschillen in intercepten tussen beide groepen. De berekening is als volgt: In het model met de intercepten apart geschat per groep, krijgen leerlingen onder en boven de zes jaar voor de OBIS-passieve woordkennis respectievelijk 21.522 en 26.806 als intercept, een verschil van 5.28. Om het effect uit te drukken in standaarddeviaties, kijken we naar de standaarddeviatie van de variabele OBIS-passieve woordkennis, welke 6.99 bedraagt (N=344). Duidelijk is dat het onterechte verschil tussen de schattingen voor leerlingen respectievelijk jonger en ouder dan zes gemiddeld .76 standaarddeviatie bedraagt, als we alleen ongelijke intercepten nodig hadden in beide groepen. We weten echter dat ook de factorladingen in beide groepen significant verschillen. Dit impliceert niet alleen dat het onterechte verschil tussen beide groepen nog groter is dan driekwart standaarddeviatie, maar ook dat we in beide groepen ongelijke constructen hebben gemeten, wat uiteraard een veel groter probleem oplevert.

Om een overzicht per taaltoets te krijgen van de ernst van het gevonden gebrek aan meetinvariantie is in tabel 24 aangegeven hoe groot de verschillen in intercept

zijn tussen beide groepen voor de groepsindeling naar zowel de thuistaal (vreemde taal of niet) als leeftijd (4-6 jaar versus 6-8 jaar). Om de gevonden verschillen in intercept over taaltoetsen heen onderling te kunnen vergelijken, zijn de verschillen gestandaardiseerd door ze te delen door de standaarddeviatie van de betreffende taalscore. Naast deze gestandaardiseerde verschillen geven we ook de gemiddelden en de standaarddeviaties van de variabelen. Deze gepresenteerde effectgrootten zijn onderschattingen van de werkelijke effectgrootten, daar we de invloed van verschil in factorladingen hier buiten beschouwing laten. Overigens kunnen we het verschil in factorladingen eigenlijk ook niet meer in een effectgrootte uitdrukken, omdat we dan verschillende grootheden proberen te vergelijken. Als de factorladingen significant verschillen over groepen meten we immers verschillende constructen in de beide groepen en vooral voor de twee gehanteerde leeftijdsgroepen blijkt dat het geval te zijn.
Voor de partijdigheid die we vonden voor de groepsindelingen naar leerlinggewicht geven we geen effectgrootten gebaseerd op de verschillende intercepten, omdat voor deze groepsindelingen bleek dat alleen het model met apart geschatte residuen, intercepten en factorladingen significant beter past dan het model met alleen apart geschatte residuen en intercepten. We weten dus dat we in de groepen gevormd door de leerlinggewichten verschillende constructen meten, maar daar het model met apart geschatte residuen en intercepten niet significant beter past dan een model met alleen apart geschatte residuen, is het berekenen van de effectmaat gebaseerd op verschillende intercepten hier misleidend.
In tabel 24 kunnen we zien dat de onterechte verschillen in de scores van de beide leeftijdsgroepen variëren van een derde tot meer dan een hele standaarddeviatie, wat wijst op een groot onterecht verschil in scores tussen de oudere en de jongere leerlingen. Let wel, dit verschil betreft niet het verschil in vaardigheidsniveau tussen oudere en jongere leerlingen, het is echt een onterecht extra verschil, dat niet veroorzaakt wordt door een verschil in vaardigheid. Er is dus een systematische schattingsfout van ongeveer een derde tot een hele standaarddeviatie groot gerelateerd aan de leeftijd van de leerling, zonder zelfs de verschillende factorladingen in de berekening op te nemen. De onterechte verschillen zijn in werkelijkheid dus nog groter daar we bij jonge en oudere leerlingen met dezelfde toetsen ook nog verschillende constructen (hier dus verschillende aspecten van taalvaardigheid) meten.

Tabel 24 Gegeven worden de gemiddelde scores (M), de standaarddeviatie (s.d.), de steekproefgroottes (N), en de grootte van het verschil in intercepten van beide groepen leerlingen gedeeld door de standaarddeviatie van de variabele ofwel de effectgrootte (D)

Subtest	M	s.d.	N	D thuistaal (Nederlands of vreemde taal)	D leeftijd (onder en boven 6 jaar)
TvK	24.33	5.07	341	1.023	1.079
Peabody	71.05	15.65	349	1.285	.926
TAK - klankonderscheiding	41.76	8.52	352	.723	.795
TAK - klankarticulatie	42.14	3.95	352	.425	.508
TAK – passieve woordenschat	52.66	18.93	352	1.226	.909
OBIS – passieve woordenschat	24.46	6.99	344	1.208	.756
OBIS - klankarticulatie	6.49	2.06	344	.516	.590
OBIS - klankonderscheiding	6.42	3.12	344	.806	.813
OBIS - kennis van letters	6.61	7.94	344	.364	.704
OBIS - leesvaardigheid woorden	2.33	3.25	344	.286	.813
OBIS – kennis van getallen	9.07	5.03	344	.340	.919

Omdat het gebrek aan meetinvariantie over de onderscheiden leeftijdsgroepen ernstig is, is nader onderzoek gedaan naar de modelfit per leeftijdsgroep. Nu blijkt uit deze analyses dat het model bij leerlingen ouder dan zes jaar uitstekend en zelfs exact past (N=441; χ^2 =52.518; df=41; p=.107; CFI=.977; TLI=.969; RMSEA=.025; 95% c.i. RMSEA= .000-.044; SRMR=.080). Voor leerlingen jonger dan zes past het model echter veel minder goed (N=442; χ^2 =80.155; df=41; p=.000; CFI=.930; TLI=.906; RMSEA=.046; 95% c.i. RMSEA= .031-.062; SRMR=.085). Naast dat de taaltests in beide leeftijdsgroepen dus verschillende constructen meten op een verschillende schaal, blijkt de passing van het model in de jongste groep ook veel minder goed te zijn dan in de oudste groep.

Een mogelijke oorzaak van de matige fit bij jonge leerlingen, kan de invloed van de gehanteerde meetmethode zijn, tenminste als het effect van de gehanteerde meetmethode voor jongere leerlingen anders is dan voor oudere leerlingen. De TvK, TAK en Peabody werken alle drie met papieren versies, tekeningen en door

testleiders gesproken opdrachten. De OBIS werkt als enige toets met een p.c. Om na te gaan of een methodefactor de oorzaak zou kunnen zijn van de matige fit in de jongste groep, is in deze jongste groep nogmaals een model gepast met voor alle OBIS-toetsen naast hun lading op één van de drie eerste factoren ook nog een extra lading op een aparte methodefactor, de p.c.-factor zeg maar (N=442; χ^2 =56.821; df=35; p=.011; CFI=.961; TLI=.939; RMSEA=.038; 95% c.i. RMSEA= .018-.055; SRMR=.075). Hoewel dit vierfactormodel significant beter past dan het driefactormodel zonder methodefactor ($\Delta\chi^2$ =23.334; Δdf=6; p<.001), blijkt de extra factor ongeïdentificeerd, oftewel, alle factorladingen op deze factor blijken niet significant. Een methodefactor is dus waarschijnlijk niet de verklaring van de misfit bij jonge leerlingen.

4.7 Selectie van leerlingen met de laagste taalvaardigheid

Een ander aspect genoemd in de eerste onderzoeksvraag is in hoeverre de taaltoetsen die zijn afgenomen binnen het onderhavige onderzoek in staat zijn leerlingen met de laagste taalvaardigheid Nederlands te selecteren en welk type leerling met de toetsen geselecteerd kan worden. Het bepalen welke toetsen het meest geschikt zijn om de leerlingen met de laagste taalvaardigheid Nederlands te selecteren, blijkt deels uit de resultaten van de factoranalyses. Uitgaand van de geselecteerde toetsen moeten we onderscheid maken tussen drie aspecten van taalvaardigheid: passieve woordkennis, fonologische vaardigheid en de vaardigheid die we voor het gemak 'technisch leesvaardigheid' hebben genoemd. De beste toetsen per onderscheiden vaardigheid zijn de toetsen die het meest betrouwbaar zijn, de hoogste gestandaardiseerde factorlading vertonen (zie tabel 23) en die daarnaast de meeste spreiding vertonen, en dan met name spreiding in de onderste regionen van de vaardigheidsverdelingen. Om inzicht in dit laatste te krijgen, worden in het onderstaande per subtoets verschillende boxplots gepresenteerd.

Een andere vraag is welk type leerlingen we met de toetsen kunnen selecteren. We hebben deze vraag naar typen leerlingen in eerste instantie opgevat als de vraag naar in hoeverre de toetsen twee variabelen voorspellen: leerlinggewicht (.00, .25 en .90) en thuistaal (Nederlands, dialect/Fries en een vreemde taal). De variabele 'leerlinggewicht' is de beste operationalisatie van het begrip achterstand-

leerling in onze data. De .00-leerlingen zijn de 'gewone' leerlingen, .25-leerlingen zijn leerlingen waarvan de ouders een laag opleiding- en/of beroepsniveau hebben, de .90-leerlingen zijn de allochtone leerlingen waarvan de ouders een laag opleidingsniveau hebben. De variabele 'thuistaal' voegt aan het leerlinggewicht nog iets toe. Het is immers waarschijnlijk dat leerlingen die thuis een dialect of een vreemde taal spreken, lagere scores op de taaltoetsen halen dan leerlingen die thuis de standaardvorm van het Nederlands spreken.

Om na te gaan in hoeverre de taaltoetsen de bovengenoemde groepen leerlingen kunnen onderscheiden, zijn per taaltoets boxplots per groep leerlingen gemaakt en covariantieanalyses verricht. Deze boxplots worden in het onderstaande per taaltoets gepresenteerd. Daar leerlingen random zijn toegewezen aan toetscondities en random klassen zijn getrokken uit de deelnemende scholen, is het niet waarschijnlijk dat de leeftijd van de onderscheiden groepen leerlingen een rol speelt bij de boxplot-resultaten. Toch zouden toevallige verschillen in leeftijd per onderscheiden groep de resultaten kunnen vertekenen. De leeftijden van de leerlingen in de steekproef variëren van 4.4 tot 8.1 jaar met een gemiddelde van 6.1 jaar en een standaarddeviatie van .75.

Eerst is nagegaan of de leeftijd voor de naar thuistaal en naar leerlinggewicht onderscheiden groepen verschilt. Hiertoe is een variantieanalyse verricht met leeftijd als afhankelijke variabele en thuistaal als factor. Uit de resultaten blijkt dat thuistaal een significante samenhang vertoont met leeftijd ($F(2, 833)=8.33$; $p=.000$). Uit post-hoc analyses blijkt dat de groep dialect/Friessprekers significant ouder is dan de leerlingen die thuis Nederlands spreken ($p=.000$) en ook ouder is dan de groep vreemde taalsprekers ($p=.003$). De dialect/Friessprekende leerlingen en de leerlingen die een vreemde taal spreken, verschillen onderling niet significant qua leeftijd. De dialect/Friessprekende leerlingen zijn gemiddeld ongeveer vier maanden ouder dan de overige leerlingen.

Doen we vergelijkbare analyses voor de groepen ingedeeld naar leerlinggewicht, dan vinden we een niet-significant resultaat ($F(2, 819)=2.13$; $p=.120$). Wel blijkt de aanname van gelijke varianties geschonden (Levene's statistic=5.49; df1=2; df2=819; $p=.004$). Uiteraard is een non-parametrische toets echter evenmin significant (Kruskal-Wallis $\chi 2=3.79$; df=2; $p=.150$), zodat we toch kunnen concluderen dat de indeling naar leerlinggewicht niet samenhangt met de leeftijd van de leerlingen.

We kunnen dus concluderen dat bij de boxplots naar thuistaal de groep dialect of Friessprekers een klein beetje ouder is dan de groep die thuis Nederlands spreekt en dat bij de boxplots naar leerlinggewicht verschil in leeftijd geen rol van betekenis speelt.
Om na te gaan in hoeverre de gemiddelden op de taaltoetsen verschillen over leerlingen met een verschillend leerlinggewicht, dan wel over leerlingen die thuis Nederlands, een dialect of Fries of een vreemde taal spreken, zijn zoals gezegd bij de boxplots covariantie-analyses verricht. Telkens is als covariaat de leeftijd opgenomen, zodat de resultaten niet kunnen worden geweten aan leeftijdsverschillen tussen de onderscheiden groepen. Als factoren zijn telkens het leerlinggewicht dan wel de thuistaal gebruikt. De resultaten van deze covariantie-analyses staan bij elke boxplot vermeld, evenals de ongecorrigeerde en de voor leeftijd gecorrigeerde gemiddelde scores. De partiële eta^2 in de tabellen geeft de proportie door de betreffende variabele verklaarde variantie in de afhankelijke variabele, waarbij de andere onafhankelijke variabele (leeftijd) constant wordt gehouden. De aanname dat de errorvarianties in de onderscheiden groepen gelijk zijn, wordt getoetst met Levene's test, waarvan de resultaten onder elke tabel vermeld staan. De taaltoetsen die het best in staat zijn om de verschillende naar leerlinggewicht en thuistaal onderscheiden groepen te oormerken, zijn de toetsen waarbij de partiele gekwadrateerde eta van respectievelijk leerlinggewicht dan wel thuistaal het hoogst is.
Naast de hierboven beschreven analyses is ook per (sub)toets gekeken welke leerlingen de 25% slechtst presterende leerlingen op de betreffende (sub)toets zijn. Hiertoe wordt aangegeven welke percentages .90- en .25-leerlingen, meisjes en leerlingen die thuis een vreemde taal spreken voorkomen bij de 25% slechtst presterende leerlingen en eveneens voor dezelfde categorieën hoeveel procent uit deze categorie in het onderste kwartiel terecht komt (zie tabel 68).

Boxplots en covariantie-analyses

De boxplots voor de TvK staan in figuur 3 en 4. De boxplots verdelen de scores in vier delen. De leerlingen met scores behorend bij respectievelijk het kwart hoogste en het kwart laagste scores, worden weergegeven door de 'whiskers', de dunne verticale lijnen die begrensd worden door een horizontaal streepje. De helft van de leerlingen die overblijft, en die dus rond de mediaan scoren, worden gerepresenteerd door de grijze rechthoek. De zwarte streep die horizontaal door deze

rechthoek loopt, is de mediaan (de score waar de éne helft van de leerlingen boven en andere helft onder scoort).

Uit figuur 3 blijkt dat de TvK goed discrimineert tussen leerlingen van verschillende vaardigheid. De boxplots laten in geen van de drie groepen plafond- of bodemeffecten zien. Ook blijkt dat de .00-leerlingen gemiddeld het hoogst scoren en de .90-leerlingen het laagst. De .25-leerlingen scoren gemiddeld iets lager dan de .00-leerlingen. Verder is het opvallend dat de beste .25-leerlingen net zo hoog scoren als de beste .00-leerlingen. De slechtst presterende .00-leerlingen scoren zelfs veel lager dan de slechtst scorende .25-leerlingen. Mogelijk wordt dit veroorzaakt door de iets hogere gemiddelde leeftijd van de .25-leerlingen. De hoogste scores van .90-leerlingen liggen iets lager dan de hoogste scores van .00- en .25-leerlingen, maar het verschil is gering. Wel zijn de laagste scores van .90-leerlingen veel lager dan die van de .00- en .25-leerlingen.

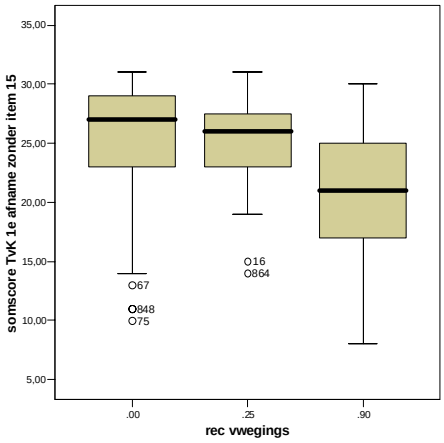
Kijken we naar tabel 25 dan zien we dat de leeftijd significant samenhangt met de scores op de TvK en dat de groepen naar leerlinggewicht na correctie voor leeftijd eveneens significant verschillende scores laten zien. De door het leerlinggewicht verklaarde proportie variantie in de TvK-scores, na controle voor leeftijdsverschillen, bedraagt .28 (zie partial eta^2). Krap een derde van de verdeling over TvK scores valt dus te voorspellen uit het leerlinggewicht. Dit betekent dus ook dat tweederde van de variatie in de TvK-scores niet met de groepsindeling naar leerlinggewicht samenvalt. Ook aan de boxplots is dit te zien. Duidelijk is dat als we een taaltoets zouden willen gebruiken voor een gewichtenregeling, de leerlingen die extra financiering zouden krijgen voor een aanzienlijk deel andere leerlingen zijn dan de leerlingen die momenteel[1] extra gefinancierd worden. Ook valt op dat het voor leeftijd gecorrigeerde verschil tussen .00- en .25-leerlingen vrijwel nihil is.

In Figuur 4 zien we de TvK-boxplots voor leerlingen die thuis Nederlands spreken, een dialect of Fries, dan wel een vreemde taal. Uit Figuur 4 blijkt dat de TvK ook voor de drie groepen die we creëren op grond van de thuistaal ook redelijk discrimineert. We zien dat dialect/Friessprekers relatief weinig lage scores vertonen en dat de verschillen in prestatie voor de vreemde taalsprekers het grootst

[1] Waar we spreken over de huidige gewichtenregeling, bedoelen we de regeling die indeelt in de gewichten .00-, .25-. .40-, .70- en .90-leerlingen (ook wel aangeduidt als 1.00, 1.25, 1.40, 1.70 en 1.90). Deze regeling is inmiddels vervangen door een andere gewichtenregeling.

zijn. Ook is opvallend dat de hoogst scorende vreemde taalsprekers nauwelijks lager scoren dan de beste Nederlandssprekende of dialect/Friessprekende leerlingen.

Figuur 3 - Boxplots TvK naar leerlinggewicht

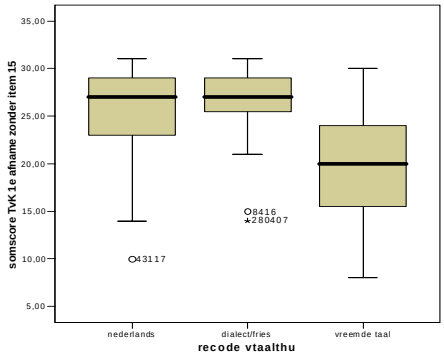


Tabel 25 Resultaten covariantie-analyse TvK - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	101.915	.000	.501
intercept	.023	.879	.000
leeftijd	189.504	.000	.383
leerlinggewicht	60.307	.000	.283
Gemiddelden	ongecorr.	gecorrig.	N
.00	25.6	25.9	169
.25	25.2	24.6	51
.90	20.7	20.6	89

Levene's test Eq. of Error-Var.: F=2.111 ; df1=2 ; df2=306 ; p=.123.

Figuur 4 - Boxplots TvK naar thuistaal



Kijken we naar tabel 26 dan zien we dat de TvK-scores beter voorspeld worden door de variabele thuistaal dan door het leerlinggewicht. De proportie door thuistaal verklaarde variantie in de TvK-scores bedraagt .33. Aan de gecorrigeerde gemiddelden is te zien dat de leerlingen die thuis Nederlands spreken na correctie voor leeftijdsverschillen nauwelijks verschillen van leerlingen die thuis een dialect of Fries spreken. De TvK blijkt dus voor wat betreft de leerlinggewichten alleen enigszins te onderscheiden tussen .90- en de overige leerlingen. De onderscheiding naar thuistaal voorspelt de Tvk beter dan de onderscheiding naar leerlinggewicht.

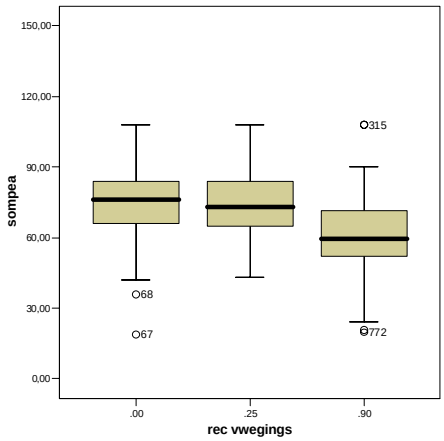
Tabel 26 Resultaten covariantie-analyse TvK - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	126.213	.000	.542
intercept	.239	.625	.001
leeftijd	197.382	.000	.382
thuistaal	80.108	.000	.334
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	25.6	25.8	193
dialect/Fries	26.8	25.3	51
vreemde taal	19.9	20.1	80

Levene's test Eq. of Error-Var.: F=3.899; df1=2; df2=321; p=.021.

In figuur 5 en 6 staan de boxplots voor de Peabody. Hieruit blijkt dat de Peabody net als de TvK geen plafond- of bodemeffecten vertoont en dus eveneens goed discrimineert in alle onderscheiden groepen. Opvallend is verder dat in vergelijking met de TvK de hoogste scores van de groep .90-leerlingen, en ook die van de vreemde taalsprekers, nu duidelijk lager uitpakken dan die van de andere groepen. Verder zien we dat de .00- en de .25-leerlingen vrijwel identieke verdelingen kennen. Het grote verschil in spreiding dat de TvK liet zien tussen .25- en .00-leerlingen is bij de Peabody afwezig.

Figuur 5 - Boxplots Peabody naar leerlinggewicht



Kijken we naar tabel 27 dan zien we dat het leerlinggewicht 22% van de variatie in de scores op de Peabody voorspelt. Weer zien we dat de .00- en de .25-leerlingen na correctie voor leeftijdsverschillen nauwelijks afwijken qua gemiddelde score.

Uit figuur 6 en tabel 28 blijkt dat de variabele thuistaal 26% van de spreiding in de Peabody-scores voorspelt. Verder zien we dat leerlingen die thuis Nederlands spreken, na correctie voor leeftijdsverschillen (tabel 28) niet verschillen van de dialect/Friessprekers.

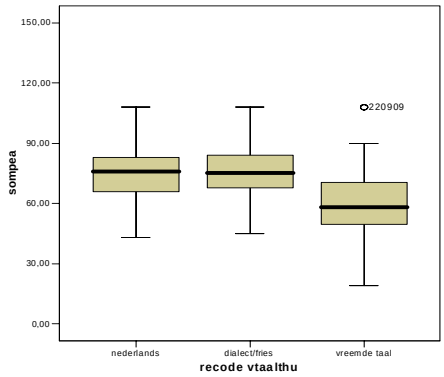
Tabel 27 Resultaten covariantie-analyse Peabody - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	69.406	.000	.400
intercept	2.080	.000	.007
leeftijd	128.536	.000	.292
leerlinggewicht	44.336	.000	.221
Gemiddelden	ongecorr.	gecorrig.	N
.00	74.7	75.2	190
.25	74.2	73.1	38
.90	60.8	60.3	88

Levene's test Eq. of Error-Var.: F=.115; df1=2 ; df2=313; p=.891.

Ook zien we, net als bij de TvK, dat de verdelingen van respectievelijk de .90-leerlingen en de vreemde taalsprekers, grotendeels overlappen met de verdelingen van de andere groepen, al scoren .90-leerlingen en leerlingen die thuis een vreemde taal spreken gemiddeld wel lager dan de andere groepen. Gewichtentoekenning op grond van deze taaltoets zal dus eveneens wezenlijk afwijken van de indeling in gewichten. Een flink deel van de .90-leerlingen scoort zelfs hoger dan de mediane scores van .00- en .25-leerlingen.

Figuur 6 - Boxplots Peabody naar thuistaal



Tabel 28 Resultaten covariantie-analyse Peabody - Thuistaal als factor, leeftijd als covariaat – Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	85.180	.000	.438
intercept	2.327	.128	.007
leeftijd	149.847	.000	.314
thuistaal	57.453	.000	.259
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	74.9	75.0	216
dialect/Fries	75.5	74.8	37
vreemde taal	59.4	58.9	79

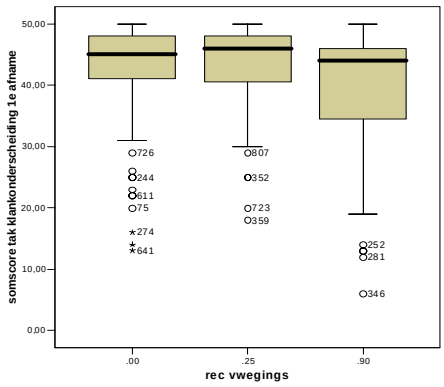
Levene's test Eq. of Error-Var.: F=2.423; df1=2; df2=329; p=.090.

In figuur 7 en 8 staan de boxplots voor de subtoets klankonderscheiding van de TAK met in tabel 29 en 30 de resultaten van de bijbehorende covariantie-analyses. Uit de boxplots blijkt dat deze subtoets relatief makkelijk is voor de leerlingen. De verdelingen zijn scheef naar boven. Ook lijken de scores voor de onderscheiden groepen elkaar niet erg te ontlopen. Wel is duidelijk dat de laagst scorende .90-leerlingen en de laagst scorende leerlingen uit de groep die thuis een

vreemde taal spreken, veel lager uitkomen dan de slechtst scorende leerlingen uit andere groepen.

Uit tabel 29 blijkt verder dat deze subtoets van de TAK niet goed door de groepsindeling naar leerlinggewicht voorspeld wordt (slechts 6% verklaarde variantie). De scores op deze subtoets vertonen dus geen samenhang met de gewichten. Voor de indeling naar thuistaal geldt min of meer hetzelfde; slechts 11% van de variantie in de scores komt overeen met de taal die thuis gesproken wordt. Wel moeten we een kanttekening maken bij de resultaten van de covariantie-analyses in tabel 29 en 30. De aanname betreffende gelijke error-varianties is bij deze beide analyse geschonden. De proporties verklaarde variantie zijn echter zo laag, dat we de conclusie handhaven dat deze subtoets nauwelijks samenhangt met de twee groepsindelingen. Opvallend is dat, ook na correctie voor leeftijdsverschillen, de dialect/Friessprekende leerlingen gemiddeld hoger scoren dan de leerlingen die thuis Nederlands spreken.

Figuur 7 - Boxplots klankonderscheiding TAK naar leerlinggewicht

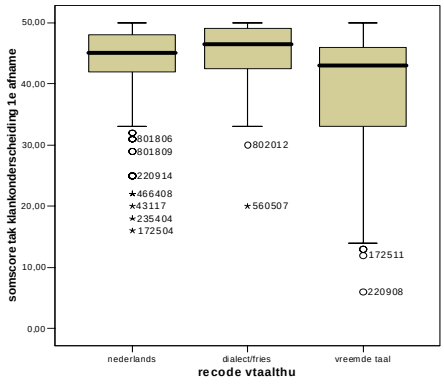


Tabel 29 Resultaten covariantie-analyse TAK-klankonderscheiding - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	P	partial eta^2
Model	30.222	.000	.225
intercept	7.428	.007	.023
leeftijd	78.008	.000	.200
leerlinggewicht	9.145	.000	.055
Gemiddelden	ongecorr.	gecorrig.	N
.00	42.6	43.0	147
.25	42.4	42.3	51
.90	39.4	39.0	119

Levene's test Eq. of Error-Var.: F=8.706; df1=2 ; df2=314; p=.000.

Figuur 8 - Boxplots klankonderscheiding TAK naar thuistaal



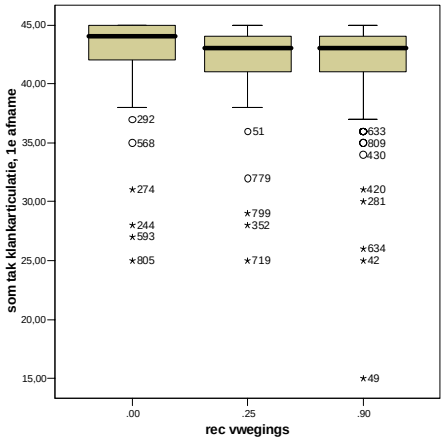
Tabel 30 Resultaten covariantie-analyse TAK-klankonderscheiding - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	43.029	.000	.293
intercept	7.415	.007	.023
leeftijd	98.472	.000	.240
thuistaal	19.846	.000	.113
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	43.3	43.6	188
dialect/Fries	44.5	44.1	36
vreemde taal	38.6	38.1	92

Levene's test Eq. of Error-Var.: F=15.252; df1=2; df2=313; p=.000.

In figuur 9 en 10 staan de boxplots voor de subtoets klankarticulatie van de TAK. Duidelijk is dat ook deze subtoets van de TAK nauwelijks discrimineert tussen de verschillende onderscheiden groepen. Uit tabel 31 en 32 blijkt dit eveneens: we vinden slechts vier en 3% verklaarde variantie in de scores door respectievelijk leerlinggewicht en thuistaal. De conclusie die we trekken op grond van de resultaten is dat de klankarticulatietoets van de TAK niet samenhangt met de groepsindeling naar leerlinggewicht of thuistaal.

Figuur 9 - Boxplots klankarticulatie TAK naar leerlinggewicht



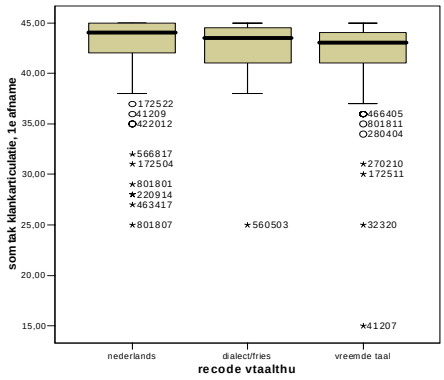
In Figuur 11 en 12 staan de boxplots voor de subtoets passieve woordkennis van de TAK. Uit de boxplots kunnen we lezen dat de subtoets passieve woordkennis geen plafond- of bodemeffecten vertoont in de onderscheiden groepen. Verder valt op dat de toets veel variatie vertoont in de scores voor .00- en .90-leerlingen. De beste .90-leerlingen scoren net zo hoog als de beste .00-leerlingen en de beste vreemde taalsprekers scoren zelfs hoger dan de beste thuis Nederlands sprekende leerlingen.

Tabel 31 Resultaten covariantie-analyse TAK-klankarticulatie - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	15.502	.000	.129
intercept	331.673	.000	.514
leeftijd	37.218	.000	.106
leerlinggewicht	6.385	.002	.039
Gemiddelden	ongecorr.	gecorrig.	N
.00	42.8	42.9	147
.25	41.5	41.5	51
.90	41.5	41.4	119

Levene's test Eq. of Error-Var.: F=3.086; df1=2 ; df2=314; p=.047.

Figuur 10 -: Boxplots klankarticulatie TAK naar thuistaal



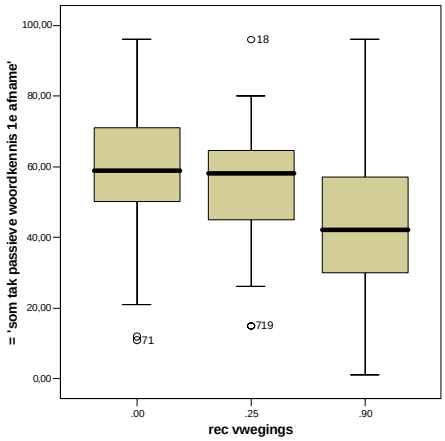
Tabel 32 Resultaten covariantie-analyse TAK-klankarticulatie - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	17.379	.000	.143
intercept	338.760	.000	.521
leeftijd	45.055	.000	.126
thuistaal	5.404	.005	.033
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	42.6	42.8	188
dialect/Fries	42.4	42.3	36
vreemde taal	41.4	41.3	92

Levene's test Eq. of Error-Var.: F=.973; df1=2; df2=313; p=.379.

Kijken we naar de resultaten van de covariantie-analyses, dan zien we dat de passieve woordkennistoets van de TAK voor 23% verklaard wordt door leerlinggewicht en zelfs voor 34% door thuistaal. We zien dat deze subtoets van de TAK voor wat betreft de covariantie-analyses betere resultaten geeft dan de Peabody en iets minder mooie resultaten dan de TvK. Opvallend is verder dat de .00- en de .25-leerlingen nu wel verschillen qua gemiddelde, al zien we de grootste verschillen tussen de .90- en de overige leerlingen. De indeling naar thuistaal toont dat dialect/Friessprekers ook na correctie voor leeftijdsverschillen hoger scoren dan de leerlingen die thuis Nederlands spreken.

Figuur 11 - Boxplots passieve woordkennis TAK naar leerlinggewicht



Tabel 33 Resultaten covariantie-analyse TAK-passieve woordkennis - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

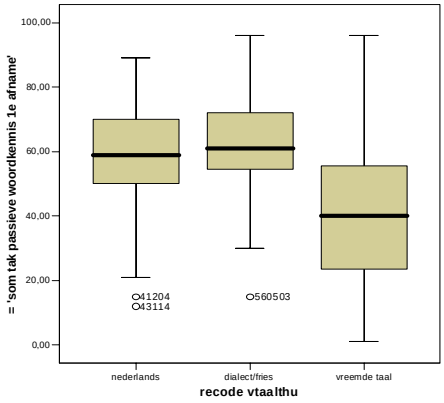
Variantiebron	F	p	partial eta^2
Model	67.485	.000	.393
intercept	12.950	.000	.040
leeftijd	124.595	.000	.285
leerlinggewicht	47.567	.000	.233
Gemiddelden	ongecorr.	gecorrig.	N
.00	58.6	59.5	147
.25	54.8	54.4	51
.90	42.5	41.6	119

Levene's test Eq. of Error-Var.: F=4.822; df1=2 ; df2=314; p=.009.

In Figuur 13 en 14 staan de boxplots voor de subtoets aanvankelijk schrijven van de OBIS. Een opvallend verschil met de vorige boxplots is dat alle verdelingen de maximale range vertonen. Kijken we naar de resultaten van de covariantie-analyses (tabel 35 en 36) dan is echter meteen duidelijk dat deze subtoets niet dif-

ferentieert tussen de onderscheiden groepen (respectievelijk 6 en 3% verklaarde variantie).

Figuur 12 - Boxplots passieve woordkennis TAK naar thuistaal

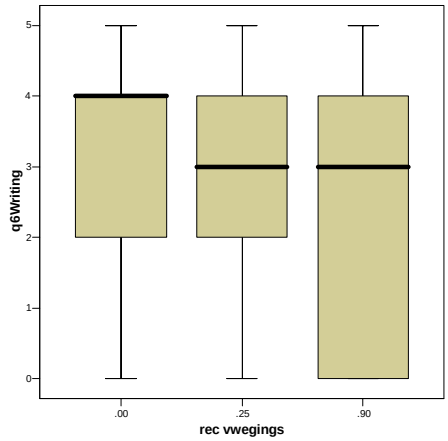


Tabel 34 Resultaten covariantie-analyse TAK-passieve woordkennis - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	97.894	.000	.485
intercept	18.286	.000	.055
leeftijd	158.832	.000	.337
thuistaal	79.455	.000	.337
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	58.4	59.3	188
dialect/Fries	61.4	60.3	36
vreemde taal	39.2	37.9	92

Levene's test Eq. of Error-Var.: F=7.236; df1=2; df2=313; p=.001.

Figuur 13 - Boxplots aanvankelijk schrijven van de OBIS naar leerlinggewicht

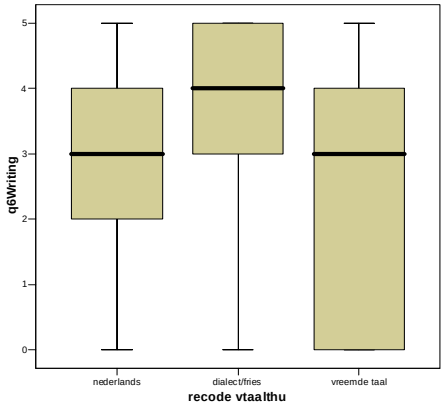


Tabel 35 Resultaten covariantie-analyse OBIS-aanvankelijk schrijven - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	50.488	.000	.331
intercept	52.329	.000	.146
leeftijd	129.420	.000	.297
leerlinggewicht	10.301	.000	.063
Gemiddelden	ongecorr.	gecorrig.	N
.00	3.1	3.2	153
.25	3.0	2.9	55
.90	2.3	2.3	102

Levene's test Eq. of Error-Var.: F=1.284; df1=2 ; df2=307; p=.278.

Figuur 14 - Boxplots aanvankelijk schrijven van de OBIS naar thuistaal



Tabel 36 Resultaten covariantie-analyse OBIS-aanvankelijk schrijven - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

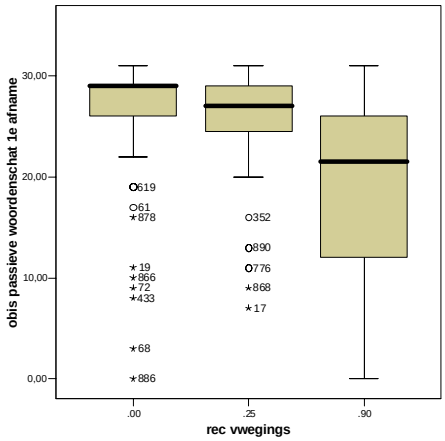
Variantiebron	F	p	partial eta^2
Model	48.648	.000	.312
intercept	48.950	.000	.132
leeftijd	128.949	.000	.286
thuistaal	5.090	.007	.031
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	3.0	3.0	202
dialect/Fries	3.4	3.2	41
vreemde taal	2.4	2.5	83

Levene's test Eq. of Error-Var.: F=.466; df1=2 ; df2=323 ; p=.628.

In figuur 15 en 16 staan de boxplots voor de subtoets passieve woordkennis van de OBIS. Hier zien we een resultaat dat duidelijk afwijkt van dat in de boxplots voor de woordenschattoetsen TvK, Peabody en passieve woordkennis van de TAK. In figuur 15 en nog sterker in figuur 16 zien we dat de bulk van de scores voor .90-leerlingen en ook voor vreemde taalsprekers onder de verdelingen van de andere groepen uitkomen. Deze groepen vertonen een licht plafondeffect. De

.25-leerlingen scoren weer nauwelijks verschillend van de .00-leerlingen. Kijken we naar de resultaten van de covariantie-analyses, dan zien we dat de passieve woordkennis van de OBIS in het differentiëren tussen alle onderscheiden groepen het bijna even goed doet als de TvK (leerlinggewicht en thuistaal verklaren respectievelijk 27 en 29% van de variantie in de scores).

Figuur 15 - Boxplots passieve woordkennis van de OBIS naar leerlinggewicht

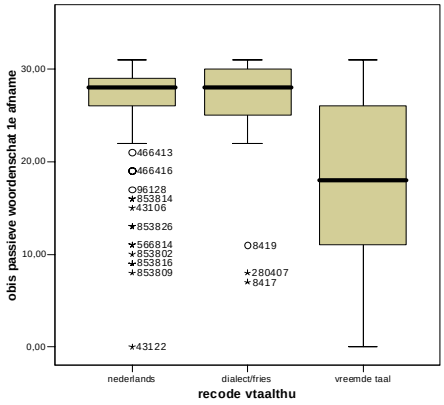


Tabel 37 Resultaten covariantie-analyse OBIS-passieve woordkennis - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	66.888	.000	.396
intercept	.117	.732	.000
leeftijd	87.229	.000	.222
leerlinggewicht	55.376	.000	.266
Gemiddelden	ongecorr.	gecorrig.	N
.00	27.1	27.1	153
.25	25.2	24.9	55
.90	19.5	19.6	102

Levene's test Eq. of Error-Var.: F=19.489; df1=2 ; df2=307; p=.000.

Figuur 16 - Boxplots passieve woordkennis van de OBIS naar thuistaal



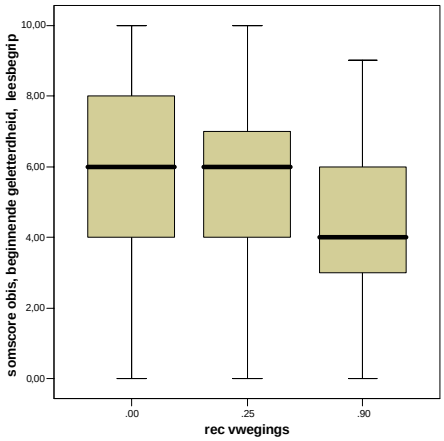
Tabel 38 Resultaten covariantie-analyse OBIS-passieve woordkennis - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	77.638	.000	.420
intercept	.003	.957	.000
leeftijd	87.419	.000	.214
thuistaal	65.736	.000	.290
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	26.8	26.7	202
dialect/Fries	26.4	25.8	41
vreemde taal	18.4	18.7	83

Levene's test Eq. of Error-Var.: F=18.312; df1=2; df2=323; p=.000.

In figuur 17 en 18 staan de boxplots voor de subtoets leesbegrip van de OBIS. Deze subtoets vertoont geen plafond- of bodemeffecten. De boxplots overlappen weer grotendeels. Uit de covariantie-analyses blijkt eveneens dat deze subtoets niet hoog samenhangt met de groepsindelingen. Leerlinggewicht en thuistaal verklaren respectievelijk negenen 14% van de variantie in de scores.

Figuur 17 - Boxplots leesbegrip van de OBIS naar leerlinggewicht

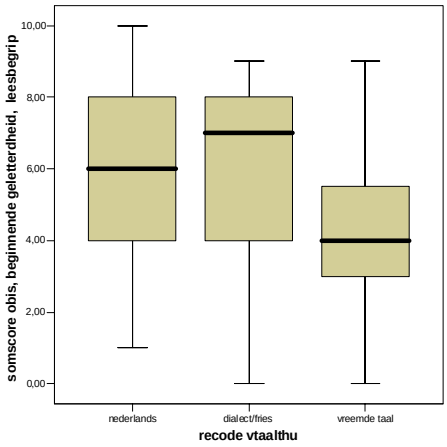


Tabel 39 Resultaten covariantie-analyse OBIS-leesbegrip - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	47.417	.000	.317
intercept	21.107	.000	.065
leeftijd	111.599	.000	.267
leerlinggewicht	14.491	.000	.087
Gemiddelden	ongecorr.	gecorrig.	N
.00	5.8	5.8	153
.25	5.6	5.4	55
.90	4.5	4.5	102

Levene's test Eq. of Error-Var.: F=.016 ; df1=2 ; df2=307; p=.984.

Figuur 18 - Boxplots leesbegrip van de OBIS naar thuistaal



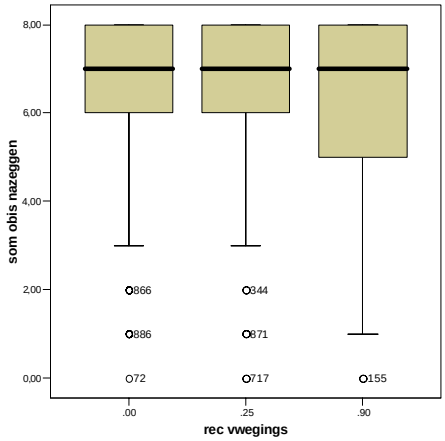
Tabel 40 Resultaten covariantie-analyse OBIS-leesbegrip - Thuistaal als factor, leeftijd als covariaat -Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	60.124	.000	.359
intercept	19.959	.000	.058
leeftijd	117.470	.000	.267
thuistaal	25.344	.000	.136
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	5.8	5.8	202
dialect/Fries	6.2	6.0	41
vreemde taal	4.1	4.2	83

Levene's test Eq. of Error-Var.: F=.073; df1=2; df2=323; p=.930.

In figuur 19 en 20 staan de boxplots voor de subtoets klankarticulatie van de OBIS. Deze subtoets vertoont net als de klankarticulatie van de TAK plafondeffecten. De verdelingen overlappen ook hier grotendeels. De variabelen leerlinggewicht en thuistaal hangen nauwelijks met de scores samen. De percentages verklaarde variantie bedragen voor leerlinggewicht en thuistaal beide slechts 4%.

Figuur 19 - Boxplots klankarticulatie van de OBIS naar leerlinggewicht

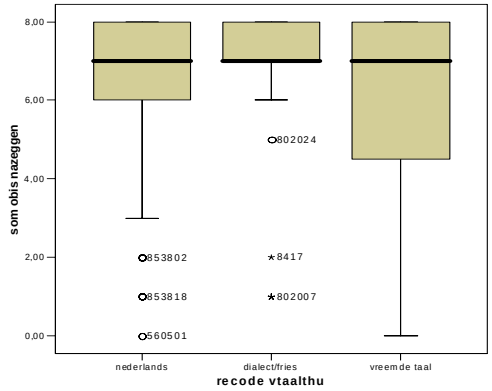


Tabel 41 Resultaten covariantie-analyse KA-OBIS - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	20.506	.000	.167
intercept	.074	.785	.000
leeftijd	50.448	.000	.142
leerlinggewicht	5.774	.003	.036
Gemiddelden	Ongecorr.	gecorrig.	N
.00	6.8	6.8	153
.25	6.2	6.1	55
.90	6.0	6.0	102

Levene's test Eq. of Error-Var.: F=4.086; df1=2 ; df2=307; p=.018.

Figuur 20 - Boxplots klankarticulatie van de OBIS naar thuistaal



Tabel 42 Resultaten covariantie-analyse OBIS-klankarticulatie - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

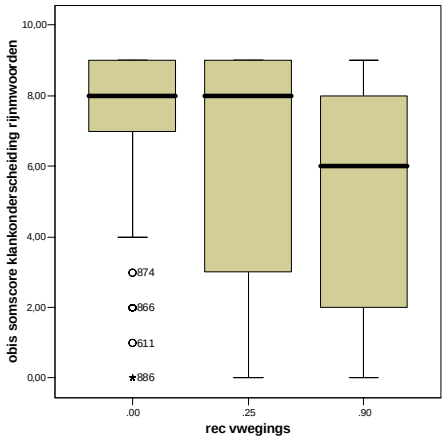
Variantiebron	F	p	partial eta^2
Model	22.103	.000	.171
intercept	.001	.971	.000
leeftijd	50.349	.000	.135
thuistaal	6.268	.002	.037
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	6.7	6.7	202
dialect/Fries	6.8	6.7	41
vreemde taal	5.8	5.8	83

Levene's test Eq. of Error-Var.: F=4.468; df1=2; df2=323 ; p=.012.

In figuur 21 en 22 staan de boxplots voor de subtoets klankonderscheiding van de OBIS. De verdelingen overlappen ook hier, maar wel zien we nu dat de .90-leerlingen en de leerlingen die thuis een vreemde taal spreken meer moeite hebben met deze toets dan de overige leerlingen. Dit is ook te zien aan de percentages verklaarde variantie door groepslidmaatschap (respectievelijk 14 en 12%), die

hoger liggen dan de percentages die we rapporteerden voor de toets klankonderscheiding van de TAK (6 en 11%).

Figuur 21 - Boxplots klankonderscheiding van de OBIS naar leerlinggewicht

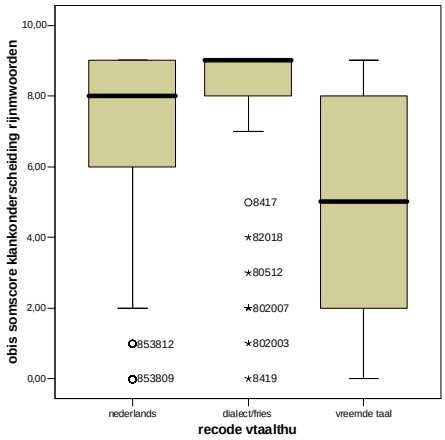


Tabel 43 Resultaten covariantie-analyse OBIS-klankonderscheiding - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	56.891	.000	.358
intercept	34.007	.000	.100
leeftijd	120.298	.000	.282
leerlinggewicht	25.091	.000	.141
Gemiddelden	ongecorr.	gecorrig.	N
.00	7.3	7.3	153
.25	6.2	6.0	55
.90	5.0	5.1	102

Levene's test Eq. of Error-Var.: F=8.141; df1=2; df2=307; p=.000.

Figuur 22 - Boxplots klankonderscheiding van de OBIS naar thuistaal



Tabel 44 Resultaten covariantie-analyse OBIS-klankonderscheiding - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

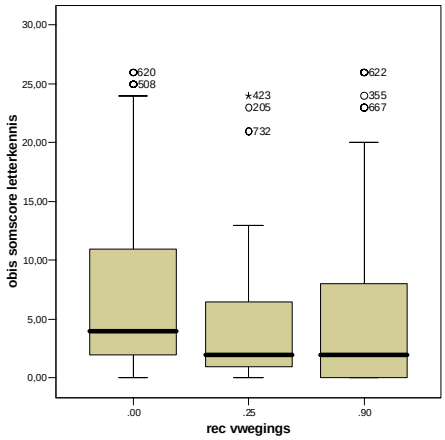
Variantiebron	F	p	partial eta^2
Model	58.003	.000	.351
intercept	28.205	.000	.081
leeftijd	116.947	.000	.266
thuistaal	22.739	.000	.124
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	7.0	7.0	202
dialect/Fries	7.5	7.2	41
vreemde taal	4.7	4.9	83

Levene's test Eq. of Error-Var.: F=3.831; df1=2; df2=323; p=.023.

In figuur 23 en 24 zien we de boxplots voor de subtoets letterkennis van de OBIS. Deze subtoets vertoont bodemeffecten die uiteraard het ergst zijn bij de .90-leerlingen en de leerlingen die thuis een vreemde taal spreken. Ook zien we aan de boxplots en aan de tabellen 45 en 46 dat deze subtoets niet discrimineert tussen

de verschillende groepen. Er wordt maar 4% verklaard door leerlinggewicht en slechts 2% door thuistaal.

Figuur 23 - Boxplots letterkennis van de OBIS naar leerlinggewicht

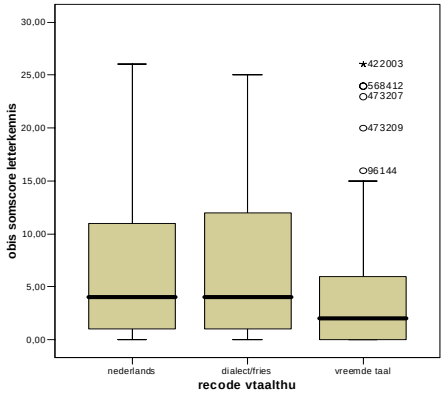


Tabel 45 Resultaten covariantie-analyse OBIS-letterkennis - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	27.738	.000	.214
intercept	43.580	.000	.125
leeftijd	71.517	.000	.189
leerlinggewicht	6.783	.001	.042
Gemiddelden	ongecorr.	gecorrig.	N
.00	7.9	8.0	153
.25	4.7	4.3	55
.90	5.6	5.7	102

Levene's test Eq. of Error-Var.: F=4.527; df1=2 ; df2=307; p=.012.

Figuur 24 - Boxplots letterkennis van de OBIS naar thuistaal



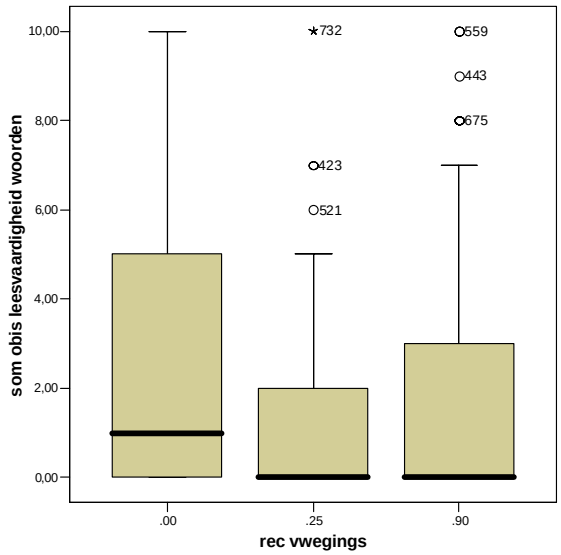
Tabel 46 Resultaten covariantie-analyse OBIS-letterkennis - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	25.545	.000	.192
intercept	38.831	.000	.108
leeftijd	68.605	.000	.176
thuistaal	2.641	.073	.016
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	7.5	7.5	202
dialect/Fries	7.8	7.2	41
vreemde taal	5.0	5.4	83

Levene's test Eq. of Error-Var.: F=3.545; df1=2; df2=323; p=.030.

In figuur 25 en 26 en tabel 47 en 48 staan de resultaten voor de subtoets leesvaardigheid woorden van de OBIS. Weer zien we bodemeffecten en zeer geringe proporties verklaarde variantie door zowel leerlinggewicht als thuistaal. Gegeven dat deze toets binnen de OBIS expliciet bedoeld zijn voor de taalvaardiger leerlingen, is dit geen verrassing.

Figuur 25 - Boxplots leesvaardigheid woorden van de OBIS naar leerlinggewicht

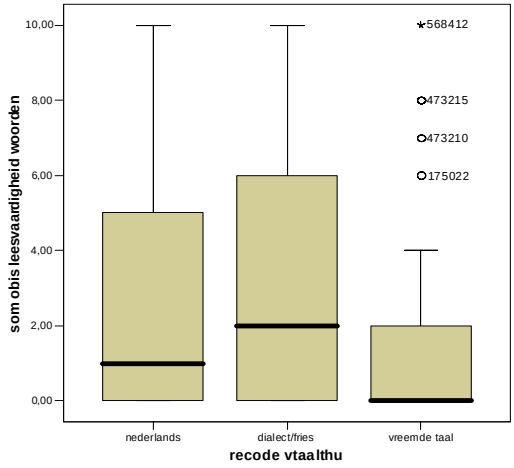


Tabel 47 Resultaten covariantie-analyse OBIS-leesvaardigheid woorden - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	28.271	.000	.217
intercept	49.541	.000	.139
leeftijd	74.256	.000	.195
leerlinggewicht	6.016	.003	.038
Gemiddelden	ongecorr.	gecorrig.	N
.00	2.8	2.8	153
.25	1.7	1.6	55
.90	1.8	1.9	102

Levene's test Eq. of Error-Var.: F=.8.434; df1=2; df2=307; p=.000.

Figuur 26 - Boxplots leesvaardigheid woorden van de OBIS naar thuistaal



Tabel 48 Resultaten covariantie-analyse OBIS-leesvaardigheid woorden - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

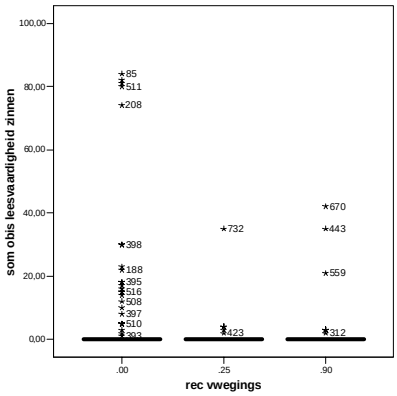
Variantiebron	F	p	partial eta^2
Model	25.442	.000	.192
intercept	42.079	.000	.116
leeftijd	69.445	.000	.177
thuistaal	1.798	.167	.011
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	2.5	2.5	202
dialect/Fries	3.2	2.9	41
vreemde taal	1.8	1.9	83

Levene's test Eq. of Error-Var.: F=3.829; df1=2; df2=323; p=.023.

In de figuren 27 en 28 en tabellen 49 en 50 staan de resultaten voor de subtoets leesvaardigheid zinnen en in figuur 29 en 30 en tabel 51 en 52 staan de resultaten voor de subtoets leesvaardigheid teksten. Zoals al eerder opgemerkt zijn deze subtoetsen bedoeld voor taalvaardiger leerlingen en ze blijken dan ook te moeilijk

voor de jonge leerlingen in onze steekproef. De boxplots laten dan ook extreme bodemeffecten zien en ook de proporties door leerlinggewicht of thuistaal verklaarde variantie zijn laag.

Figuur 27 - Boxplots leesvaardigheid zinnen van de OBIS naar leerlinggewicht

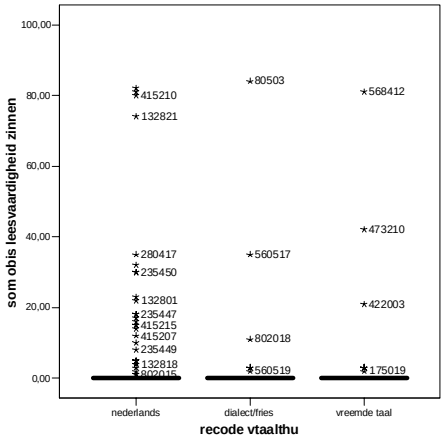


Tabel 49 Resultaten covariantie-analyse OBIS-leesvaardigheid zinnen - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	5.393	.001	.050
intercept	6.668	.010	.021
leeftijd	9.003	.003	.029
leerlinggewicht	3.756	.024	.024
Gemiddelden	ongecorr.	gecorrig.	N
.00	4.4	4.4	153
.25	.9	.8	55
.90	1.1	1.1	102

Levene's test Eq. of Error-Var.: F=9.259; df1=2; df2=307; p=.000.

Figuur 28 - Boxplots Leesvaardigheid zinnen van de OBIS naar thuistaal

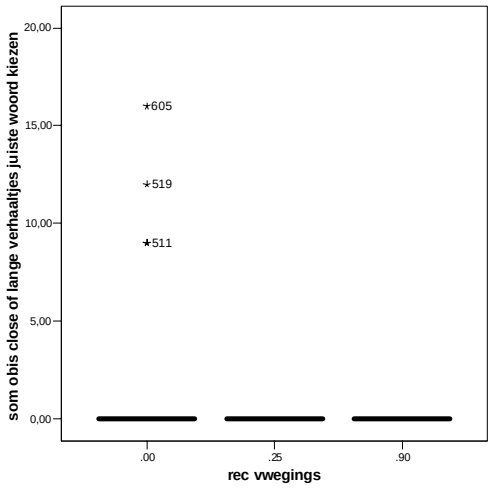


Tabel 50 Resultaten covariantie-analyse OBIS-leesvaardigheid zinnen - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	2.874	.036	.026
intercept	5.073	.025	.016
leeftijd	7.735	.006	.023
thuistaal	.296	.744	.002
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	3.2	3.3	202
dialect/Fries	3.4	3.0	41
vreemde taal	1.9	2.1	83

Levene's test Eq. of Error-Var.: F=.991; df1=2; df2=323; p=.372.

Figuur 29 - Boxplots leesvaardigheid teksten van de OBIS naar leerlinggewicht

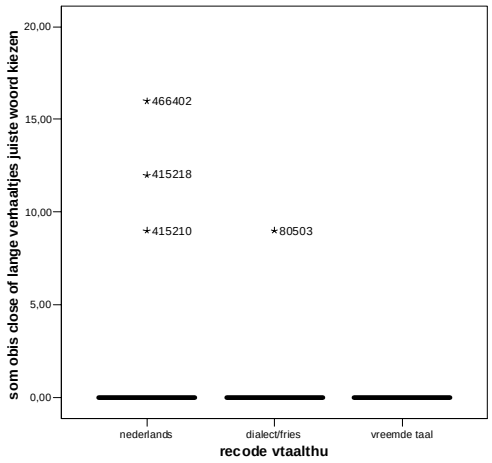


Tabel 51 Resultaten covariantie-analyse OBIS-leesvaardigheid teksten - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	1.708	.165	.016
intercept	.864	.353	.003
leeftijd	1.198	.275	.004
leerlinggewicht	2.002	.137	.013
Gemiddelden	ongecorr.	gecorrig.	N
.00	.3	.3	153
.25	.0	-.0	55
.90	.0	.0	102

Levene's test Eq. of Error-Var.: F=6.247; df1=2 ; df2=307; p=.002.

Figuur 30 - Boxplots leesvaardigheid teksten van de OBIS naar thuistaal



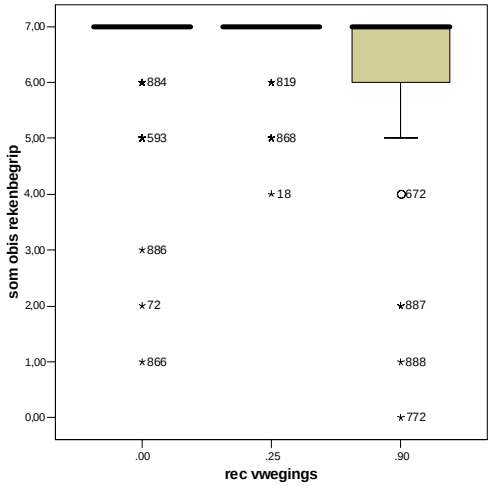
Tabel 52 Resultaten covariantie-analyse OBIS-leesvaardigheid teksten - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	.710	.547	.007
intercept	.456	.500	.001
leeftijd	.810	.369	.003
thuistaal	.584	.558	.004
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	.2	.2	202
dialect/Fries	.2	.2	41
vreemde taal	.0	.0	83

Levene's test Eq. of Error-Var.: F=1.891; df1=2; df2=323; p=.153.

In figuur 31 en 32 en tabel 53 en 54 staan de resultaten voor de subtoets rekenjargon. Helder is dat de subtoets voor deze populatie leerlingen veel te makkelijk is.

Figuur 31 - Boxplots rekenjargon van de OBIS naar leerlinggewicht

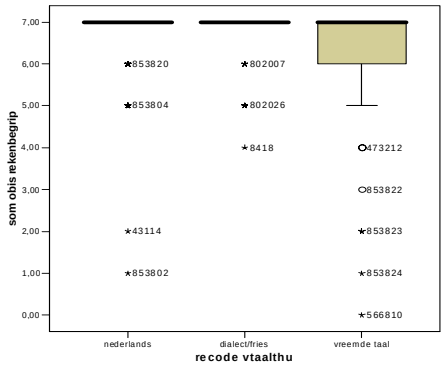


Tabel 53 Resultaten covariantie-analyse OBIS-rekenjargon - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	9.869	.000	.088
intercept	91.598	.000	.230
leeftijd	19.458	.000	.060
leerlinggewicht	4.884	.008	.031
Gemiddelden	ongecorr.	gecorrig.	N
.00	6.7	6.7	153
.25	6.6	6.6	55
.90	6.3	6.3	102

Levene's test Eq. of Error-Var.: F=5.690; df1=2; df2=307; p=.004.

Figuur 32 - Boxplots rekenjargon van de OBIS naar thuistaal



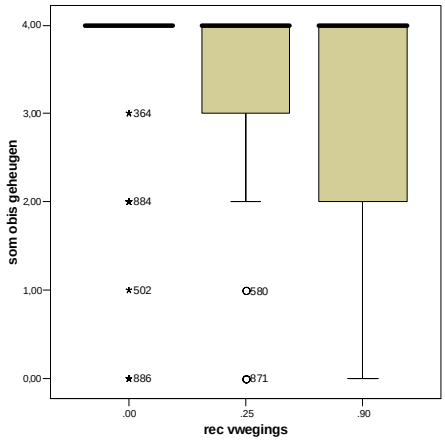
Tabel 54 Resultaten covariantie-analyse OBIS-rekenjargon - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	13.684	.000	.113
intercept	102.419	.000	.241
leeftijd	19.310	.000	.057
thuistaal	9.894	.000	.058
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	6.7	6.7	202
dialect/Fries	6.6	6.6	41
vreemde taal	6.2	6.2	83

Levene's test Eq. of Error-Var.: F=14.084; df1=2; df2=323; p=.000.

Kijken we naar figuur 33 en 34 en tabel 55 en 56, dan zien we dat de subtoets geheugen van de OBIS ook forse plafondeffecten laat zien en dus voor deze leerlingen te makkelijk is, en dat de groepsindelingen naar leerlinggewicht en thuistaal nauwelijks samenhangen met de toetsscores.

Figuur 33 - Boxplots geheugen van de OBIS naar leerlinggewicht

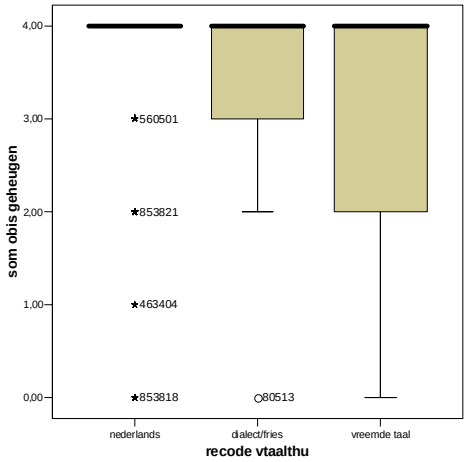


Tabel 55 Resultaten covariantie-analyse OBIS-geheugen - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.).

Variantiebron	F	p	partial eta^2
Model	25.851	.000	.202
intercept	2.810	.095	.009
leeftijd	63.562	.000	.172
leerlinggewicht	6.888	.001	.043
Gemiddelden	ongecorr.	gecorrig.	N
.00	3.5	3.5	153
.25	3.3	3.2	55
.90	3.0	3.0	102

Levene's test Eq. of Error-Var.: F=5.807; df1=2; df2=307; p=.003.

Figuur 34 - Boxplots geheugen van de OBIS naar thuistaal



Tabel 56 Resultaten covariantie-analyse OBIS-geheugen - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

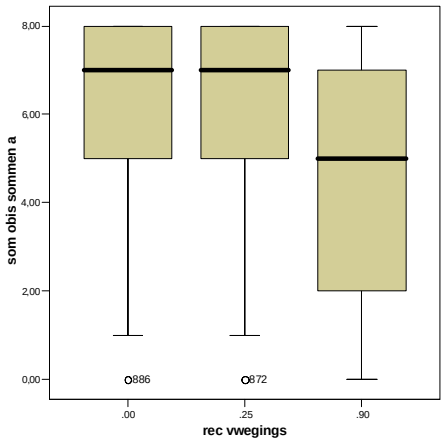
Variantiebron	F	p	partial eta^2
Model	25.308	.000	.191
intercept	2.084	.150	.006
leeftijd	63.182	.000	.164
thuistaal	4.728	.009	.029
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	3.5	3.4	202
dialect/Fries	3.5	3.4	41
vreemde taal	3.0	3.0	83

Levene's test Eq. of Error-Var.: F=4.321; df1=2; df2=323; p=.014 .

In figuur 35 en 36 en tabel 57 en 58 zien we de resultaten voor sommen A van de OBIS. Dit is eigenlijk de eerste rekentoets die we beschouwen. Op voorhand ligt het voor de hand te verwachten dat de groepsindeling naar leerlinggewicht hoger met deze subtoets samenhangt dan de indeling naar thuistaal. Het omgekeerde

blijkt echter het geval. De proportie verklaarde variantie door de groepsindeling naar thuistaal is groter (.17) dan die door de indeling naar leerlinggewicht (.14). Mogelijk speelt voor tweede taalsprekers taalvaardigheid nederlands een rol bij het maken van deze subtoets.

Figuur 35 - Boxplots sommen A van de OBIS naar leerlinggewicht

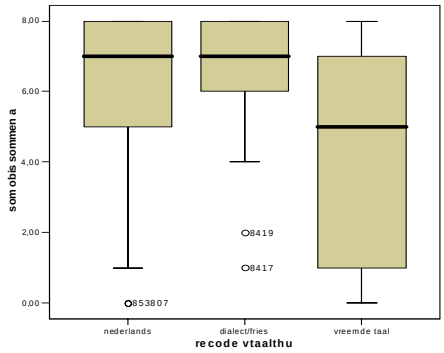


Tabel 57 Resultaten covariantie-analyse OBIS-sommen A - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	80.678	.000	.442
intercept	56.438	.000	.156
leeftijd	190.820	.000	.384
leerlinggewicht	23.813	.000	.135
Gemiddelden	ongecorr.	gecorrig.	N
.00	6.1	6.1	153
.25	5.9	5.7	55
.90	4.4	4.5	102

Levene's test Eq. of Error-Var.: F=5.517; df1=2; df2=307; p=.004.

Figuur 36 - Boxplots sommen A van de OBIS naar thuistaal



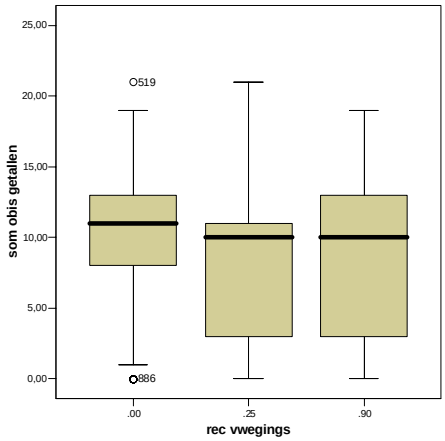
Tabel 58 Resultaten covariantie-analyse OBIS-sommen A - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	93.137	.000	.465
intercept	51.298	.000	.137
leeftijd	194.133	.000	.376
thuistaal	33.349	.000	.172
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	6.0	6.0	202
dialect/Fries	6.6	6.3	41
vreemde taal	4.1	4.2	83

Levene's test Eq. of Error-Var.: F=14.362; df1=2; df2=323; p=.000 .

In figuur 37 en 38 en tabel 59 en 60 staan de resultaten voor de subtoets getallenkennis. Duidelijk is dat de verdelingen over groepen erg op elkaar lijken. De proportie verklaarde variantie door de groepsindelingen is dan ook zeer laag. Opvallend is dat het voor leeftijd gecorrigeerde gemiddelde voor .90-leerlingen hoger is dan voor .25-leerlingen.

Figuur 37 - Boxplots getallenkennis van de OBIS naar leerlinggewicht

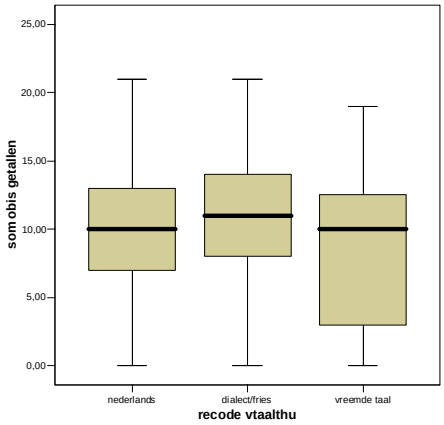


Tabel 59 Resultaten covariantie-analyse OBIS-getallenkennis - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	57.038	.000	.359
intercept	62.544	.000	.170
leeftijd	156.334	.000	.338
leerlinggewicht	8.956	.000	.055
Gemiddelden	ongecorr.	gecorrig.	N
.00	10.0	10.0	153
.25	7.9	7.5	55
.90	8.4	8.5	102

Levene's test Eq. of Error-Var.: F=6.178; df1=2; df2=307; p=.002.

Figuur 38 - Boxplots getallenkennis van de OBIS naar thuistaal



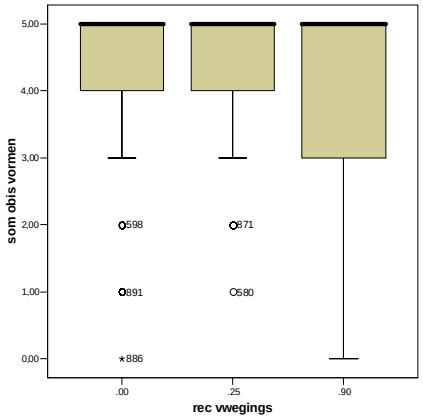
Tabel 60 Resultaten covariantie-analyse OBIS-getallenkennis - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	54.304	.000	.336
intercept	54.173	.000	.144
leeftijd	151.329	.000	.320
thuistaal	2.866	.058	.017
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	9.5	9.5	202
dialect/Fries	10.4	9.8	41
vreemde taal	8.0	8.3	83

Levene's test Eq. of Error-Var.: F=2.160; df1=2; df2=323; p=.117.

In figuur 39 en 40 en tabel 61 en 62 staan de resultaten voor de subtoets vormen. De plafondeffecten geven aan dat deze subtoets te makkelijk is voor de leerlingen. De proporties verklaarde variantie door leerlinggewicht en thuistaal zijn dan ook bijna nihil.

Figuur 39 - Boxplots vormen van de OBIS naar leerlinggewicht

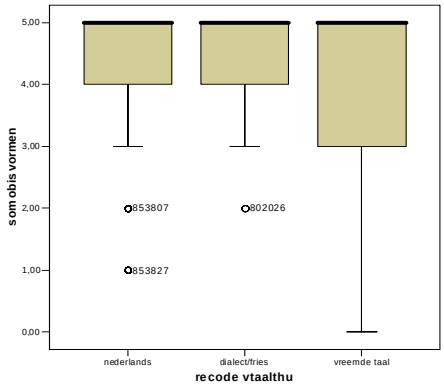


Tabel 61 - Resultaten covariantie-analyse OBIS-vormen -eerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	18.019	.000	.150
intercept	.197	.657	.001
leeftijd	49.997	.000	.140
leerlinggewicht	2.047	.131	.013
Gemiddelden	ongecorr.	gecorrig.	N
.00	4.3	4.3	153
.25	4.2	4.1	55
.90	4.0	4.0	102

Levene's test Eq. of Error-Var.: F=4.541; df1=2; df2=307; p=.011.

Figuur 40 - Boxplots vormen van de OBIS naar thuistaal



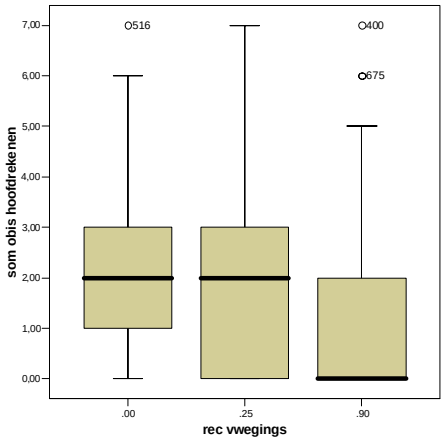
Tabel 62 Resultaten covariantie-analyse OBIS-vormen - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	19.974	.000	.157
intercept	.341	.560	.001
leeftijd	52.824	.000	.141
thuistaal	2.110	.123	.013
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	4.3	4.3	202
dialect/Fries	4.5	4.4	41
vreemde taal	3.9	4.0	83

Levene's test Eq. of Error-Var.: F=8.154; df1=2; df2=323; p=.000.

De resultaten voor de toets hoofdrekenen staan weergegeven in figuur 41 en 42 en tabel 63 en 64. Deze subtoets is moeilijk voor de leerlingen, wat te zien is aan de bodemeffecten.

Figuur 41 - Boxplots hoofdrekenen van de OBIS naar leerlinggewicht

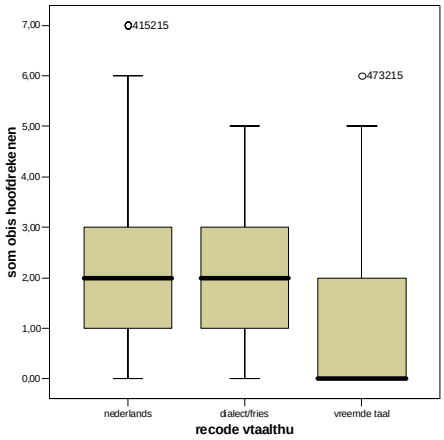


Tabel 63 Resultaten covariantie-analyse OBIS-hoofdrekenen - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	57.380	.000	.360
intercept	86.503	.000	.220
leeftijd	141.380	.000	.316
leerlinggewicht	14.836	.000	.088
Gemiddelden	ongecorr.	gecorrig.	N
.00	2.1	2.2	153
.25	1.8	1.7	55
.90	1.1	1.2	102

Levene's test Eq. of Error-Var.: F=1.331; df1=2; df2=307; p=.266.

Figuur 42 - Boxplots hoofdrekenen van de OBIS naar thuistaal



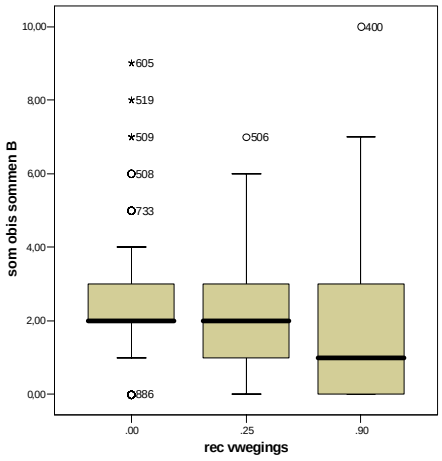
Tabel 64 Resultaten covariantie-analyse OBIS-hoofdrekenen - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	58.608	.000	.353
intercept	82.390	.000	.204
leeftijd	137.495	.000	.299
thuistaal	15.184	.000	.086
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	2.1	2.1	202
dialect/Fries	2.0	1.8	41
vreemde taal	1.0	1.1	83

Levene's test Eq. of Error-Var.: F=5.206; df1=2; df2=323; p=.006.

De laatste subtoets betreft sommen B. ook hier zien we bodemeffecten en lage proporties verklaarde variantie.

Figuur 43 - Boxplots sommen B van de OBIS naar leerlinggewicht

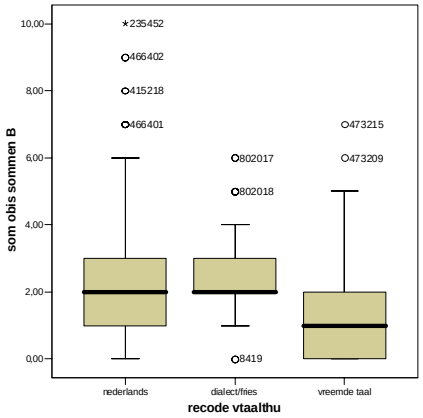


Tabel 65 Resultaten covariantie-analyse OBIS-sommen B - Leerlinggewicht als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	38.307	.000	.273
intercept	45.583	.000	.130
leeftijd	90.786	.000	.229
leerlinggewicht	12.394	.000	.075
Gemiddelden	ongecorr.	gecorrig.	N
.00	2.5	2.6	153
.25	1.9	1.8	55
.90	1.6	1.6	102

Levene's test Eq. of Error-Var.: F=1.756; df1=2; df2=307; p=.174.

Figuur 44 - Boxplots sommen B van de OBIS naar thuistaal



Tabel 66 Resultaten covariantie-analyse OBIS-sommen B - Thuistaal als factor, leeftijd als covariaat - Gegeven worden de F-waarde (F), de overschrijdingskans (p), de proportie verklaarde variantie (partial eta^2), de steekproefgrootte per groep (N), de gemiddelde score per groep (ongecorr.) en de gemiddelde score per groep na correctie voor leeftijdsverschillen (gecorr.)

Variantiebron	F	p	partial eta^2
Model	37.781	.000	.260
intercept	37.778	.000	.105
leeftijd	83.729	.000	.206
thuistaal	11.318	.000	.066
Gemiddelden	ongecorr.	gecorrig.	N
Nederlands	2.4	2.4	202
dialect/Fries	2.7	2.5	41
vreemde taal	1.4	1.5	83

Levene's test Eq. of Error-Var.: F=3.021; df1=2; df2=323; p=.050.

In tabel 67 staat een samenvattend overzicht van de proporties verklaarde variantie in de scores van de verschillende subtoetsen door respectievelijk de groepsindeling naar leerlinggewicht en de groepsindeling naar thuistaal. Voor leerlinggewicht zijn, zoals eerder aangegeven, alleen de groepen .00-, .25- en .90-leerlingen onderscheiden en voor thuistaal onderscheiden we de groep die thuis voornamelijk Nederlands spreekt, zij die voornamelijk een dialect dan wel Fries

spreken thuis en als derde groep zij die thuis hoofdzakelijk een vreemde taal spreken. De proporties verklaarde variantie zijn berekend na eerst te hebben gecontroleerd voor leeftijdsverschillen.
Daarnaast wordt in de tabel aangegeven of, afgaand op de boxplots, de (sub)toetsen voldoende spreiding vertonen (+=goed; +/-=matig; - = slecht) en of er plafond- dan wel bodemeffecten zijn (+=ja). Voor het detecteren van taalzwakke leerlingen zijn plafondeffecten, mits gecombineerd met voldoende spreiding, minder erg dan bodemeffecten. In de laatste twee kolommen staan ook nogmaals de homogeniteit (alfa) en de test-hertestbetrouwbaarheid van de toetsen.
Uit tabel 67 blijkt dat de woordkennistoetsen duidelijk superieur zijn als het gaat om samenhang met groepslidmaatschap naar leerlinggewicht en thuistaal. De beste toets blijkt naar dit criterium gemeten de TvK. De OBIS-passieve woordkennis en de TAK-passieve woordkennis doen het ongeveer even goed en de Peabody is hekkesluiter. Verder valt op dat de .00- en de .25-leerlingen nauwelijks verschillen qua taalvaardigheid en hetzelfde zien we bij de leerlingen die thuis Nederlands dan wel een dialect of Fries spreken. De laagste taalvaardigheid vinden we bij de leerlingen die thuis een vreemde taal spreken.
Kijken we naar de spreiding in de toetsscores, dan blijkt dat voor het detecteren van leerlingen in groep één en twee die het Nederlands slecht beheersen of die niet goed kunnen rekenen de meeste (sub)toetsen bruikbaar lijken. Alleen de OBIS-toetsen leesvaardigheid zinnen, leesvaardigheid teksten en rekenjargon spreiden onvoldoende. De scores van de subtoetsen TAK-klankarticulatie, OBIS-geheugen en OBIS-vormen zijn ook wat minder goed gespreid. Verder blijken de OBIS-toetsen letterkennis, leesvaardigheid woorden, leesvaardigheid zinnen en leesvaardigheid teksten ongeschikt voor het detecteren van slecht presterende leerlingen uit groep één en twee omdat ze een bodemeffect vertonen.
Kijken we nogmaals naar de betrouwbaarheid van de toetsen, dan blijken de test-hertests te laag van de Peabody, alle drie de TAK-subtoetsen en de OBIS-subtoetsen leesbegrip, rekenjargon, geheugen en vormen. Van de OBIS-toetsen leesbegrip, rekenjargon en vormen zijn (ook) de alfa's aan de lage kant.

Tabel 67 Percentages verklaarde variantie in de toetsscores door groepslidmaatschap (leerlinggewicht en thuistaal), na controle voor leeftijdsverschillen - Daarnaast oordeel spreiding in data en aanwezigheid van plafond- of bodemeffecten - Ook de al eerder gerapporteerde homogeniteit (alfa) en test-hertestbetrouwbaarheid worden gegeven

(Sub)toets	leer-ling ge-wicht	thuis-taal	sprei-ding	pla-fond	bodem	alfa	test-hertest
TvK	.283	.334	+			.84	.90
Peabody	.221	.259	+			.95	.74
TAK-Klankonderscheiding	.055	.113	+	+		.93	.74
TAK-Klankarticulatie	.039	.033	+/-	+		.86	.65
TAK-Passieve woordkennis	.233	.337	+			.97	.54
OBIS-Aanvankelijk schrijven	.063	.031	+			n.v.t.	.83
OBIS-Passieve woordkennis	.266	.290	+			.95	.91
OBIS-Leesbegrip	.087	.136	+			.73	.70
OBIS-Klankarticulatie	.036	.037	+	+		.83	.82
OBIS-Klankonderscheiding	.141	.124	+	+/-		.92	.91
OBIS-Letterkennis	.042	.016	+		+	.97	.96
OBIS-Leesvaardigheid woorden	.038	.011	+		+	.93	.85
OBIS-Leesvaardigheid zinnen	.024	.002	-		++	.95	.99
OBIS-Leesvaardigheid teksten	.013	.004	-		++	.98	.96
OBIS-Rekenjargon	.031	.058	-	++		.72	.77
OBIS-Geheugen	.043	.029	+/-	+		.84	.77
OBIS-Sommen A	.135	.172	+	+		.86	.87
OBIS-Getallenkennis	.055	.017	+			.93	.95
OBIS-Vormen	.013	.013	+/-	+		.73	.65
OBIS-Hoofdrekenen	.088	.086	+	+		.75	.83
OBIS-Sommen B	.075	.066	+		+	.71	.81

Toetsen die op grond van de gegevens in tabel 67 goed lijken zijn dus de TvK en de OBIS-toets passieve woordkennis. De andere taaltoetsen hangen niet samen met de groepsindelingen naar leerlinggewicht of thuistaal, vertonen bodemeffecten of zijn onvoldoende betrouwbaar.

In tabel 68 worden vergelijkbare gegevens over samenhang tussen toetsscores en leerlingkenmerken gegeven als in tabel 67, maar op een andere wijze gepresen-

teerd. In deze tabel geven we per taaltoets de percentages leerlingen die te vinden zijn in het onderste kwartiel van de scoreverdeling. Zo kunnen we in tabel 68 lezen dat voor de TvK geldt dat 61% van de leerlingen in het onderste kwartiel van de scores .90-leerling is en dat van alle .90-leerlingen in de steekproef die de TvK maakten er 49% in het onderste kwartiel van de scores terechtkomt.
Uit tabel 68 blijkt wat we al eerder zagen. De indeling naar score op een taaltoets en de oude gewichtenindeling komen niet sterk overeen. De taaltoetsen met het grootste deel .90-leerlingen in het onderste kwartiel van de scores zijn de passieve woordkennistoetsen van TAK en OBIS (69%). Voor beide toetsen geldt ook dat ongeveer 50% van de .90-leelringen hoger scoort dan het onderste kwartiel. Kijken we naar de verdeling van de .25-leerlingen, dan zien we bij de OBIS-toets getallenkennis het hoogste percentage .25-leerlingen in het onderste kwartiel (24%). Verder zien we dat ongeveer een derde van de .25-leerlingen bij het laagste kwart van de scores hoort op de klankonderscheidingtoetsen van TAK en OBIS, op de klankarticulatietoets van de TAK en op de OBIS-toets getallenkennis. Kijken we naar de percentages meisjes in het onderste kwartiel van de scores, dan blijken dat er steeds minder dan de helft te zijn, behalve voor de passieve woordkennistoets van de TAK. Daar de meisjes 50% van de steekproef uitmaken, kunnen we dus concluderen dat meisjes iets hoger scoren dan jongens. Kijken we naar het percentage meisjes dat in het onderste kwartiel van de scores zit, dan zien we hetzelfde: een enkele keer is het ongeveer een kwart van de meisjes, maar meestal minder. De percentages vreemde taalsprekers zijn het hoogst in de onderste kwartielen van de vocabulairetoetsen (TvK, Peabody, en passieve woordkennis van TAK en OBIS) en van de vreemde taalsprekers komt meer dan de helft in het onderste kwartiel van deze toetsen terecht. Bij de andere (sub)toetsen is dat aandeel veel lager (29-48%).

Tabel 68 Percentages .90-leerlingen, .25-leerlingen, meisjes en leerlingen die thuis een vreemde taal spreken bij de per toets geselecteerde 25% slechtst presterende (dus minst taalvaardige) leerlingen - Voor de .90-leerlingen geven we het percentage .90-leerlingen in het onderste kwartiel (% van kwartiel) en het percentage van de .90-leerlingen dat bij de onderste 25% terecht komt (% van .90) - De andere percentages worden op gelijke wijze gegeven

Taaltoets	.90-leerlingen		.25-leerlingen		meisjes		vreemde taal-sprekers	
	% van kwar-tiel	% van .90-lln	% van kwar-tiel	% van .25-lln	% van kwar-tiel	% van meisjes	% van kwar-tiel	% van vreemde t.spr.
TvK	61	49	7	10	32	14	58	53
Peabody	59	55	5	11	48	25	57	58
TAK-Klankonderscheiding	48	35	15	26	40	22	46	38
TAK-Klankarticulatie	47	30	21	31	40	20	38	30
TAK-Passieve woord-kennis	69	48	8	14	51	26	63	52
OBIS-Schrijven	49	37	15	22	39	20	37	35
OBIS-Passieve woord-kennis	69	53	14	20	46	23	64	58
OBIS-Leesbegrip	55	36	15	18	43	20	55	43
OBIS-Klankarticulatie	50	31	19	22	41	17	42	34
OBIS-Klankonderscheiding	52	42	22	33	42	22	48	47
OBIS-Letterkennis	49	32	19	24	40	18	38	30
OBIS-Rekenjargon	51	37	16	22	43	20	45	41
OBIS-Geheugen	47	35	16	22	48	24	39	36
OBIS-Sommen A	52	45	15	24	43	24	47	48
OBIS-Getallenkennis	43	35	24	36	45	25	36	36
OBIS-Vormen	39	26	19	24	47	20	35	29
OBIS-Sommen B	59	40	16	20	42	18	55	45

Conclusie

Gegeven de resultaten van de bovenbeschreven analyses kunnen we bij leerlingen van groep één en twee dus woordkennis het best meten met de TvK en de OBIS-passieve woordkennis. De andere twee woordkennistoetsen vertonen een te lage test-hertestbetrouwbaarheid. Gegeven dat deze toetsen dezelfde latente trek bleken te meten, kunnen deze twee toetsen geëquivaleerd worden met behulp van een item-responsmodel. Verder kunnen we de fonologische trek het best meten

met de OBIS-klankonderscheiding en de OBIS-klankarticulatie, al hangen deze taaltoetsen slecht samen met de groepsindelingen en vertonen ze ook plafondeffecten. Van de technisch-lezenfactor blijken alleen getallenkennis en aanvankelijk schrijven voldoende betrouwbaar en geen bodemeffect te vertonen. Van de rekenfactor is alleen sommen A voldoende betrouwbaar. Deze toetsen hangen niet met de groepsindelingen naar leerlinggewicht of thuistaal samen. Of taaltoetsen met de groepsindelingen samenhangen, betekent niet per se dat de toetsen invalide zijn. Het geeft alleen aan dat de door de toets als taalzwak gekenmerkte leerlingen voor een groot deel niet .90-leerlingen zijn en ook niet leerlingen die thuis een vreemde taal spreken, al zien we wel dat de thuistaal hoger samenhangt met de toetsscores dan het leerlinggewicht. De laagste taalvaardigheid vinden we dus bij de leerlingen die thuis geen Nederlands spreken. Dit kan als argument opgevat worden om de ouders Nederlands te leren.
Toetsen die op grond van samenhang met leerlinggewicht of thuistaal, op grond van de betrouwbaarheid en op grond van de spreiding (zie tabel 67) goed lijken, zijn dus de TvK en de OBIS-toets passieve woordkennis. De andere taaltoetsen hangen niet samen met de groepsindelingen naar leerlinggewicht of thuistaal, vertonen bodemeffecten of zijn onvoldoende betrouwbaar. De TvK bleek wel iets minder betrouwbaar dan de norm (.78 i.p.v. .80) bij de leerlingen van vier tot vijf (zie tabel 13).
Zijn we iets minder streng ten aanzien van de betrouwbaarheid (homogeniteit en test-hertest) en ten aanzien van plafond- en bodemeffecten, dan kunnen wellicht ook de Peabody, de TAK-klankonderscheiding en de OBIS-toetsen letterkennis, leesvaardigheid woorden, hoofdrekenen en sommen B meegenomen worden. Uiteraard zijn deze conclusies gericht op het gebruik van de toetsen voor het detecteren van taalzwakke leerlingen in groep één en twee. We vellen hier geen oordeel over de kwaliteit van de toetsen in het algemeen.

Bij deze conclusies past wel een nuancering. Bedacht moet worden dat uit de analyses ter bepaling van de mate van meetinvariantie bleek dat er geen meetinvariantie is tussen leerlingen van een verschillend leerlinggewicht (.00 vs. rest en .90 vs. rest) en tussen leerlingen die thuis Nederlands, een dialect of Fries spreken en leerlingen die thuis een (andere) vreemde taal spreken. En wat wellicht het grootste probleem levert, er was evenmin meetinvariantie tussen de groep leerlingen van vier tot zes jaar oud enerzijds en de groep leerlingen van zes

tot acht jaar oud anderzijds. Het gebrek aan meetinvariantie naar leeftijd blijft ook bij een coulante toetsing (RDR) significant. Overigens is de grens van zes jaar een relatief arbitraire keus. De steekproef is gesplitst op de mediane leeftijd (zes jaar) zodat er in beide leeftijdsgroepen evenveel leerlingen voorkomen. We weten dus dat de taaltoetsen bij leerlingen van vier tot zes iets anders meten dan bij leerlingen van zes tot acht en dat het factormodel bij leerlingen jonger dan zes relatief slecht past. We weten niet op welke leeftijd de toetsen wel meetinvariant worden. Het is mogelijk dat de grens van zes jaar wellicht iets omlaag kan (of, wat ook mogelijk is, iets omhoog moet).
De conclusie over het gebrek aan meetinvariantie luidt dat de onderzochte toetsen in de voornoemde groepen verschillende vaardigheden meten en dus niet bruikbaar zijn om leerlingen uit deze verschillende groepen leerlingen onderling te vergelijken.

De mate van overeenstemming tussen verschillende (sub)toetsen in het oormerken van de slechtst presterende leerlingen

Naast de bovenbeschreven analyses, is nog een reeks analyses verricht om na te gaan in hoeverre de taaltoetsen overeenstemmen als het gaat om het identificeren van de laagst taalvaardige leerlingen. Indien de toetsen hierin verschillen, is duidelijk dat er een zekere willekeur ontstaat gekoppeld aan de keuze voor een specifieke toets. Als bijvoorbeeld twee toetsen die allebei passieve woordkennis meten en allebei ongeveer even betrouwbaar zijn, onvoldoende overeenstemmen in het oormerken van de 25% leerlingen met de kleinste woordenschat Nederlands, dan is onduidelijk welk van beide toetsen de voorkeur verdient.
Per taaltoets is nagegaan welke score het onderste kwartiel van de verdeling scheidt van de bovenste 75%. De scores op de toetsen zijn vervolgens gehercodeerd in een nieuwe dichotome variabele waarbij de score 1 staat voor behoren tot het onderste kwartiel en de score 2 voor behoren bij de bovenste 75% van de scores. Vervolgens zijn kruistabellen gemaakt voor elke combinatie van twee taaltoetsen om na te gaan in hoeverre de toetsen afwijken als het gaat om het oormerken van de 25% slechtst presterende leerlingen (zie Appendix voor kruistabellen).
In tabel 69 staan de kappa's die aangeven in hoeverre verschillende toetsen dezelfde dan wel verschillende leerlingen in het onderste kwartiel van de verdeling hebben.

Uit tabel 69 blijkt dat van de taaltoetsen de Peabody en de OBIS-passieve woordenschat het best overeenstemmen in het oormerken van het laagst presterende kwart van de leerlingen. De kappa voor deze twee toetsen is .77. De Obis-klankonderscheiding en de Peabody scoren de op één na hoogste kappa (.71) voor taaltoetsen. Overigens stemmen de OBIS-sommen A en B ook redelijk overeen met verschillende taaltoetsen, wat te denken geeft over wat we eigenlijk precies meten bij deze jonge leerlingen.
Deze kappa's laten zien dat het veel uitmaakt welke toets wordt gebruikt om de minst taalvaardige leerlingen te oormerken. Als we een minimale betrouwbaarheid van .80 willen, dan geldt dat eigenlijk ook als norm voor de hier gepresenteerde kappa's (opgevat als convergente validiteitscoëfficiënten). Gerekend naar dat criterium voldoet dus geen van deze validiteitcoëfficiënten aan de norm, behalve de toetsen sommen A en sommen B van de OBIS, maar die meten geen taalvaardigheid.
Ter illustratie geven we de kruistabel van de TvK en de OBIS-passieve woordkennis (zie Appendix voor alle overige kruistabellen). Deze beide toetsen meten dezelfde vorm van taalvaardigheid (passieve woordkennis) en bleken het best naar voren te komen uit alle betrouwbaarheidsanalyses.

Kruistabel TVK en OBIS-passieve woordkennis, oormerken 25% minst taalvaardige leerlingen (1=onderste kwartiel; 2=bovenste 75%)

TvK + OBIS-passieve woordkennis (Kappa=.53)		OBIS-passieve woordkennis: Onderste kwartiel	OBIS-passieve woordkennis: Bovenste 75%	Totaal
TvK	Onderste kwartiel	12	4	16
	Bovenste 75%	10	51	61
Total		22	55	77

We zien hier dat van de 77 leerlingen die beide toetsen maakten er volgens de TvK 16 bij de slechtste 25% horen en volgens de OBIS 22 van de 77. Dat deze

aantallen niet voor beide toetsen op precies 25% uitkomen, komt doordat grensscore ter bepaling van het onderste kwartiel is berekend over alle leerlingen die een toets maakten (veel meer dan 77 dus). Dit is echter geen probleem, daar alle leerlingen random zijn toegewezen aan toetspakketten. Als er een landelijke norm gesteld zou worden om te bepalen of een leerling bij de 25% slechtste leerlingen hoort, zullen normen ook landelijk en niet per school of klas bepaald worden. We zien dus dat beide toetsen het over 12 leerlingen eens zijn dat deze taalzwak zijn (bij de over de gehele steekproef heen slechtste 25% horen), en over 14 (10+4) het oneens zijn. Verder stemmen beide toetsen voor 51 leerlingen overeen die worden toegewezen aan de bovenste 75% van de scores. Duidelijk is dat als we een leerlinggebonden budget zouden willen vaststellen, of verwijzing naar de VVE, op grond van een taaltoets waarbij de normen landelijk bepaald worden, er relatief veel leerlingen zullen zijn die met de éne toets wel en met de andere niet in de prijzen vallen. Overigens is dit resultaat nog geflatteerd. We gaan hier immers voorbij aan het feit dat we hier uitgaan van twee toetsen die passieve woordenschat meten. Onduidelijk is echter welk aspect van taalvaardigheid de norm zou moeten bepalen. Overeenstemming tussen toetsen die verschillende aspecten van taalvaardigheid meten, valt uiteraard veel lager uit (zie tabel 69 en de Appendix).

Tabel 69 De mate van overeenstemming tussen verschillende toetsen in het oormerken van de 25% minst taal- of rekenvaardige leerlingen - Voor elke combinatie van twee (sub)toetsen wordt de kappa gegeven die de mate van overeenstemming weergeeft in het oormerken van de 25% slechtst presterende leerlingen

(Sub)toets	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1 TvK	.68	.37	.22	.56	.27	.52	.31	.45	.52	.21	.17	.17	.35	.17	.28	.24
2 Peabody	-	.41	.46	.61	.39	.77	.58	.33	.71	.62	.38	.45	.69	.45	.47	.74
3 TAK-klankond.		-	.37	.49	.27	.22	.26	.34	.34	.20	.13	.22	.56	.42	.54	.36
4 TAK-klankart.			-	.36	.45	.28	.17	.47	.32	.26	.04	.32	.47	.53	.38	.49
5 TAK-pass. woordk.				-	.38	.49	.47	.38	.18	.32	-.06	.33	.40	.40	.23	.49
6 OBIS-aanv. schr.					-	.43	.33	.42	.47	.51	.17	.36	.53	.53	.34	.54
7 OBIS-pass. woordk.						-	.49	.43	.53	.43	.32	.36	.53	.38	.38	.57
8 OBIS-leesbegr.							-	.32	.42	.44	.23	.39	.41	.36	.37	.47

Vervolg tabel 69

(Sub)toets	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
9 OBIS-klankart.								-	.43	.37	.18	.31	.40	.26	.27	.42
10 OBIS-klankond.									-	.50	.21	.41	.51	.36	.44	.55
11 OBIS-letterkennis										-	.22	.38	.51	.54	.36	.52
12 OBIS-rekenjargon											-	.16	.25	.30	.19	.21
13 OBIS-geheugen												-	.49	.42	.35	.49
14 OBIS-som. A													-	.57	.44	.80
15 OBIS-getallenkennis														-	.41	.54
16 OBIS-vormen															-	.41
17 OBIS sommen B																-

Leerkrachtoordelen over de taalvaardigheid van de leerling

Behalve de overeenstemming tussen verschillende toetsen in het onderscheiden van de 25% slechtst presterende leerlingen, is ook gekeken in hoeverre het oordeel van leerkrachten over de spreek- en luistervaardigheid Nederlands van de leerlingen samenhangt met de toetsscores. In tabel 70 staan de correlaties tussen de leerkrachtoordelen (gegeven op de in scholen traditionele tienpuntsschaal) en de toetsscores.

Duidelijk is dat de leerkrachtoordelen redelijk hoog samenhangen met de toetsscores. Met name de woordkennistoetsen vertonen hoge correlaties met de leerkrachtoordelen. Voor alle vier de opgenomen woordkennistoetsen (TvK, Peabody, TAK-passieve woordenschat en OBIS-passieve woordenschat) blijken de correlaties zelfs boven .50 uit te komen. Deze docentoordelen hangen dus even hoog of hoger samen met deze vier toetsscores dan de indelingen naar thuistaal en leerlinggewicht. Immers, de hoogste proportie door deze groepsindelingen verklaarde variantie in de toetsscores bedraagt .33 (TvK * thuistaal, zie tabel 67). De hoogste correlatie in tabel 70 is .59 (TAK-passieve woordenschat * spreekvaardigheid), wat overeenkomt met een proportie verklaarde variantie van .35. Overigens betekent dit ook dat 65% van de toetsscores niet door de docentoordelen verklaard wordt.

In de Appendix is ook een aantal kruistabellen opgenomen waarin de leerkrachtoordelen over de spreek- en luistervaardigheid Nederlands van de leerlingen wordt afgezet tegen de dichotome variabelen die zijn gemaakt per toetsscore om aan te geven of een leerling bij de onderste 25% van de verdeling hoort. Uit deze kruistabellen (zie Appendix) blijkt dat leerlingen die bij de slechtste scorende 25% leerlingen horen volgens de leraren soms nog heel taalvaardig zijn en andersom.

Tabel 70 Correlaties tussen leerkrachtoordelen over spreek- en luistervaardigheid Nederlands van de leerlingen (10-puntsschaal) en de toetsscores (N=steekproefgrootte)

Subtoets (steekproefgrootte)	N	luistervaardigheid	spreekvaardigheid
TvK	318	.512(**)	.553(**)
Peabody	325	.520(**)	.531(**)
TAK-klankonderscheiding	327	.407(**)	.466(**)
TAK-klankarticulatie	327	.483(**)	.537(**)
TAK-passieve woordkennis	327	.518(**)	.590(**)
OBIS-aanv. Schrijven	305	.315(**)	.298(**)
OBIS-passieve woordkennis	305	.558(**)	.581(**)
OBIS-leesbegrip	305	.385(**)	.375(**)
OBIS-klankarticulatie	305	.452(**)	.479(**)
OBIS-klankonderscheiding	305	.421(**)	.416(**)
OBIS-letterkennis	305	.358(**)	.369(**)
OBIS-leesvaardigheid woorden	305	.357(**)	.359(**)
OBIS-leesvaardigheid zinnen	305	.245(**)	.195(**)
OBIS-leesvaardigheid teksten	305	.161(**)	.141(*)
OBIS-rekenjargon	305	.359(**)	.318(**)
OBIS-geheugen	305	.328(**)	.337(**)
OBIS-sommen A	305	.465(**)	.463(**)
OBIS-getallen	305	.426(**)	.398(**)
OBIS-vormen	305	.428(**)	.407(**)
OBIS-hoofdrekenen	305	.326(**)	.321(**)
OBIS-sommen B	305	.382(**)	.325(**)

** Correlatie is significant op 0.01 (2-zijdig).

* Correlatie is significant op 0.05 (2-zijdig).

Zouden we de leerkrachtoordelen als criterium nemen, dan zijn de beste toetsen voor het meten van begrip van het Nederlands in aflopende volgorde van kwaliteit OBIS-passieve woordkennis, Peabody, passieve woordkennis van de TAK en de TvK. Het betreft dus alleen vocabulairetoetsen. Spreekvaardigheid Nederlands, zoals beoordeeld door de leerkracht, hangt ook het hoogst samen met de vier passieve woordkennistoetsen van de TAK, de OBIS, de Peabody en de TvK en met de toets klankarticulatie van de TAK.

Preselectie van leerlingen voorafgaand aan taaltoetsing

Bij het beschrijven van de geplande kwantitatieve analyses (paragraaf 3.1) is aangegeven dat aan de hand van logistische regressie nagegaan zou worden of preselectie van leerlingen op grond van andere leerlingkenmerken dan toetsscores, zinvol lijkt. De logistische regressies die in tabel 72 gerapporteerde worden, zijn verricht met als afhankelijke variabele telkens de dichotome variabele die per toets aangeeft of een leerling ja dan nee behoort bij de 25% laagst scorende leerlingen. Omdat de leeftijd van de leerling uiteraard meegewogen moet worden, leerlingen worden met het stijgen van de leeftijd immers vaardiger, is de leeftijd van de leerling telkens als eerste covariaat ingevoerd. Vervolgens zijn in een tweede blok de overige onafhankelijke variabelen in de logistische regressievergelijking opgenomen. Deze variabelen zijn beide continue variabelen die het oordeel van de leerkracht weergeven over de receptieve en productieve taalvaardigheid Nederlands (verstaan en spreken) van de leerling en de categorische variabelen leerlinggewicht (.00, .25 en .90) thuistaal (Nederlands, dialect of Fries en vreemde taal) en sekse. In de onderstaande tabel staat per op het onderste kwartiel gedichotomiseerde toetsscore het percentage van de variantie in deze variabele dat voorspeld wordt door de predictoren. De laatste kolom (% verklaard door alleen de predictoren na correctie voor leeftijd (2-1)) geeft aan hoe goed de predictoren voorspellen of een leerling bij het minst taalvaardige kwart van de leerlingen hoort.

Tabel 71 Resultaten logistische regressies - Percentages verklaarde variantie (Nagelkerke's R^2) van alleen leeftijd, leeftijd en predictoren samen en predictoren alleen na correctie voor verschillen in prestatie samenhangend met leeftijdsverschillen - Predictoren zijn naast leeftijd het leerlinggewicht (.00; .25; .90), de thuistaal (Nederlands; dialect/Fries; vreemde taal), sekse en het leerkrachtoordeel over de beheersing van het Nederlands (spreken en verstaan apart)

Dichotome toetsscores (onderste kwartiel versus rest)	Steekproefgrootte	% verklaard door leeftijd (1)	% verklaard door leeftijd en predictoren (2)	% verklaard door alleen de predictoren na correctie voor leeftijd (2-1)
TvK	300	37	64	27
Peabody	307	23	64	41
TAK-Klankonderscheiding	290	22	41	19
TAK-Klankarticulatie	290	18	41	23
TAK-Passieve woordkennis	290	12	55	43
OBIS-Schrijven	284	35	47	12
OBIS-Passieve woordkennis	284	18	63	45
OBIS-Leesbegrip	284	26	40	14
OBIS-Klankarticulatie	284	21	44	23
OBIS-Klankonderscheiding	284	33	51	18
OBIS-Letterkennis	284	32	46	14
OBIS-Rekenjargon	284	7	24	17
OBIS-Geheugen	284	23	28	5
OBIS-Sommen A	284	45	63	18
OBIS-Getallenkennis	284	36	52	16
OBIS-Vormen	284	24	40	16
OBIS-Sommen B	284	41	66	25

Uit de tabel blijkt dat de proporties variantie die verklaard worden door de predictoren na correctie voor verschillen veroorzaakt door leeftijd, lopen van 5% (OBIS-geheugen) tot 45% (OBIS-passieve woordkennis). Hoewel 45% verklaarde variantie best veel is, betekent het wel dat indien we de preselectie op grond

van de gebruikte variabelen zo zouden toepassen dat we alleen leerlingen zouden toetsen waarvan we voorspellen dat ze bij de 25% meest taalzwakke leerlingen behoren, er minimaal 55% van de taalzwakke leerlingen onterecht niet getoetst wordt, wat wel erg veel is. Overigens blijken alleen de passieve woordkennistoetsen van de OBIS, TAK en Peabody dergelijke relatief hoge percentages verklaarde variantie te laten zien. De meeste toetsen scoren rond 20% verklaarde variantie.
Natuurlijk kan het percentage fouten verkleind worden door bijvoorbeeld een logistische regressie te verrichten om te voorspellen welke leerling onder de mediaan scoort (dus bij de minst taalvaardige 50% van de leerlingen hoort). Uit de zo geselecteerde groep, waarbij een ongeveer gelijk foutenpercentage verwacht kan worden als bij voorspelling van wie bij het onderste kwartiel hoort, kan dan via taaltoetsen opnieuw de onderste helft bepaald worden. Zo houdt men ook een kwart van de leerlingen over en wellicht is het aantal foute classificaties dan geringer.
In de Appendix zijn kruistabellen opgenomen waarin de op het onderste kwartiel gesplitste toetsscores worden afgezet tegen de oordelen van leerkrachten over de luister- en spreekvaardigheid Nederlands van de leerlingen. Uit de kruistabellen met de dichotome scores van de woordkennistoetsen blijkt dat foute classificaties onvermijdelijk zijn als we de leerkrachtoordelen als preselectievariabele zouden nemen. Leerlingen die zeer taalvaardig zijn volgens de leerkracht scoren relatief vaak bij het onderste kwartiel en leerlingen met scores bij de bovenste 75% zijn volgens de leerkracht soms zeer laag taalvaardig. Het lijkt dus niet goed mogelijk om foute classificaties te voorkomen, zelfs niet bij een zeer voorzichtige preselectie waarbij bijvoorbeeld alleen leerlingen getoetst worden die volgens hun leerkracht zeer laag taalvaardig zijn in het Nederlands.

Op grond van deze resultaten lijkt het af te raden de toetsen bij eventuele afnamen alleen af te nemen bij een selectie van de leerlingen, tenminste, als die selectie geschiedt aan de hand van de variabelen die wij hier gebruikt hebben voor de logistische regressies.

4.8 Verschil in afnamekwaliteit testleiders en leerkrachten

Voor het beantwoorden van de onderzoeksvragen twee en drie is het van belang om te weten of de (sub)toetsen afgenomen moeten worden door testleiders, of dat ze ook afgenomen zouden kunnen worden door leerkrachten. Om dit te onderzoeken zijn de test-hertestdata voor ongeveer 50% door leerkrachten afgenomen en voor 50% door testleiders. Om een eventueel verschil in afnamekwaliteit te verifiëren zijn er verschillende analyses verricht. Per (sub)toets zijn boxplots gemaakt voor de toetsgegevens verzameld door de testleider en de gegevens verzameld door de leerkracht. Daar leerlingen aselect aan condities zijn toegewezen en dus ook aan testleiders of leerkrachten voor het afnemen van de hertestdata, verwachten we als er geen verschil is in afnamekwaliteit gelijke boxplots en gelijke gemiddelden voor testleiders en leerkrachten. Verder is nagegaan of de betrouwbaarheid van de metingen voor testleiders en leerkrachten verschilt. Deze vergelijking is gemaakt voor zowel de homogeniteit van de itemverzameling per (sub)toets als voor de test-hertestcorrelatie. Verder is met behulp van t-toetsen nagegaan of de gemiddelden op de afgenomen toetsen voor testleiders en leerkrachten significant verschillen.
In de Appendix geven we eerst per (sub)toets en zowel voor de eerste afname als voor de hertest, de boxplots, apart voor testleiders en voor leerkrachten (zie Appendix, figuur A45 t/m A86). In tabel 72 staan de homogeniteiten en test-hertestbetrouwbaarheden van testleiders en leerkrachten, evenals de resultaten van de t-toetsen.

Tabel 72 Resultaten van homogeniteitanalyses, t-tests en test-hertest, apart voor leerkrachten en testleiders

(Sub)toets (aantal items)	homogeniteit leerkracht (N)	homogenteit testleider (N)	Verschil in gemiddelde leerkracht-testleider[24]	test-hertest lrkr	test-hertest testl
TvK (31)	.82 (42)	.84 (50)	-.56	.91	.90
TvK-her (31)	.86 (44)	.88 (51)	-.63		
Peabody (108)	.92 (47)	.94 (49)	.95	.73	.75
Peabody-her (108)	.94 (47)	.95 (49)	-.26		
TAK-Klank-onderscheiding (50)	.95 (47)	.95 (47)	.13	.68	.75
TAK-Klank-onderscheiding-her (50)	.96 (47)	.95 (47)	.66		
TAK-Klankarticulatie (45)	.91 (47)	.81 (47)	-.40	.56	.81
TAK-Klank-articulatie-her (45)	.90 (47)	.83 (47)	.68		
TAK-Passieve woordkennis (96)	.96 (96)	.97 (47)	-5.66	.51	.58
TAK-Passieve woordkennis-her (96)	.98 (47)	.98 (47)	-1.66		
OBIS-Schrijven (1)	-	-			
OBIS-Schrijven-her (1)	-	-			
OBIS-Passieve woordkennis (31)	.94 (44)	.97 (45)	1.04	.86	.94
OBIS-Passieve woordkennis-her (31)	.94 (42)	.97 (40)	.21		

[24] Levene's test blijkt alleen significant voor de OBIS-toetsen klankarticulatie en klankonderscheiding op zowel test- als hertestdata. Voor deze vier t-tests rapporteren we de voor ongelijke varianties gecorrigeerde resultaten.

Vervolg tabel 72

(Sub)toets (aantal items)	homogeniteit leerkracht (N)	homogenteit testleider (N)	Verschil in gemiddelde leerkracht-testleider[24]	test-hertest lrkr	test-hertest testl
OBIS-Leesbegrip (10)	.75 (44)	.74 (45)	.87	.77	.60
OBIS-Leesbegrip-her (10)	.83 (42)	.79 (40)	.82		
OBIS-Klankarticulatie (8)	.74 (44)	.86 (45)	.37	.82	.81
OBIS-Klankarticulatie-her (8)	.82 (42)	.92 (40)	.66		
OBIS-Klankonderscheiding (9)	.85 (44)	.94 (45)	1.37*	.81	.95
OBIS-Klankonderscheiding-her (9)	.91 (42)	.96 (40)	1.49*		
OBIS-Letterkennis (27)	.96 (44)	.97 (45)	-.89	.98	.94
OBIS-Letterkennis-her (27)	.96 (42)	.97 (40)	-.21		
OBIS-Leesvaardigheid woorden (10)	.92 (44)	.95 (45)	.33	.82	.88
OBIS-Leesvaardigheid woorden-her (10)	.92 (42)	.96 (40)	.50		
OBIS-Leesvaardigheid zinnen (10)	.95 (44)	.97 (45)	-.40	.99	.98
OBIS-Leesvaardigheid zinnen-her (10)	.94 (42)	.97 (40)	-.91		

* = p<.05; ** = p<.01; *** = p< .001.

[24] Levene's test blijkt alleen significant voor de OBIS-toetsen klankarticulatie en klankonderscheiding op zowel test- als hertestdata. Voor deze vier t-tests rapporteren we de voor ongelijke varianties gecorrigeerde resultaten.

Vervolg tabel 72

(Sub)toets (aantal items)	homogeniteit leerkracht (N)	homogeniteit testleider (N)	Verschil in gemiddelde leerkracht-testleider	t-ht lrkr	t-ht testl
OBIS-Leesvaardigheid teksten (18)	.94 (44)	.97 (45)	-.06	1.00	1.00
OBIS-Leesvaardigheid teksten-her (18)	.98 (42)	.95 (40)	.06		
OBIS-Rekenjargon (7)	.86 (44)	.80 (45)	-.25	.91	.76
OBIS-Rekenjargon-her (7)	.50 (42)	.90 (40)	-.01		
OBIS-Geheugen (4)	.90 (44)	.90 (45)	.32	.68	.83
OBIS-Geheugen-her (4)	.90 (42)	.89 (40)	.33		
OBIS-Sommen A (8)	.84 (44)	.86 (45)	.60	.84	.90
OBIS-Sommen A-her (8)	.87 (42)	.96 (40)	.38		
OBIS-Getallenkennis (21)	.93 (44)	.93 (45)	-.22	.60	.96
OBIS-Getallenkennis-her (21)	.92 (42)	.94 (40)	.38		
OBIS-Vormen (5)	.85 (44)	.76 (45)	-.09	.87	.39
OBIS-Vormen-her (5)	.85 (42)	.72 (40)	.08		
OBIS-Hoofdrekenen (8)	.82 (44)	.75 (45)	.57	.83	.84
OBIS-Hoofdrekenen-her (8)	.78 (42)	.81 (40)	.61		
OBIS-Sommen B (16)	.71 (44)	.78 (45)	.16	.69	.89
OBIS-Sommen B-her (16)	.71 (42)	.84 (40)	.18		

* = p<.05; ** = p<.01; *** = p< .001.

Uit tabel 72 kunnen we lezen dat de gemiddelden op de toetsen die zijn afgenomen door testleiders enerzijds en leerkrachten anderzijds, slechts eenmaal significant verschillen (OBIS-klankonderscheiding). We kunnen dus concluderen dat leerkrachten en testleiders waarschijnlijk even streng zijn bij het afnemen van de toetsen. Leerkrachten lijken dus bijvoorbeeld ook niet hun leerlingen te helpen bij het geven van antwoorden, tenminste, niet meer dan dat testleiders dat doen.
Verder zien we dat de homogeniteiten van de (sub)toetsen afgenomen door leerkrachten dan wel testleiders een grote mate van overeenstemming vertonen. Daarnaast blijkt uit de geringe verschillen die we wel vinden, dat soms de leerkrachten homogener resultaten leveren en soms de testleiders.

Kijken we naar de test-hertestbetrouwbaarheden, apart voor leerkrachten en testleiders, dan zien we wel aanzienlijke verschillen. Over het algemeen zitten

testleiders iets hoger. Dit geldt voor alle drie de TAK-subtoetsen en voor de OBIS-toetsen passieve woordenschat, klankonderscheiding, geheugen, sommen A, getallenkennis en sommen B. Leerkrachten doen het beter op leesbegrip, rekenjargon en vormen. Een eenduidig resultaat dat wijst op betere afnames door toetsleiders dan wel leerkrachten vinden we dus niet.

Kijken we naar de boxplots in figuur 45 tot en met 86 in de Appendix, dan zien we dat testleiders en leerkrachten wel enigszins verschillen. Op de Peabody zien we meer spreiding bij testleiders, maar alleen op de eerste afname. Bij de hertest is dat verschil verdwenen. Bij de TAK zien we voor klankonderscheiding een plafond op de hertest, wat op een leereffect zou kunnen wijzen. Bij klankarticulatie zien we dit verschijnsel voor zowel testleiders als leerkrachten. Bij de OBIS-toetsen zien we voor schrijven dat de hertestscores van testleiders slechter zijn geworden. Bij leesbegrip zien we voor testleiders minder spreiding op de hertest dan bij de eerste afname, bij klankarticulatie zijn de leerkrachtscores hoger en minder gespreid, bij klankonderscheiding zien we voor zowel de eerste afname als de hertest dat de spreiding bij testleiders veel groter is dan bij leerkrachten, bij rekenjargon zien we op de eerste afname bij leerkrachten minder plafondeffect dan bij testleiders, op de hertest is het plafond er voor leerkrachten echter ook. Voor de geheugentoets zien we het omgekeerde; het plafondeffect is veel extremer bij leerkrachten op de eerste afname dan bij testleiders en op de hertest zien we dat het verschil tussen testleiders en leerkrachten minder is geworden, al blijft het zichtbaar. Bij sommen A is er ook een extreem verschil tussen testleiders en leerkrachten. Scores voor de eerste afname verzameld via leerkrachten vertonen een sterker plafondeffect en laten meer spreiding zien dan de scores verzameld via testleiders. Bij de hertestdata is dit verschil verdwenen. Bij de vormentoets zien we alleen op de hertest een extreem plafond voor de scores verzameld door leerkrachten terwijl dat verschil er eerst niet was. Ook hier zou dit kunnen wijzen op een leereffect dat sterker is als leerkrachten de scores verzamelen. Dat dit leereffect optreedt, wisten we al na de t-toetsen die we uitvoerden op de gemiddelden van de (sub)toetsscores van de eerste afname en de hertest (zie tabel 14).

We concluderen dat leerkrachten en testleiders slechts geringe verschillen in afnamekwaliteit laten zien, waarbij niet één van beide groepen consistent beter lijkt

te presteren. De toetsen kunnen dus zonder verlies aan kwaliteit door leerkrachten afgenomen worden.

4.9 Afnamegemak

Aan de 13 testleiders die de toetsen voor het onderzoek hebben afgenomen, is een vragenlijst voorgelegd met vragen over hoe gemakkelijk de (sub)toetsen af te nemen zijn, over of de leerlingen de toetsinstructie begrepen, of er misverstanden rezen tijdens de afnames, of de leerlingen voldoende aandacht konden opbrengen tijdens de afnames, of zij serieus meewerkten en of de antwoorden makkelijk te scoren waren voor de testleider. Alle vragen zijn gesteld apart voor allochtone en autochtone leerlingen en apart voor leerlingen in groep één en twee. De resultaten staan in tabel 73, 74, 75 en 76.

Uit tabel 73 blijkt dat volgens de testleiders de Peabody op alle bevraagde aspecten voor alle onderscheiden groepen leerlingen goed scoort. Het laagste gemiddelde bedraagt 3.8, wat op een positief oordeel wijst. Op twee aspecten na wijzen de gemiddelden zelfs op een oordeel dat ligt tussen mee eens (4) en helemaal mee eens (5). De twee aspecten waar gemiddeld onder vier gescoord wordt, betreffen de vraag of respectievelijk allochtone en autochtone leerlingen in groep één voldoende aandacht op konden brengen tijdens de toetsafname.
Kijken we naar de laagste scores die gegeven zijn (minimum), dan zien we slechts vijf keer een negatief testleideroordeel. Hoewel de oordelen van andere testleiders de negatieve oordelen dus ruimschoots compenseren, zien we in de data dat één testleider misverstanden bij allochtone leerlingen in groep één vermoedde, twee testleiders vonden onvoldoende aandacht bij autochtone leerlingen in groep één, één vond onvoldoende aandacht bij autochtone leerlingen in groep twee, twee vonden onvoldoende aandacht bij allochtone leerlingen in groep één en een vond dit ook voor allochtone leerlingen in groep twee.
We kunnen concluderen dat de afname van de Peabody over het algemeen geen problemen opleverde, noch voor leerlingen in groep één en twee, noch voor allochtone of autochtone leerlingen.
Naast de lijst met gesloten vragen, kregen de toetsleiders een open vraag naar gerezen problemen, mogelijke oplossingen voor geconstateerde problemen en een

verzoek om relevante op- of aanmerkingen te noteren. Als antwoorden voor de Peabody kregen we de onderstaande opmerkingen.

Gemaakte opmerkingen over Peabody:

- Toets duurt niet te lang, is pluspunt. Leerlingen begrijpen direct de bedoeling, is pluspunt. Toets is gemakkelijk af te nemen, is pluspunt. De norm van 9 is hoog. Soms heeft men er per ongeluk door zomaar wat te wijzen slechts 8 fout en dan moet er toch nog een hele sectie afgewerkt worden.
- Geen problemen. Ik merkte wel soms dat er makkelijk van plaatjes gewisseld werd waardoor de vraag toch opnieuw gesteld moest worden. Soms is het ook nodig te zeggen dat ze eerst alle 4 de plaatjes moeten bekijken alvorens een aan te wijzen. Anders slordigheid en gegok. Ik heb zodra ik onoplettendheid bespeurde duidelijk gemaakt: niet zomaar zwabberen met je vinger, goed kijken, precies aanwijzen.
- Geen problemen kinderen vonden het zelf geweldig.
- Geen problemen. Peabody wordt door de kleuters als erg leuk ervaren.
- Peabody is voor de kinderen leuker dan de TAK toets. Ze vinden de plaatjes leuker en houden er daardoor beter de aandacht bij.
- Geen noemenswaardige problemen.
- Erg leuk voor de kinderen. Goed afwisselend waardoor ze geboeid bleven. Een enkeling vond het iets te veel.
- Uitstekende toets, nadat je als testleider enige ervaring hebt opgedaan kan je aan de hand van ervaring op enig moment vaststellen na 2 testen (3) 4 en 5 hoe ver het kind gaat komen.
- De limiet 9 fouten is te groot voor 12 vragen. De kinderen weten het niet in bv set 7 8 en 9. Ze raken gefrustreerd en dan ongemotiveerd! Ze wijzen maar lukraak wat aan! Bepaalde woorden zijn achterhaald, b.v. huwen!!
- Geen problemen bij de eerste afname, alle lln konden doorgaan na set 3. Bij erg moeilijke woorden verblikken en verbloosden ze niet, maar wezen prompt een plaatje aan. De mogelijkheid van WN werd bijna niet gebruikt.
- Geen problemen kinderen vonden het zelf geweldig.
- Waarom geen mogelijk tot WEET NIET!! aan te kruisen ?

Tabel 73 Afnamegemak Peabody - Gegeven worden het aantal testleiders (N), de minimale en maximale score, het gemiddelde (M), de standaarderror van het gemiddelde (S.E.) en de standaarddeviatie (Std.) (1=helemaal niet mee eens, 3=neutraal; 5=helemaal mee eens.)

Peabody	N	min	max	M	S.E.	Std.
gemakkelijk af te nemen bij autochtone leerlingen groep 1	12	4	5	4.58	.149	.515
gemakkelijk af te nemen bij autochtone leerlingen groep 2	12	4	5	4.58	.149	.515
gemakkelijk af te nemen bij allochtone leerlingen groep 1	10	3	5	4.30	.213	.675
gemakkelijk af te nemen bij allochtone leerlingen groep 2	10	3	5	4.30	.213	.675
instructie goed begrepen door autochtone leerlingen groep 1	12	4	5	4.67	.142	.492
instructie goed begrepen door autochtone leerlingen groep 2	12	4	5	4.67	.142	.492
instructie goed begrepen door allochtone leerlingen groep 1	10	4	5	4.40	.163	.516
instructie goed begrepen door allochtone leerlingen groep 2	10	4	5	4.50	.167	.527
geen misverstanden bij autochtone leerlingen groep 1	12	4	5	4.58	.149	.515
geen misverstanden bij autochtone leerlingen groep 2	12	4	5	4.67	.142	.492
geen misverstanden bij allochtone leerlingen groep 1	10	2	5	4.10	.277	.876
geen misverstanden bij allochtone leerlingen groep 2	10	3	5	4.20	.200	.632
voldoende aandacht bij autochtone leerlingen groep 1	12	1	5	3.83	.366	1.267
voldoende aandacht bij autochtone leerlingen groep 2	12	2	5	4.17	.271	.937
voldoende aandacht bij allochtone leerlingen groep 1	10	2	5	3.80	.359	1.135
voldoende aandacht bij allochtone leerlingen groep 2	10	2	5	4.10	.348	1.101
serieus meewerken door autochtone leerlingen groep 1	12	4	5	4.58	.149	.515
serieus meewerken door autochtone leerlingen groep 2	12	4	5	4.58	.149	.515
serieus meewerken door allochtone leerlingen groep 1	10	4	5	4.50	.167	.527
serieus meewerken door allochtone leerlingen groep 2	10	4	5	4.50	.167	.527
makkelijk te scoren bij autochtone leerlingen groep 1	12	4	5	4.83	.112	.389
makkelijk te scoren bij autochtone leerlingen groep 2	12	4	5	4.83	.112	.389
makkelijk te scoren bij allochtone leerlingen groep 1	10	4	5	4.80	.133	.422
makkelijk te scoren bij allochtone leerlingen groep 2	10	4	5	4.80	.133	.422

In tabel 74 staan de resultaten voor de toets Taal voor Kleuters. De gemiddelde scores liggen hier nog hoger dan bij de Peabody. Het laagste gemiddelde bedraagt 4.22, wat wijst op een score tussen 'mee eens' en 'helemaal mee eens'.
Kijken we naar de minimumscores, dan zien we dat er voor zeven aspecten minstens één testleider een negatief oordeel gaf. Eén testleider gaf aan misverstanden bij autochtone leerlingen in groep één tegen te zijn gekomen, ook één testleider zegt dit voor autochtone leerlingen in groep twee, twee testleiders geven aan dat er gebrek aan aandacht voorkwam bij autochtone leerlingen in groep één, één constateert hetzelfde voor allochtone leerlingen in groep één en één vond aandachtsproblemen voor autochtone leerlingen in groep twee. Tenslotte is er één testleider die vindt dat autochtone leerlingen in groep één en ook groep twee niet serieus meededen.
De conclusie luidt dat over het algemeen de TvK probleemloos werd afgenomen bij zowel allochtone als autochtone leerlingen in zowel groep één als groep twee.

Losse opmerkingen die werden toegevoegd naar aanleiding van afnames van de TvK waren:

- De test duurt net niet te lang (m.n. voor groep 1). De test is duidelijk.
- Plaatjes zijn niet allemaal even duidelijk bv 13 14 15 17 19 21 23 28. Ik vond deze plaatjes lastig voor een kind van 4 jaar. Ook al herhaal je de vraag 2 x dan nog luisteren ze selectief. B.v. 29 groeit in het bos heeft een steel en een hoed. Horen alleen maar hoed en strepen kabouter aan!
- Alleen jammer dat veel kinderen deze test eerder dit jaar hadden gehad.
- Geen problemen. Vaak toets herkenning. Wordt vaker afgenomen. Sterke punt is dat de kleuter actief is met het onderstrepen, waardoor de aandacht optimaal is. Verder is de toets erg kort zodat nooit vermoeidheid optreedt.
- Je merkt dat de toets bekend is bij de meeste leerlingen.
- Vraag 31 leverde vaak twijfel op: afwasborstel is dat de 2e of de 4e?
- De vraag is wat de meerwaarde van de test is. Alle door mij geteste leerlingen hebben deze test al eerder gedaan. De instructie is derhalve makkelijk. De herkenning door sommige leerlingen ook.
- De scholen die enkel de TvK toetsen moesten doen begrepen de doelstelling niet goed! 2 x toetsen uit het boekje van jongste kleuters en dat moesten de oudste kleuters invullen.

- Het was een toets voor de jongste kleuters en dit was in groep 2. Bovendien waren deze kinderen bekend met de CITO-toetsen, dus het was voor hen beslist niet moeilijk. Een makkie zeiden sommigen.
- Ik had met deze toets verwacht het gr 1 aan te bieden, maar misschien was dit wel juist de bedoeling. Ik bedoel, dat het voor gr 2 wel erg simpel was.

Tabel 74 Afnamegemak TvK - Gegeven worden het aantal testleiders (N), de minimale en maximale score, het gemiddelde (M), de standaarderror van het gemiddelde (S.E.) en de standaarddeviatie (Std.) (1=helemaal niet mee eens, 3=neutraal; 5=helemaal mee eens.)

Taal voor Kleuters	N	min	max	M	S.E.	Std.
gemakkelijk af te nemen bij autochtone leerlingen groep 1	13	4	5	4.69	.133	.480
gemakkelijk af te nemen bij autochtone leerlingen groep 2	11	4	5	4.82	.122	.405
gemakkelijk af te nemen bij allochtone leerlingen groep 1	9	4	5	4.44	.176	.527
gemakkelijk af te nemen bij allochtone leerlingen groep 2	9	4	5	4.56	.176	.527
instructie goed begrepen door autochtone leerlingen groep 1	13	4	5	4.69	.133	.480
instructie goed begrepen door autochtone leerlingen groep 2	12	4	5	4.75	.131	.452
instructie goed begrepen door allochtone leerlingen groep 1	9	4	5	4.56	.176	.527
instructie goed begrepen door allochtone leerlingen groep 2	9	4	5	4.67	.167	.500
geen misverstanden bij autochtone leerlingen groep 1	13	2	5	4.46	.243	.877
geen misverstanden bij autochtone leerlingen groep 2	12	2	5	4.50	.261	.905
geen misverstanden bij allochtone leerlingen groep 1	9	4	5	4.56	.176	.527
geen misverstanden bij allochtone leerlingen groep 2	9	4	5	4.67	.167	.500
voldoende aandacht bij autochtone leerlingen groep 1	13	2	5	4.31	.308	1.109
voldoende aandacht bij autochtone leerlingen groep 2	12	2	5	4.58	.260	.900
voldoende aandacht bij allochtone leerlingen groep 1	9	2	5	4.22	.364	1.093
voldoende aandacht bij allochtone leerlingen groep 2	9	3	5	4.67	.236	.707
serieus meewerken door autochtone leerlingen groep 1	13	2	5	4.62	.241	.870
serieus meewerken door autochtone leerlingen groep 2	12	2	5	4.58	.260	.900
serieus meewerken door allochtone leerlingen groep 1	9	4	5	4.78	.147	.441
serieus meewerken door allochtone leerlingen groep 2	9	4	5	4.78	.147	.441
makkelijk te scoren bij autochtone leerlingen groep 1	7	4	5	4.71	.184	.488
makkelijk te scoren bij autochtone leerlingen groep 2	6	4	5	4.67	.211	.516
makkelijk te scoren bij allochtone leerlingen groep 1	5	4	5	4.60	.245	.548
makkelijk te scoren bij allochtone leerlingen groep 2	5	4	5	4.60	.245	.548

In tabel 75 staan de resultaten voor de TAK. Uit de tabel blijkt dat de afname van de TAK voor de meeste leerlingen zonder problemen verliep. De enige vragen waarbij het gemiddelde duidt op een negatief oordeel betreft de vragen naar of allochtone leerlingen voldoende aandacht konden opbrengen. Zowel voor allochtone leerlingen in groep één als in groep twee blijkt het gemiddelde oordeel negatief (resp. 2.56 en 2.90).
Kijken we naar de minimumscores, dan zien we voor de TAK relatief veel lage minimumscores. Dertien keer oordeelt minstens één testleider negatief over een geboden item. Het betreft negatieve oordelen over afnamegemak bij allochtone leerlingen in groep één en twee en bij autochtone leerlingen in groep één, het voorkomen van misverstanden en gebrek aan aandacht bij zowel allochtone als autochtone leerlingen in groep één en twee, en het niet serieus meewerken voor allochtone en autochtone leerlingen in groep één.
Ook uit de opmerkingen die testleiders opschreven, blijkt dat de afname van de TAK soms problemen gaf.

Gemaakte opmerkingen over de TAK:

- Toets duurde nogal eens te lang. Concentratie was weg. Woordjes moeilijk te verstaan.
- Het is niet altijd duidelijk of ze begrijpen bij de klankonderscheiding wat anders of hetzelfde is. Aandacht erbij houden is moeilijk.
- Klankonderscheiding is wat moeilijk. Sommige kinderen dachten aan rijmwoorden. Dikwijls 2 x uitleg geven.
- De instructie voor klankherkenning vergt extra tijd en uitleg. Vergt veel van de concentratie. Er moest soms een korte pauze worden genomen voor opnieuw uitleg.
- De toets klankonderscheiding is voor sommige kinderen niet te begrijpen.
- bij de klankonderscheiding speelt dialect een rol. Daar heb ik rekening mee gehouden.
- Voor sommige jonge of allochtone leerlingen was het teveel achter elkaar>>>moeite om de concentratie vast te houden.
- Passieve woordenschat was voor een aantal kinderen iets te langdradig.
- Een prima toets. De instructie hetzelfde niet hetzelfde is bij sommige leerlingen niet over te brengen. Dat heeft wellicht met het niveau te maken.

- 3 verschillende onderdelen was te veel gevraagd van de kinderen. Woordenschat toets in dit geval te lang. Afbreeknorm is hoog!!
- Zeer onrustige sfeer, oude school, geen goede ruimte om rustig te zitten. Zowel de leerkracht als de testleider vonden dat bij deze kinderen de TAK-toets niet achter elkaar afgenomen kan worden. Toets in twee dagen afgenomen. Deze kinderen waren vaak snel afgeleid, ik wijt dit aan bovenstaande omstandigheden.
- Het nadrukkelijk uitspreken, zeker niet te vlug gaan en telkens de aandacht erbij halen. 'luister goed' is het anders of hetzelfde?
- Afname in delen splitsen voor de jongste kleuters.
- Inkorten!! tot bv. 70 plaatjes.

Tabel 75 Afnamegemak TAK - Gegeven worden het aantal testleiders (N), de minimale en maximale score, het gemiddelde (M), de standaarderror van het gemiddelde (S.E.) en de standaarddeviatie (Std.) (1=helemaal niet mee eens, 3=neutraal; 5=helemaal mee eens.)

Taal voor alle Kinderen	N	min	max	M	S.E.	Std.
gemakkelijk af te nemen bij autochtone leerlingen groep 1	12	2	5	4.08	.229	.793
gemakkelijk af te nemen bij autochtone leerlingen groep 2	12	3	5	4.33	.188	.651
gemakkelijk af te nemen bij allochtone leerlingen groep 1	9	2	5	3.67	.289	.866
gemakkelijk af te nemen bij allochtone leerlingen groep 2	10	2	5	3.80	.291	.919
instructie goed begrepen door autochtone leerlingen groep 1	12	3	5	4.08	.149	.515
instructie goed begrepen door autochtone leerlingen groep 2	12	4	5	4.33	.142	.492
instructie goed begrepen door allochtone leerlingen groep 1	9	3	5	3.89	.200	.601
instructie goed begrepen door allochtone leerlingen groep 2	10	4	5	4.20	.133	.422
geen misverstanden bij autochtone leerlingen groep 1	12	2	5	4.00	.213	.739
geen misverstanden bij autochtone leerlingen groep 2	12	2	5	4.17	.241	.835
geen misverstanden bij allochtone leerlingen groep 1	9	2	5	3.78	.324	.972
geen misverstanden bij allochtone leerlingen groep 2	10	2	5	3.90	.277	.876
voldoende aandacht bij autochtone leerlingen groep 1	12	2	5	3.17	.322	1.115
voldoende aandacht bij autochtone leerlingen groep 2	12	2	5	3.58	.260	.900
voldoende aandacht bij allochtone leerlingen groep 1	9	2	5	2.56	.338	1.014
voldoende aandacht bij allochtone leerlingen groep 2	10	2	5	2.90	.314	.994
serieus meewerken door autochtone leerlingen groep 1	12	2	5	4.17	.241	.835
serieus meewerken door autochtone leerlingen groep 2	12	4	5	4.33	.142	.492
serieus meewerken door allochtone leerlingen groep 1	9	2	5	3.89	.309	.928
serieus meewerken door allochtone leerlingen groep 2	10	3	5	4.20	.200	.632
makkelijk te scoren bij autochtone leerlingen groep 1	12	3	5	4.33	.188	.651
makkelijk te scoren bij autochtone leerlingen groep 2	12	3	5	4.33	.188	.651
makkelijk te scoren bij allochtone leerlingen groep 1	9	3	5	4.22	.222	.667
makkelijk te scoren bij allochtone leerlingen groep 2	10	3	5	4.20	.200	.632

In tabel 76 staan de resultaten voor de OBIS. Uit de tabel blijkt dat de afname van de OBIS zonder problemen verliep. Alle gemiddelde scores op de vragen wijzen op een positief tot zeer positief oordeel. Het laagste gemiddelde is 4.25, waarmee de OBIS en de TvK als de gemakkelijkst af te nemen toetsen naar voren komen.
Kijken we naar de minimumscores, dan blijkt de OBIS eveneens zeer goed af te nemen te zijn. De enige negatieve minimumscores betreffen gebrek aan aandacht van autochtone en allochtone leerlingen in groep twee.
Ook uit de opmerkingen die testleiders opschreven, blijkt dat de afname van de OBIS probleemloos verliep, behalve dat soms het afnemen per pc problemen gaf op scholen. Wellicht is dit in de toekomst te ondervangen door testleiders een eigen laptop mee te geven.

Tabel 76 Afnamegemak OBIS - Gegeven worden het aantal testleiders (N), de minimale en maximale score, het gemiddelde (M), de standaarderror van het gemiddelde (S.E.) en de standaarddeviatie (Std.) (1=helemaal niet mee eens, 3=neutraal; 5=helemaal mee eens.)

OBIS	N	min	max	M	S.E.	Std.
gemakkelijk af te nemen bij autochtone leerlingen groep 1	12	4	5	4.67	.142	.492
gemakkelijk af te nemen bij autochtone leerlingen groep 2	12	4	5	4.67	.142	.492
gemakkelijk af te nemen bij allochtone leerlingen groep 1	8	3	5	4.38	.263	.744
gemakkelijk af te nemen bij allochtone leerlingen groep 2	9	3	5	4.44	.242	.726
instructie goed begrepen door autochtone leerlingen groep 1	11	3	5	4.55	.207	.688
instructie goed begrepen door autochtone leerlingen groep 2	12	3	5	4.50	.195	.674
instructie goed begrepen door allochtone leerlingen groep 1	8	3	5	4.25	.250	.707
instructie goed begrepen door allochtone leerlingen groep 2	9	3	5	4.56	.242	.726
geen misverstanden bij autochtone leerlingen groep 1	11	4	5	4.64	.152	.505
geen misverstanden bij autochtone leerlingen groep 2	12	4	5	4.67	.142	.492
geen misverstanden bij allochtone leerlingen groep 1	8	4	5	4.50	.189	.535
geen misverstanden bij allochtone leerlingen groep 2	9	4	5	4.67	.167	.500
voldoende aandacht bij autochtone leerlingen groep 1	11	3	5	4.55	.207	.688
voldoende aandacht bij autochtone leerlingen groep 2	12	2	5	4.42	.260	.900
voldoende aandacht bij allochtone leerlingen groep 1	8	3	5	4.38	.324	.916
voldoende aandacht bij allochtone leerlingen groep 2	9	2	5	4.33	.333	1.000
serieus meewerken door autochtone leerlingen groep 1	11	4	5	4.64	.152	.505
serieus meewerken door autochtone leerlingen groep 2	12	4	5	4.58	.149	.515
serieus meewerken door allochtone leerlingen groep 1	8	4	5	4.63	.183	.518
serieus meewerken door allochtone leerlingen groep 2	9	4	5	4.56	.176	.527
makkelijk te scoren bij autochtone leerlingen groep 1	11	3	5	4.64	.203	.674
makkelijk te scoren bij autochtone leerlingen groep 2	12	4	5	4.75	.131	.452
makkelijk te scoren bij allochtone leerlingen groep 1	8	4	5	4.75	.164	.463
makkelijk te scoren bij allochtone leerlingen groep 2	9	4	5	4.78	.147	.441

Gemaakte opmerkingen over de OBIS:

- Bij sommen B mocht men kladpapier gebruiken. Was niet goed aangegeven waar B begon.Test net lang genoeg.
- PC problemen ook met geluid. Scholen willen pc niet beschikbaar stellen.
- Kinderen vinden het erg leuk. Geen problemen.
- De rijmwoorden waren voor sommige leerlingen van groep 1 moeilijk te begrijpen.
- De kinderen vinden deze toets prachtig en zijn volledig geboeid door de beelden.
- Soms was de geheugentest ineens afgelopen en kon ik dat niet terughalen.
- PC problemen!
- Bij allochtone kinderen kun je niet aangeven hoelang ze in Nederland verblijven.
- Geen problemen. Prima toets.
- Leuke toets voor de kinderen. Komen veel onderdelen aan de orde!
- Prima verlopen! Ik vond alleen dat ze toch nog wel wat woordjes, letters en zinnen moesten benoemen. M.n. waar begint een zin, waar eindigt een zin? En het lezen van zinnen zelfs losse woordjes, was wel heel moeilijk. Zelfs al waren het maar 1 of 2 zinnen. Ook veel sommen kregen ze nog voorgeschoteld, die voor beide groepen onbegrijpelijk waren.
- Leuke toets om af te nemen.

Conclusie

Het afnemen van de toetsen verliep zonder veel problemen. Over het afnamegemak van de TvK, de Peabody en de OBIS zijn de testleiders zeer positief, over de TAK is men ook positief maar iets minder positief dan over de andere drie toetsen. Wel moeten we bedenken dat we slechts oordelen van maar 13 testleiders hebben en dat we deze oordelen dus niet goed kunnen generaliseren.

4.10 Oordelen schoolleiders over toetsen van jonge leerlingen

De schoolleiders van de aan het onderzoek deelnemende scholen is gevraagd naar hun oordeel over een aantal aspecten van het afnemen van taaltoetsen in het basisonderwijs (zie tabel 77).

De overgrote meerderheid van de 40 schoolleiders die hun oordeel hebben gegeven over het afnemen van taaltoetsen in het basisonderwijs, is daar positief over. Alleen worden vraagtekens geplaatst bij de belasting die individuele afname van toetsen in groep 1/2 vormt voor de leerkrachten.

Driekwart vindt het van belang om meteen bij de start van de schoolloopbaan een taaltoets af te nemen om taalachterstand vast te stellen. Bijna iedereen vindt het een goede zaak om het onderwijs aan individuele leerlingen mede te baseren op de resultaten van een taaltoets en om taaltoetsen te gebruiken om het onderwijs te evalueren. Rond acht van de tien schoolleiders vinden dat dit laatste ook geldt voor het evalueren van het onderwijs in groep 1/2. Bijna iedereen vindt het goed als voor deze groepen landelijk uniforme taaltoetsen beschikbaar zijn. Een grote groep schoolleiders vindt ook dat het afnemen van taaltoetsen bij leerlingen in groep 1/2 een belangrijke bijdrage kan leveren aan de kwaliteit van het onderwijs. Slechts twee van de 40 schoolleiders zijn van mening dat het afnemen van taaltoetsen bij leerlingen in groep 1/2 niet nodig is.

Tabel 77 Oordeel van schoolleiders over het afnemen van taaltoetsen in het onderwijs (percentages); 40 respondenten

	helemaal mee oneens	tamelijk mee oneens	niet mee oneens en niet mee eens	tamelijk mee eens	helemaal mee eens
• Het is van belang taalachterstand direct bij de start van de schoolloopbaan vast te stellen met behulp van een taaltoets	8	3	15	45	30
• Het is een goede zaak om het onderwijs aan individuele leerlingen mede te baseren op de resultaten van een taaltoets	3	5	0	55	38
• Het is een goede zaak om taaltoetsen te gebruiken om het onderwijs te evalueren	3	5	3	25	65
• Het afnemen van taaltoetsen bij leerlingen in groep 1/2 kan een belangrijke bijdrage leveren aan de kwaliteit van het onderwijs	8	3	10	43	38
• Het afnemen van taaltoetsen bij leerlingen in groep 1/2 stuit op veel bezwaren bij leerkrachten	18	30	25	23	5
• Ook in groep 1/2 is het van belang het onderwijs te evalueren aan de hand van taaltoetsen	3	3	13	50	33
• Het afnemen van taaltoetsen bij leerlingen in groep 1/2 is niet nodig	58	30	8	0	5
• Het is een goede zaak als er voor groep 1/2 landelijk uniforme taaltoetsen beschikbaar zijn	5	5	5	40	45
• Het individueel afnemen van toetsen in groep 1/2 vormt een grote belasting voor de leerkrachten	5	8	13	36	39

Vervolg tabel 77

	helemaal mee oneens	tamelijk mee oneens	niet mee oneens en niet mee eens	tamelijk mee eens	helemaal mee eens
• Het is goed mogelijk taaltoetsen in te zetten om gelden voor de bestrijding van onderwijsachterstanden te verdelen	5	5	23	45	23
• Goede taaltoetsen zijn van belang bij het nemen van beslissingen over de overgang van leerlingen naar een volgende groep	3	5	20	43	30
• Goede taaltoetsen kunnen een belangrijke informatiebron vormen bij het informeren van ouders over de ontwikkeling van hun kind	3	5	5	45	43

Bijna iedereen is van mening dat goede taaltoetsen een belangrijke informatiebron kunnen vormen bij het informeren van ouders over de ontwikkeling van hun kind. Driekwart vindt dat goede taaltoetsen van belang zijn bij het nemen van beslissingen over de overgang van leerlingen naar een volgende groep. Verder vindt twee derde dat het goed mogelijk is taaltoetsen in te zetten om gelden voor de bestrijding van onderwijsachterstanden te verdelen.

Over de opvattingen van leerkrachten over het afnemen van taaltoetsen, zijn de meningen van de schoolleiders verdeeld. Ruim een kwart denkt dat het afnemen van taaltoetsen bij leerlingen in groep 1/2 op veel bezwaren van leerkrachten stuit, terwijl de helft denkt dat dit niet het geval is. De rest heeft daarover geen duidelijke mening. Als taaltoetsen individueel moeten worden afgenomen in groep 1/2, denkt driekwart van de schoolleiders dat dit een grote belasting voor de leerkrachten vormt.

Naast de scores op de items kunnen we ook kijken naar de attitude van schoolleiders ten aanzien van het toetsen van jonge leerlingen. De alfa van de items bedraagt .89 (zie tabel 78), dus kunnen we de somscore opvatten als een betrouwbare meting van de attitude van de schoolleider ten aanzien van het toetsen van leerlingen op jonge leeftijd.

Tabel 78 Homogeniteitanalyse attitude schoolleiders ten aanzien van toetsen jonge leerlingen

Alfa = .89	Corrected Item-Total Correlation
Het is van belang taalachterstand direct bij de start van de schoolloopbaan vast te stellen met behulp van een taaltoets	.733
Het is een goede zaak om het onderwijs aan individuele leerlingen mede te baseren op de resultaten van een taaltoets	.757
Het is een goede zaak om taaltoetsen te gebruiken om het onderwijs te evalueren	.742
Het afnemen van taaltoetsen bij leerlingen in groep 1/2 kan een belangrijke bijdrage leveren aan de kwaliteit van het onderwijs	.618
Ook in groep 1/2 is het van belang het onderwijs te evalueren aan de hand van taaltoetsen	.756
Het is een goede zaak als er voor groep 1/2 landelijk uniforme taaltoetsen beschikbaar zijn	.718
Het is goed mogelijk taaltoetsen in te zetten om gelden voor de bestrijding van onderwijsachterstanden te verdelen	.457
Goede taaltoetsen zijn van belang bij het nemen van beslissingen over de overgang van leerlingen naar een volgende groep	.761
Goede taaltoetsen kunnen een belangrijke informatiebron vormen bij het informeren van ouders over de ontwikkeling van hun kind	.716
Het afnemen van taaltoetsen bij leerlingen in groep 1/2 stuit op veel bezwaren bij leerkrachten (gespiegeld)	.176
Het afnemen van taaltoetsen bij leerlingen in groep 1/2 is niet nodig (gespiegeld)	.786
Het individueel afnemen van toetsen in groep 1/2 vormt een grote belasting voor de leerkrachten (gespiegeld)	.240

Als we kijken naar de verdeling van de somscores over schoolleiders (zie tabel 79), dan zien we dat slechts 7.7% van de 39 schoolleiders negatief tot neutraal staat tegenover toetsen. Ruim 40% is uitgesproken positief (weer gemiddeld over alle items). Het gemiddelde bedraagt 3.86 (s.e.= .11; std=.70) en wijkt significant van de neutrale waarde drie af. We kunnen dus ook op grond van dit gemiddelde stellen dat de schoolleiders in onze steekproef over het algemeen positief staan tegenover het toetsen van jonge leerlingen.

Tabel 79 Frequentietabel somscores attitude schoolleiders ten aanzien van toetsen jonge leerlingen

Gemiddelde score over alle items		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.25	1	2.5	2.6	2.6
	2.08	1	2.5	2.6	5.1
	3.00	1	2.5	2.6	7.7
	3.25	3	7.5	7.7	15.4
	3.33	2	5.0	5.1	20.5
	3.42	1	2.5	2.6	23.1
	3.58	1	2.5	2.6	25.6
	3.75	6	15.0	15.4	41.0
	3.83	3	7.5	7.7	48.7
	3.92	2	5.0	5.1	53.8
	4.00	1	2.5	2.6	56.4
	4.08	2	5.0	5.1	61.5
	4.17	1	2.5	2.6	64.1
	4.25	2	5.0	5.1	69.2
	4.33	5	12.5	12.8	82.1
	4.42	2	5.0	5.1	87.2
	4.50	1	2.5	2.6	89.7
	4.58	1	2.5	2.6	92.3
	4.67	1	2.5	2.6	94.9
	4.83	1	2.5	2.6	97.4
	5.00	1	2.5	2.6	100.0
	Total	39	97.5	100.0	

4.11 Praktische aspecten van de toetsing van taalvaardigheid

Bij eventuele grootschalige invoering van de afname van taaltoetsen in groep één en twee van het basisonderwijs zijn een aantal praktische aspecten van belang. Door wie moeten de toetsen worden afgenomen? Gebeurt dat klassikaal of individueel? Hoeveel tijd kost de afname? Wie verwerkt de gegevens? Wat zijn de kosten? Wat de kosten betreft, maken we een schatting gebaseerd op afname bij

50 leerlingen in twee onderbouwgroepen. Dit geeft een indicatie van de kosten bij toetsafname op kleine schaal.

TAK

De Taaltoets Alle Kinderen voor groep één tot en met vier bestaat uit een doos met handleiding, vier platenboeken, twee stripverhalen, de observatielijst 'Sociale taalvaardigheid', een leerlingboek en een cd-rom. De toets kan vijf maal worden afgenomen: oktober/november in groep één, september/oktober groep twee, mei/juni groep twee, mei/juni groep drie en mei/juni groep vier. Omdat met de TAK kan worden vastgesteld of leerlingen bepaalde taalonderdelen beheersen, kan op basis van de uitkomsten worden afgeweken van de geadviseerde data. Er wordt gestart met de observatielijst 'Sociale taalvaardigheid'. Op grond van de uitkomst daarvan wordt bepaald welke taak de leerling krijgt. Het is niet noodzakelijk dat alle kinderen alle taken doen. Door de diversiteit aan taken is het volgens de ontwikkelaars van de toets mogelijk een gerichte diagnose te stellen. Alle taken worden mondeling afgenomen. Bij een aantal taken wordt gebruik gemaakt van een platenboek of een stripverhaal, andere taken worden in een gesprek met het kind verwerkt. De taken 'klankonderscheiding', 'passieve woordenschat' en 'zinsbegrip' (1 en 2) kunnen desgewenst digitaal worden afgenomen. In het hier beschreven onderzoek zijn drie onderdelen van de TAK afgenomen: 'passieve woordenschat', 'klankonderscheiding' en 'klankarticulatie'. Dit duurde vijf tot 60 minuten, met een gemiddelde afnameduur van 21 minuten. Na afloop van de afname moeten de gegevens van elk kind worden vastgelegd in een leerlingoverzicht, dat is opgenomen in het leerlingenboek. Het materiaal voor de afname van de TAK op papier (leerkrachtmap en 50 leerlingboeken) bij 50 leerlingen in de onderbouw komt uit op een bedrag van € 278.

Peabody

De Peabody is een receptieve woordenschattest die individueel wordt afgenomen. De test bevat 204 testplaten met vier afbeeldingen per plaat. Het kind kiest de afbeelding die past bij een woord dat mondeling wordt aangeboden. De platen zijn verdeeld in 17 sets. Voor leerlingen in groep één en twee worden de eerste negensets gebruikt. De eigenlijke afname start bij item 1 van set 3. Blijkt deze set te moeilijk, dan moet worden teruggegaan naar set 1. Op basis van het aantal opeenvolgende fouten wordt bepaald op welk moment de toetsafname wordt

afgebroken. Voor de afname is geen tijdslimiet gesteld. In dit onderzoek varieerde de afnameduur van vier tot 55 minuten. Gemiddeld kostte het afnemen van de Peabody 15 minuten. Op het afnameformulier worden de antwoorden aangekruist, evenals een vermelding indien het gegeven antwoord fout is. Na afloop wordt het aantal juist gemaakte items geteld. Afname van de toets bij 50 leerlingen kost € 435 (een set met platenboek en leerlingformulieren en een extra set van 25 leerlingformulieren).

TvK

De TvK omvat een toets voor de jongste en een voor de oudste kleuters. Beide toetsen dienen in januari en in juni te worden afgenomen. Dat kan groepsgewijs, individueel of in kleinere groepjes. Bij iedere opdracht moeten de leerlingen het juiste plaatje aanstrepen. De afname van de volledige toets duurt ongeveer 60 minuten. Dat kan worden verdeeld over twee of drie afnamemomenten. In dit onderzoek is alleen het onderdeel 'passieve woordenschat' afgenomen. Dat duurde twee tot 25 minuten. De gemiddelde afnameduur van dit onderdeel was 12 minuten. Het materiaal voor de afname van de TvK op papier (1 leerkrachtmap met handleiding, 50 platenboeken en 50 leerlingoverzichten) bij 50 leerlingen in de onderbouw komt uit op een bedrag van € 196. Na afname moeten de uitkomsten worden gecorrigeerd en verwerkt. Als men het observatie- en hulpprogramma wil gebruiken, kost dit extra geld en extra tijd voor afname en verwerking. In plaats van het maken van toetsen op papier, kan worden gekozen voor afname met behulp van de computer. Hiervoor dient de school te beschikken over het Cito-LVS.

OBIS

De OBIS kan worden afgenomen bij de start van het kind in groep één, gevolgd door een tweede afname, ongeveer een jaar later, in groep twee. Voor de afname van de OBIS dient een computer beschikbaar te zijn. De afname vindt individueel plaats, onder leiding van een begeleider of leerkracht. Het voordeel van afname door de leerkracht is dat deze dan al tijdens de afname (meer) inzicht krijgt in het niveau van de leerling. Per leerling neemt de afname ongeveer 20 minuten in beslag. In verband met het gebruik van de computer zijn door de testleiders in het onderzoek geen aantekeningen gemaakt over de feitelijke afnameduur per leerling. Voorafgaand aan de afname dienen de namen en een aantal achtergrondgegevens van de leerlingen te worden ingevoerd. Het verwerken van

de resultaten gebeurt extern, zodat dit geen extra tijdsinvestering van de school vraagt. Voor de levering, verwerking van de gegevens en terugrapportage van de OBIS geldt een richtbedrag van € 100,- per locatie, plus € 6,- per leerling. Voor een school die bij 50 leerlingen in de onderbouw de OBIS afneemt, komt dit uit op een bedrag van € 400.

Conclusie

De afname van de vier toetsen levert in de praktijk geen belemmeringen op, zo blijkt uit het onderzoek. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400. Daarnaast is inzet van personeel nodig om de gegevens te verwerken en de resultaten te berekenen. Bij OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Bij de Peabody kost de verwerking van de gegevens weinig tijd, maar is de verkregen informatie ook summier.

De afnameduur varieert. Gemiddeld kan worden uitgegaan van een afnameduur van 15 à 20 minuten. Daarbij moet worden aangetekend dat van de TAK en TvK delen zijn geselecteerd voor dit onderzoek. Afname van de gehele toetsen kost (aanzienlijk) meer tijd. Ook bleek uit het onderzoek dat er een grote spreiding was in de benodigde tijd.

Geschiktheid voor diagnostische doeleinden door leerkrachten

Indien de taaltoetsen geschikt zijn voor diagnostische doeleinden, kan op de uitkomsten een verbeterprogramma voor individuele leerlingen worden gebaseerd, waarbij gericht wordt gewerkt aan de geconstateerde tekorten. De toets moet dan concrete aanwijzingen geven over het benodigde verbeterprogramma.

TAK

De TAK is volgens de handleiding specifiek ontwikkeld als een diagnostische toets voor de mondelinge vaardigheid Nederlands bij kinderen van groep één tot en met vier. Hij is bedoeld om taalleerproblemen op te sporen en nader te diagnosticeren. Er kunnen daarbij twee groepen van leerlingen worden onderscheiden. In de eerste plaats zijn dat de kinderen die het Nederlands als moedertaal verwerven. Binnen deze groep gaat het om kinderen die in algemene zin 'taalzwak' zijn en om kinderen met specifieke taalleerproblemen (bv. op het

terrein van woordenschat en betekenisrelaties). In de tweede plaats richt de toets zich op kinderen die het Nederlands als tweede taal leren. Om die reden is bij de constructie van de toets zoveel mogelijk rekening gehouden met mogelijke culturele bias in toetsopgaven en de begrijpelijkheid van de instructie.
De toets is genormeerd naar het niveau van kinderen die het Nederlands als moedertaal hebben; volgens de auteurs moet deze norm voor alle kinderen het richtsnoer zijn waar in het onderwijs naar toe dient te worden gewerkt. Daarnaast is voor tweede-taalleerders een vergelijking met kinderen met eenzelfde taalachtergrond mogelijk. De toetshandleiding voorziet per subtoets in een interpretatie van de resultaten en op basis daarvan uitgebreide suggesties voor vormen van taalstimulering. Met behulp van het onderdeel Lessuggesties kan gerichte hulp worden gegeven aan kinderen die achterblijven. Ook zijn daarin verwijzingen opgenomen naar bestaande hulpprogramma's.

Peabody
De Peabody is eveneens een diagnostische test, die één domein van de taalontwikkeling meet, namelijk het begrip van gesproken woorden (door de constructeurs ook aangeduid als verbale intelligentie). Doelgroep in het basisonderwijs zijn kinderen bij wie men snel een globale indruk wil krijgen van het begrip van het gesproken Nederlands. Bij de samenstelling van de items is nadrukkelijk rekening gehouden met de mogelijkheid de test in te zetten voor hen die het Nederlands als tweede taal leren. Overigens telde de totale ijkingsteekproef van 4-6-jarigen slechts 300 kinderen; daaronder bevonden zich geen anderstaligen. Gegevens van Turkse en Marokkaanse kinderen zijn wel apart geanalyseerd, maar het betrof in totaal slechts 72 kinderen. De handleiding bevat gegevens op basis waarvan het niveau van het kind kan worden ingeschat; suggesties voor eventuele remediëring of stimulering ontbreken echter.

TvK
De TvK richt zich op de passieve woordenschat en kritisch luisteren van kleuters in groep één en twee. Met deze toets, die deel uitmaakt van het Cito-leerlingvolgsysteem, is het mogelijk in korte tijd na te gaan hoe de kinderen in groep één en twee zich ontwikkelen en te signaleren of ze een voorsprong dan wel achterstand vertonen. De toets vormt ook een hulpmiddel om de kwaliteit van het onderwijs te verbeteren. Op basis van de resultaten worden zowel groepsover-

zichten als individuele overzichten verkregen. Het groepsoverzicht geeft een snel overzicht van de ontwikkeling van alle leerlingen in de groep, waarbij elk kind een vaardigheidsscore krijgt. Deze varieert van score A (behoort tot de beste 25%) tot score E (behoort tot de zwakste 10%). Op basis daarvan wordt een gemiddeld groepsniveau berekend. In het leerlingenoverzicht worden de vaardigheidsscores per afnamemoment ingetekend en met een lijn verbonden. Zo kan de ontwikkeling in de loop van de tijd worden weergegeven en kan tevens een vergelijking worden gemaakt met de normgroep. Indien de school beschikt over het computerprogramma 'Leerlingvolgsysteem', kunnen de toetsresultaten geautomatiseerd worden verwerkt.
Om meer zicht te krijgen op maatregelen die nodig zijn bij laag scorende leerlingen, kan gebruik worden gemaakt van een speciaal analyseformulier. Op dat formulier worden de scores per onderdeel van de toets ingevuld en gesommeerd. Zo kan worden nagegaan op welke onderdelen van de toets het kind een risicoscore heeft behaald. Vervolgens adviseert het Cito het 'Observatie- en hulpprogramma Taalplezier' af te nemen. Met behulp van het observatieprogramma wordt nauwkeuriger vastgesteld wat de aard van de achterstand is. Het hulpprogramma geeft op uitgebreide wijze suggesties voor gerichte hulp om achterstanden weg te werken. Dit programma omvat vier blokken van vier à vijf thema's.

OBIS

Het doel van de OBIS is het opsporen van verschillen tussen goed en zwak presterende leerlingen. OBIS meet de voortgang in cognitieve ontwikkeling in verschillende contexten, in de eerste twee jaren van het basisonderwijs. De inhoud en de toetsscores kunnen worden gerelateerd aan diagnostiek en (latere) leerprestaties en aan onderwijsdoelen en instructie. Dat is zowel mogelijk voor de individuele leerling als voor de groep. De toets is zodanig samengesteld dat de inhoud correspondeert met de belangrijkste onderwijsdoelen voor beginnende geletterdheid. Bij het ontwerpen van de toets zijn de inhoud en de afbeeldingen zorgvuldig geïnspecteerd op taalgebruik, ontwikkelingsniveau en typische sociaaletnische situaties om te voorkomen dat scores het resultaat zijn van de wijze waarop de vaardigheid is gemeten. Het onderdeel 'beginnende geletterdheid' omvat schrijven, woordenschat, leesbegrip, letterkennis en het lezen van woorden. Het rekengedeelte toetst 'beginnende gecijferdheid'. Het bestaat uit de onderdelen

rekenjargon, tellen, sommen en getalsherkenning. De vragen en opdrachten zijn zodanig gerangschikt dat zij oplopen in moeilijkheid en complexiteit. Op basis van afbreekregels worden de leerlingen geen vragen of onderdelen aangeboden die duidelijk boven hun niveau liggen. De inhoud van de OBIS wordt jaarlijks geïnspecteerd op validiteit en betrouwbaarheid, evenals de standaardisering van de scores.
OBIS wordt met behulp van de computer afgenomen. De resultaten worden extern verwerkt. Scholen ontvangen voor de begin- en de vervolgmeting verschillende sets met toetsuitslagen. Alle uitslagen bestaan uit ruwe en gestandaardiseerde scores. De gestandaardiseerde scores zijn gebaseerd op een grote steekproef met een normale verdeling, zodat deze scores zowel met andere groepen als in de tijd kunnen worden vergeleken. De scores zijn ook samengevat in een lijnengrafiek en een namengrafiek, die ieder een ander aspect van de vooruitgang van de leerlingen laat zien. Er zijn aparte grafieken voor taal en voor rekenen. De ontwikkeling van elke leerling is weergegeven met een lijn die loopt van links, de score op de begintoets, naar rechts, de score op de eindtoets. In de grafiek geeft de positie ten opzichte van de middelste lijn aan of een kind zich naar verwachting ontwikkelt of sneller of langzamer. Dit kan een signaal geven dat er extra zorg en aandacht gewenst is.

Conclusie

De derde onderzoeksvraag betreft de geschiktheid van de vier taaltoetsen bij gebruik voor diagnostische doeleinden door leerkrachten. De vier toetsen hebben als doel inzicht te geven in de mate van achterstand die leerlingen hebben op taalgebied. Bij TAK en OBIS kan gedetailleerde informatie op subtoetsniveau worden verkregen. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht, het observatie- en hulpprogramma. Bij de Peabody is het inzicht in de mate van taalontwikkeling beperkt en ontbreken suggesties voor verbetering. Deze toets biedt dus de minste mogelijkheden op diagnostisch gebied.

5 Eindconclusies

Als we de uitkomsten van het onderzoek bezien, zijn een aantal conclusies te trekken.

De voor het onderzoek geselecteerde toetsen en subtoetsen blijken in de totale steekproef van leerlingen van vier tot acht jaar oud zonder uitzondering intern consistent ofwel homogeen te zijn. Bij de leerlingen jonger dan zes jaar blijken de meeste toetsen ook voldoende homogeen. De items behorend bij één (sub)toets meten dus in ruime mate dezelfde trek ofwel hebben genoeg gemeenschappelijke variantie.

De test-hertest betrouwbaarheid is voor veel toetsen goed, maar voor enkele toetsen laag. De waarden variëren van .54 (TAK passieve woordenschat) tot .99 (OBIS-leesvaardigheid zinnen). Wanneer gebruik van de toetsresultaten consequenties heeft voor het individu, wordt algemeen een ondergrens voor de betrouwbaarheid van .80 gehanteerd (de COTAN-norm). Voor het nemen van beslissingen op groepsniveau, zoals bij toewijzing van achterstandsgelden aan scholen of bij het doen van onderzoek dat gericht is op het zoeken naar verbanden op groepsniveau, is een lagere betrouwbaarheid niet direct een probleem, mits de betrouwbaarheid van de oordelen op groepsniveau voldoende is.
Test-hertestbetrouwbaarheden onder .80 vinden we bij acht van de 21 (sub)toetsen, te weten de Peabody, de TAK-subtoetsen klankarticulatie, klankonderscheiding en receptieve woordenschat en de OBIS-subtoetsen leesbegrip, rekenjargon, geheugen en vormen. Voor de Peabody en de drie TAK-subtoetsen zijn deze waarden een probleem. Voor de OBIS zijn de lage test-hertests minder bezwaarlijk, daar de OBIS zeer veel subtoetsen kent en de subtoetsen met een lage test-hertest veelal zeer weinig items hebben. Daarnaast blijkt uit de resultaten van factoranalyses, dat de OBIS-subtoetsen met een lage test-hertest in onze steekproef hetzelfde meten als andere, wel betrouwbare delen van de OBIS.
Opvallend is dat voor de OBIS de test-hertest voor leerlingen jonger dan zes jaar beter is dan die voor de totale steekproef. Naast de al eerder genoemde toetsen met een test-hertest groter dan .80, zien we dat nu ook de OBIS-subtoetsen klankonderscheiding, leesvaardigheid woorden, leesvaardigheid zinnen, geheugen,

sommen A, vormen, hoofdrekenen en sommen B een voldoende test-hertestbetrouwbaarheid vertonen.
Voor het doen van uitspraken over groepen leerlingen kunnen we een minder strenge norm hanteren. Als we uitgaan van een norm van .70, dan blijken alleen de TAK subtoetsen klankarticulatie en passieve woordenschat en de OBIS-subtoets vormen een lagere test-hertest te hebben. De vormentoets bevat echter maar vijf items en blijkt (zie hieronder) passieve woordenschat te meten. Daar de OBIS subtoets passieve woordenschat zeer betrouwbaar blijkt, is de lagere betrouwbaarheid van de vormentoets geen probleem. Voor de beide TAK subtoetsen is de test-hertest echter te laag.

Naast de bepaling van de betrouwbaarheid over de gehele scorerange van de toetsen, is ook gekeken in hoeverre toetsen stabiel zijn in het identificeren van de 25% minst vaardige leerlingen. Voor deze test-hertestbetrouwbaarheid in het onderscheiden van de 25% minst taalvaardige leerlingen, blijken alleen de TvK en de OBIS-toetsen passieve woordenschat, klankarticulatie, kennis van letters en kennis van getallen op of boven een kappa van .80 uit te komen. De OBIS-toetsen klankonderscheiding, aanvankelijk schrijven en sommen B hebben kappa's van bijna .80. De andere toetsen blijken dus niet erg stabiel in het onderscheiden van de 25% slechtst presterende leerlingen.

Tot slot is als betrouwbaarheidsbepaling ook nog nagegaan in hoeverre *verschillende* (sub)toetsen overeenstemmen bij het onderscheiden van de 25% zwakst presterende leerlingen. Het blijkt dat de Peabody en de OBIS-passieve woordenschat het best overeenstemmen in het oormerken van het laagst presterende kwart van de leerlingen. De kappa voor deze twee toetsen is .77. De OBIS-klankonderscheiding en de Peabody scoren de op één na hoogste kappa (.71) voor taaltoetsen. Deze kappa's zijn lager dan de betrouwbaarheidsnorm die de COTAN voor individuele oordelen stelt (.80). Duidelijk is dus dat het veel uitmaakt welke toets gebruikt wordt om de taalzwakke leerlingen te identificeren. Overigens stemmen de OBIS-sommen A en B ook redelijk overeen met verschillende taaltoetsen. Dit gegeven zou erop kunnen wijzen dat de verschillende vaardigheden die we meten bij deze jonge en zwak presterende leerlingen nog niet erg gedifferentieerd zijn ofwel op jonge leeftijd bij de zwakste leerlingen onderling hoger samenhangen dan bij oudere of vaardiger leerlingen.

Vervolgens is gekeken naar de validiteit van de verschillende toetsen. Principale componentenanalyses maken duidelijk dat alle (sub)toetsen die bedoeld zijn om één construct te meten, dat ook min of meer doen. De (sub)toetsen blijken dus alle voldoende unidimensioneel. Uit exploratieve en toetsende factoranalyses op somscores per (sub)toets blijkt vervolgens dat we alle (sub)toetsen kunnen zien als metingen van grofweg vier constructen: receptieve of passieve woordkennis, fonologische vaardigheid, technisch leesvaardigheid en rekenvaardigheid. De receptieve woordkennis wordt gemeten door de (sub)toetsen TvK, Peabody, TAK-passieve woordenschat en de OBIS-toetsen passieve woordenschat, rekenjargon en vormen. De fonologische factor wordt gemeten door de toetsen klankonderscheiding en klankarticulatie van zowel de TAK als de OBIS. De technisch-lezenfactor wordt gemeten door de OBIS-toetsen aanvankelijk schrijven, leesbegrip, letterkennis, leesvaardigheid woorden en getallenkennis. De OBIS-toetsen geheugen, sommen A, hoofdrekenen en sommen B vormen de rekenfactor. Daar deze structuur ook inhoudelijk goed te verdedigen is, kunnen we deze resultaten zien als een ondersteuning van de validiteit van de verschillende (sub)toetsen. Ook blijkt uit deze resultaten dat de scores op verschillende toetsen behorend tot één en dezelfde factor, in elkaar omgezet kunnen worden, ofwel dat equivalering met behulp van item-responsmodellen goed mogelijk is.

Een andere belangrijke vraag is of de toetsen die zijn afgenomen voor het onderhavige onderzoek meetinvariant zijn (ofwel geen partijdigheid of bias vertonen) voor specifieke subgroepen in de steekproef. Zonder meetinvariantie kunnen de scores op de toetsen in de onderscheiden groepen niet gezien worden als gelijkwaardige metingen van dezelfde vaardigheid. Er zijn verschillende vormen van meetinvariantie. De strengste vorm (strict factorial invariance) impliceert dat de toetsen binnen de onderscheiden groepen dezelfde vaardigheid meten op dezelfde schaal en met een gelijke betrouwbaarheid. Deze vorm van meetinvariantie hebben we niet getoetst, omdat gelijke betrouwbaarheid in alle onderscheiden groepen wel een erg strenge eis is. Een iets minder strenge vorm van meetinvariantie impliceert het meten van dezelfde trek op dezelfde schaal binnen alle onderscheiden groepen. Deze vorm van meetinvariantie is wel gecontroleerd. De derde vorm van meetinvariantie betreft alleen het meten van dezelfde trek. Als alleen deze laatste vorm van meetinvariantie houdt, is wel hetzelfde gemeten in de onderscheiden groepen, maar kunnen de gevonden scores alleen binnen en niet

over groepen vergeleken worden. Als er helemaal geen meetinvariantie is, dan meten de toetsen in de onderscheiden groepen met verschillende betrouwbaarheid op een verschillende schaal en ook nog eens niet dezelfde trek. In de wetenschappelijke literatuur over dit onderwerp wordt bij het toetsen van individuen waarbij beslissingen worden genomen die consequenties hebben voor het individu, geëist dat de strengste vorm van meetinvariantie (inclusief gelijke betrouwbaarheid) houdt (Meredith, 1993).

Er is gekeken naar meetinvariantie voor de groepen die worden gevormd door de variabelen sekse, leerlinggewicht, thuistaal en leeftijd. Meetinvariantie is een noodzakelijke voorwaarde om te mogen stellen dat de toetsen onpartijdig zijn ten aanzien van verschillende (groepen) leerlingen. Gebrek aan meetinvariantie wordt dan ook wel aangeduid als 'partijdigheid' of 'bias'. Voor het bepalen van de mate van meetinvariantie hebben we ons beperkt tot de drie taalfactoren en alleen (sub)toetsen meegenomen die goed op hun factor laden. Voor de vocabulairefactor zijn dat de TvK, de Peabody, TAK-passieve woordenschat en OBIS-passieve woordenschat. Voor de fonologische factor zijn dat de toetsen klankonderscheiding en klankarticulatie van zowel de TAK als de OBIS en voor de technisch-lezenfactor zijn dat de drie OBIS-subtoetsen kennis van letters, leesvaardigheid woorden en kennis van getallen.
De meetinvariantie is getoetst voor het gehele driefactormodel ineens. Gegeven de zeer goede fit van dit model en de hoge factorladingen van elk van de opgenomen toetsen op de factor waarop ze moeten laden, is deze keus goed te verdedigen. De meetinvariantie van de toetsen is in stappen geverifieerd. Eerst is nagegaan of de toetsen op dezelfde schaal meten in de onderscheiden groepen. Als dit niet het geval blijkt, krijgen leerlingen met een identiek vaardigheidsniveau op de gemeten trek een verschillende score. Evenzo zijn leerlingen uit verschillende groepen die een gelijke score hebben dan niet even taalvaardig. Dit soort partijdigheid is eventueel nog te corrigeren, bijvoorbeeld door de betreffende toets per groep te normeren.
Vervolgens is nagegaan of in de onderscheiden groepen wel hetzelfde construct ofwel dezelfde vaardigheid is gemeten. Het mag duidelijk zijn dat als blijkt dat in onderscheiden groepen verschillende vaardigheden gemeten worden, de toetsen niet op zinvolle wijze over deze groepen heen gebruikt kunnen worden. Men meet

dan immers in de onderscheiden groepen met dezelfde toets verschillende vaardigheden.

Uit de meergroepenanalyses blijkt dat er wel meetinvariantie is over sekse, maar niet over leerlingen die thuis een vreemde taal spreken versus zij die dat niet doen, niet over de groep .00-leerlingen[25] versus de overige leerlingen, niet over de groep .90-leerlingen versus de overige leerlingen en niet over leerlingen van vier tot zes jaar oud ten opzichte van leerlingen van zes tot acht jaar oud. Met name de meetinvariantie over de verschillende leeftijdsgroepen blijkt flink geschonden. Het gehanteerde factormodel blijkt overigens in de groep ouder dan zes jaar zeer goed te passen en in de groep jonger dan zes jaar vrij slecht. Dit doet vermoeden dat de toetsen bij leerlingen onder de zes jaar niet goed functioneren en boven de zes jaar wel.
Om een indruk te krijgen van de ernst van de partijdigheid, zijn de onterechte verschillen over groepen per toets uitgedrukt in standaarddeviaties. Dit is alleen gedaan indien werd gevonden dat de toetsen op verschillende schalen meten. De schending van meetinvariantie blijkt, afhankelijk van de toets, onterechte verschillen van krap een derde tot ruim een hele standaarddeviatie te bedragen. Deze effecten zijn zeer fors. Bij het berekenen van deze effectmaat is nog buiten beschouwing gelaten dat de toetsen in de onderscheiden groepen ook verschillende constructen blijken te meten. Uiteraard impliceert dit dat de toetsen niet voor de verschillende groepen gebruikt kunnen worden om vervolgens de scores over groepen heen te vergelijken. Men vergelijkt dan appels en peren. Het afnemen van de toetsen bij leerlingen ouder dan zes jaar lost mogelijk het probleem van gebrek aan meetinvariantie naar leeftijd op, de meetinvariantie naar thuistaal of sociaal economisch milieu echter niet.
Ook is nagegaan of de gevonden partijdigheid van de toetsen veroorzaakt wordt door verschil in meetmethoden van de gebruikte toetsen. De OBIS wordt immers afgenomen met een p.c. en de andere drie toetsen via toetsleiders en plaatjes op papier. Een methodefactor blijkt echter niet de verklaring te zijn van het gevonden gebrek aan meetinvariantie.

[25] In het onderzoek worden nog de oude leerlinggewichten gehanteerd.

Het gevonden gebrek aan meetinvariantie betekent dat de toetsen partijdig zijn voor bepaalde groepen leerlingen. Het is helder dat de toetsen dus niet op een correcte wijze gebruikt kunnen worden om individuele leerlingen te oormerken als zwak taalvaardig en zeker niet als men individuele consequenties aan de toetsuitslag verbindt. Bij het meten van de taalvaardigheid van groepen leerlingen, bijvoorbeeld voor het doen van onderzoek of het toekennen van achterstandsgelden op schoolniveau, is het gebrek aan meetinvariantie weliswaar een iets minder groot probleem. Echter, ook bij toepassingen op groepsniveau kan het gebrek aan meetinvariantie over de bovengenoemde groepen een onterechte vertekening van de resultaten geven. Zo is het denkbaar dat een effectstudie die één van de in dit opgenomen taaltoetsen gebruikt als afhankelijke variabele, wel effecten vindt voor leerlingen die thuis een vreemde taal spreken en niet voor leerlingen die thuis Nederlands spreken. Dit onderscheid tussen beide voornoemde groepen zou dan echter geheel aan het gebrek aan meetinvariantie kunnen liggen. En zo zouden bij toekenning van gelden op schoolniveau scholen benadeeld worden die een relatief grote proportie leerlingen bevatten waarvoor de partijdigheid in de toetsen leidt tot onterecht hoge scores op de gebruikte taaltoets.

Deze resultaten betekenen, kort samengevat, dat we als we deze (sub)toetsen afnemen bij de groepen waarvoor geen meetinvariantie werd gevonden, we in die groepen niet dezelfde constructen meten met de subtoetsen. Het verschil tussen allochtone en autochtone leerlingen of tussen leerlingen die thuis Nederlands spreken dan wel een vreemde taal, is goed te begrijpen. Het is niet verrassend dat het beheersen van een tweede taal iets anders blijkt te zijn dan het beheersen van je moedertaal. Dat de toetsen ook bias vertonen naar leeftijd betekent dat we de scores van leerlingen jonger dan zes niet kunnen vergelijken met de scores van leerlingen ouder dan zes. Dit is niet alleen een probleem bij het doen van onderzoek of het op scholen oormerken van taalzwakke leerlingen, het is ook een probleem dat we (nog) niet kunnen duiden waardoor deze bias ontstaat. In feite impliceert de gevonden bias dat jonge leerlingen (of allochtone leerlingen, of leerlingen van een bepaald leerlinggewicht ofwel een bepaald sociaaleconomisch milieu) een ander oplossingsproces hanteren bij het beantwoorden van de opgaven dan de andere leerlingen.

Vervolgens is gekeken in hoeverre de toetsscores te voorspellen zijn vanuit de achtergrondvariabelen leerlinggewicht (.00, .25 en .90) en thuistaal (Nederlands, dialect of Fries en een vreemde taal). Het blijkt dat de beide variabelen de toetsscores maar matig voorspellen. De grootste proporties verklaarde variantie in de toetsscores bedragen ongeveer 30%. Het oordeel van leerkrachten over de taalvaardigheid van de leerlingen is een nog iets betere voorspeller van de toetsscores. Duidelijk is dat als we de slechtst presterende leerlingen selecteren op basis van de toetsscores, we veelal andere leerlingen aanwijzen dan de leerlingen die thuis een vreemde taal spreken of die een .90-leerlinggewicht hebben.

Indien we op grond van leerlingkenmerken proberen te voorspellen welke leerlingen bij het kwart minst taalvaardige leerlingen horen, zien we dat de voorspelling, afhankelijk van de toets waarop we het slechtste kwart selecteren, slechts vijf tot 45% van de keren een juiste voorspelling oplevert. De bij deze voorspelling gehanteerde leerlingkenmerken betreffen de lerarenoordelen over de taalvaardigheid Nederlands van de leerling, de thuistaal, het leerlinggewicht en de sekse. Op grond van deze resultaten raden we preselectie voorafgaand aan taaltoetsing op grond van de hier gebruikte achtergrondvariabelen af.

Bij de analyses gericht op het bepalen van eventueel verschil in kwaliteit van de toetsafnamen door leerkrachten dan wel door getrainde testleiders, zien we geen duidelijk verschil in kwaliteit tussen beide groepen. Toetsen kunnen dus wat dat betreft ook door leerkrachten worden afgenomen.

Kijken we naar het afnamegemak van de toetsafnamen, dan blijken deze over het algemeen zonder veel problemen te zijn verlopen. De testleiders beoordelen het afnamegemak van de TAK iets minder goed dan het afnamegemak van de andere drie toetsen.

De peiling onder schoolleiders gericht op het bepalen van hun attitude ten aanzien van het toetsen van zeer jonge leerlingen laat zien dat de selecte groep schoolleiders die aan het onderzoek meedeed over het algemeen positief staat tegenover het toetsen van jonge leerlingen.

De belangrijkste conclusies zijn de volgende:

- De toetsen blijken voor leerlingen van vier tot acht jaar oud over het algemeen betrouwbare metingen van taalvaardigheid te zijn.
- Zowel wat de test-hertestbetrouwbaarheid betreft als wat betreft de betrouwbaarheid in het aanwijzen van de 25% slechtst presterende leerlingen, scoren TvK en OBIS iets beter dan TAK en Peabody. Subtoetsen waarvan de homogeniteit, de stabiliteit over de gehele scorerange en de stabiliteit in het oormerken van de 25% minst taalvaardige leerlingen voldoen aan de norm die de COTAN voor individuele beoordelingen stelt (>.80), zijn de TvK en de OBIS-toetsen passieve woordkennis, klankarticulatie, letterkennis en getallenkennis.
- Alle in het onderzoek opgenomen (sub)toetsen blijken binnen de onderzochte groep leerlingen unidimensioneel ofwel één vaardigheid te meten.
- Een toetsende factoranalyse over de verschillende toetsscores laat een goed passend model zien met een woordkennisfactor, een fonologische factor, een technisch-lezenfactor en een rekenfactor. Dit resultaat ondersteunt de validiteit van de toetsen.
- Kijken we naar de gestandaardiseerde factorladingen in het getoetste model, dan blijkt voor passieve woordkennis (receptieve woordenschat) de TvK de hoogste lading te vertonen en volgens dit criterium dus de beste woordkennistoets. De andere drie woordkennistoetsen hebben echter ook hoge factorladingen en doen wat dit betreft nauwelijks voor de TvK onder. Op de fonologische factor hebben de toetsen klankonderscheiding en klankarticulatie van de OBIS de hoogste gestandaardiseerde ladingen. De technisch-lezenfactor bestaat uitsluitend uit toetsen van de OBIS.
- Als we nagaan in hoeverre verschillende toetsen onderling overeenstemmen in het oormerken van de 25% slechtst presterende leerlingen, dan blijken de toetsen hierin flink te verschillen. Het maakt dus veel uit welke toets men kiest om de zwakst taalvaardige leerlingen te selecteren.
- Uit de samenhang tussen toetsscores en de oude indeling naar leerlinggewicht blijkt dat de toetsen niet erg hoog samenhangen met deze indeling. Ook de indeling van leerlingen naar of ze Nederlands als moedertaal spreken of een andere taal hangt niet erg hoog met de toetsscores samen. Duidelijk is dat een belangrijk deel van de .90-leerlingen in de oude gewichtenregeling niet de minst taalvaardige leerlingen zijn.

- Er is gebrek aan meetinvariantie over de toetsen geconstateerd voor de groepen leerlingen die worden gevormd door de variabelen leeftijd, leerlinggewicht en thuistaal. Deze geconstateerde partijdigheid of bias wijst op een validiteitprobleem en moet nader onderzocht worden om zicht te krijgen op de vraag wat de consequenties zijn voor de inzetbaarheid van de toetsen. Duidelijk is dat de toetsen in de bij het biasonderzoek betrokken subgroepen ongelijke vaardigheden meten op een verschillende schaal. De onterechte verschillen in de scores belopen krap een derde tot meer dan een hele standaarddeviatie. Nadere analyse van de modelfit bij leerlingen respectievelijk jonger en ouder dan zes jaar laat zien dat het model onder de zes jaar matig en boven de zes jaar zeer goed past. Dit is ook een aanwijzing dat de toetsen bij leerlingen jonger dan zes niet goed werken.
- Het voorspellen welke leerlingen tot de 25% minst taalvaardige horen op grond van de door de leerkracht ingeschatte taalvaardigheid Nederlands van de leerling, de thuistaal, het leerlinggewicht en de sekse van de leerling, blijkt slechts vijf tot 45% van de keren tot een juiste voorspelling te leiden. Het voor het afnemen van de taaltoets selecteren van leerlingen op grond van de hier gebruikte leerlingkenmerken, raden we dan ook af.
- Om na te gaan hoe goed de voorspellende waarde van de toetsen is voor later presteren, is een meting van het taalvaardigheidniveau van dezelfde leerlingen in groep drie en vier nodig. Zou men besluiten tot deze longitudinale dataverzameling, dan kan tevens nagegaan worden in hoeverre de taalscores schools presteren voorspellen en of de bias naar leeftijd, leerlinggewicht en thuistaal bij oudere leerlingen vermindert.
- De toetsen blijken makkelijk af te nemen bij de jonge leerlingen en leraren doen de afnamen niet minder goed dan getrainde testleiders.
- Schoolleiders die deelnamen aan het onderzoek staan over het algemeen positief tegenover het toetsen van leerlingen in groep één en twee van het basisonderwijs.
- De afname van de vier toetsen levert in de praktijk geen belemmeringen op. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400. Bij OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Gemiddeld kan worden uitgegaan van een afnameduur van 15 à 20 minuten voor de vier toetsen zoals afgenomen in dit onderzoek (de TAK en de TvK zijn in deze studie

niet geheel afgenomen). Bij de TvK kan worden bespaard op afnametijd door groepsgewijze afname of door afname via de computer.

- Bij de TAK en de OBIS kan gedetailleerde diagnostische informatie op subtoetsniveau worden verkregen. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht.

6 Samenvatting

Het SCO-Kohnstamm Instituut en het ITS zijn door het ministerie van OCW verzocht om een onderzoek te doen naar de geschiktheid van taaltoetsen voor het identificeren van leerlingen met een taalachterstand bij aanvang van het basisonderwijs. Daarbij werd gevraagd speciale aandacht te besteden aan de geschiktheid voor het toetsen van de meest taalzwakke leerlingen. Naast psychometrische kwaliteiten als betrouwbaarheid en validiteit, is ook de praktische bruikbaarheid van de toetsen van belang.

Het onderzoek moet antwoord geven op de vraag hoe de toetsing ter bepaling van taalachterstanden bij jonge kinderen, dat wil zeggen kinderen in groep één en twee van de basisschool, het best kan worden uitgevoerd. Na overleg met de opdrachtgever zijn voor dit onderzoek de volgende onderzoeksvragen gespecificeerd:
1) Zijn er betrouwbare en valide taaltoetsen voor het meten van de taalvaardigheid van jonge kinderen? Welke taaltoetsen of onderdelen daarvan zijn het meest geschikt om de leerlingen te identificeren met de laagste taalvaardigheid Nederlands? Welk type leerlingen kan met behulp van deze toetsen worden geselecteerd?
2) Indien er voor het hierboven genoemde doel geschikte toetsen zijn, hoe kan de taaltoetsing dan in de praktijk het best worden georganiseerd, wat zijn de kosten van grootschalige invoering in groep één en twee van het basisonderwijs en wat is de administratieve en organisatorische belasting voor de school bij het gebruik van deze toetsen?
3) Zijn deze taaltoetsen geschikt bij gebruik voor diagnostische doeleinden door leerkrachten?

Om de bovenstaande onderzoeksvragen te beantwoorden zijn bij een steekproef van ruim 800 leerlingen uit groep één en twee taaltoetsen afgenomen. De deelnemende leerlingen zaten op scholen met relatief veel achterstandsleerlingen. De voor het onderzoek geselecteerde toetsen zijn; de toets Taal voor Kleuters (TvK), versie jongste kleuters, onderdeel woordenschat; van de Taaltoets Alle Kinderen (TAK) de subtoetsen passieve woordenschat, klankonderscheiding en klankarti-

culatie; de Peabody (ook passieve woordenschat) en de gehele OBIS (Onderbouwinformatiesysteem) die 16 verschillende subtoetsen bevat, negenvoor taal en zeven voor rekenen. De leerlingen in de steekproef hebben niet alle toetsen gemaakt, maar telkens twee van de vier toetsen. Welke twee toetsen een leerling maakte, is door het toeval bepaald, zodat data geanalyseerd konden worden ervan uitgaande dat de ontbrekende waarnemingen aselect zijn. Ook is ervoor gezorgd dat een deel van de leerlingen dezelfde toets twee maal maakte met een tussenpoos van ongeveer twee weken. Dit laatste is gedaan om de test-hertestbetrouwbaarheid van de toetsen te bepalen. Verder is bij de test-hertestafnamen de helft aselect afgenomen door leerkrachten en de andere helft door de testleiders om zo na te kunnen gaan of de afnamen verricht door leerkrachten en testleiders verschillen in kwaliteit.

Eerst is nagegaan in hoeverre de toetsen betrouwbaar meten. De betrouwbaarheid van de toetsen is bepaald door per subtoets de homogeniteit (Cronbach's alfa) te bepalen en door de test-hertestbetrouwbaarheid te berekenen. De homogeniteit is op te vatten als een maat die aangeeft in hoeverre de vragen behorend bij een subtoets onderling samenhangen, de test-hertestbetrouwbaarheid geeft aan in hoeverre de toetsscores stabiel zijn over tijd.

Uit de betrouwbaarheidsanalyses blijkt dat de subtoetsen die zijn afgenomen voor het onderzoek over het algemeen voldoende homogeen zijn in de totale steekproef van leerlingen (vier tot acht jaar oud). Als we zoals de COTAN (Evers, Van Vliet-Mulder, Resing, Starren, Van Alphen de Veer, & Van Boxtel, 2002) aangeeft voor individuele oordelen als criterium een minimale betrouwbaarheid van .80 nemen, dan blijken alleen de alfa's van enkele kleine OBIS-subtoetsen lager (leesbegrip, kennis van rekenjargon, kennis van namen van vormen, hoofdrekenen en sommen B). De lage homogeniteiten van deze subtoetsen zijn mede een gevolg van het kleine aantal items van deze subtoetsen (5-10). Alfa wordt immers groter naarmate het aantal onderling positief samenhangende items toeneemt.
Omdat de betrouwbaarheid of homogeniteit niet een kenmerk is van een toets, maar van een toets afgenomen in een specifieke populatie, zijn de homogeniteiten nog eens bepaald, maar nu alleen voor de jongste leerlingen. Het onderzoek moet immers nagaan in hoeverre de toetsen voor jonge leerlingen deugen en het is denkbaar dat toetsen voor de jongste leerlingen in de steekproef minder betrouw-

baar zijn dan voor de oudere leerlingen (bv. door minder spreiding bij de jongere leerlingen als gevolg van bodemeffecten).
De homogeniteiten zijn bepaald voor leerlingen van vier tot vijf jaar en voor de leerlingen van vier tot zes jaar. Bij de leerlingen van vier tot vijf jaar blijken inderdaad meer toetsen een alfa lager dan .80 te vertonen. Het betreft Taal voor kleuters en de OBIS-subtoetsen leesbegrip, kennis van letters, leesvaardigheid woorden, geheugen, sommen A, kennis van namen van vormen, hoofdrekenen en sommen B. In de steekproef van leerlingen van vier tot zes jaar blijken dezelfde subtoetsen een alfa onder dan wel boven .80 te vertonen als die in de totale steekproef.

Ook de test-hertestbetrouwbaarheid van de toetsen blijkt over het algemeen goed, al is deze betrouwbaarheid voor enkele subtoetsen laag. Met name de subtoetsen van de TAK, en in mindere mate ook de Peabody, scoren niet hoog voor de test-hertest. Daarnaast is de test-hertest van enkele kleinere subtoetsen uit de OBIS laag. Ook de test-hertestbetrouwbaarheid is nogmaals bepaald voor alleen de jongere leerlingen. Doordat de hertestdata zijn verzameld bij minder leerlingen, is het aantal leerlingen per toets van vier tot vijf jaar oud te klein om de test-hertestbetrouwbaarheid te berekenen. Kijken we naar de substeekproef van vier tot zes jaar, dan zien we dat de test-hertestbetrouwbaarheid van zes toetsen te laag uitvalt. Het betreft de Peabody, de TAK-subtoetsen klankarticulatie, klankonderscheiding en passieve woordenschat en de OBIS-subtoetsen aanvankelijk schrijven, leesbegrip en rekenjargon. Blijkbaar zijn de scores op deze toetsen bij de leerlingen jonger dan zes jaar onvoldoende stabiel.
Voor sommige toetsen echter blijkt dat de test-hertestbetrouwbaarheid fors hoger uitvalt voor de jongere leerlingen. Stellen we weer .80 als norm, dan voldoen nu de TvK en de OBIS-toetsen passieve woordenschat, klankarticulatie, klankonderscheiding, kennis van letters, leesvaardigheid woorden, leesvaardigheid zinnen, geheugen, sommen A, kennis van getallen, kennis van vormen, hoofdrekenen en sommen B. De OBIS blijkt voor wat betreft de test-hertest dus betrouwbaarder naarmate de leerlingen jonger zijn.

De hierboven gepresenteerde resultaten van de betrouwbaarheidsanalyses, betreffen de betrouwbaarheid van de toetsen over de gehele scorerange van de toetsen. Daar we in dit onderzoek moeten kijken in welke mate de taaltoetsen voldoen als

het gaat om het detecteren van de meest taalzwakke leerlingen, is ook nagegaan hoe betrouwbaar de subtoetsen de 25% laagst taalvaardige leerlingen onderscheiden. Uit deze analyses blijkt dat deze betrouwbaarheid veel lager is. Als we berekenen in welke mate de toetsen bij de eerste afname en bij de hertest dezelfde leerlingen aanwijzen als de 25% meest taalzwakke leerlingen, dan blijkt deze overeenstemming 12 van de 17 keer onder de grenswaarde van .80 te liggen (voor vier subtoetsen met een extreem bodemeffect zijn deze analyses niet verricht). De enige toetsen die bij deze analyses een kappa (= betrouwbaarheidsmaat) groter dan .80 vertonen, zijn de toetsen Taal voor Kleuters, OBIS-passieve woordenschat, OBIS-klankarticulatie, OBIS-kennis van letters en OBIS-kennis van getallen.
De enige (sub)toetsen die een voldoende scores op alle berekende indicatoren van betrouwbaarheid, zijn de OBIS-toetsen passieve woordkennis, klankarticulatie en kennis van getallen.

Ook is met behulp van principale componentenanalyses en toetsende factoranalyses nagegaan in hoeverre de subtoetsen ieder voor zich unidimensioneel zijn. Deze analyses leveren net als de alfa's, maar dan wat strenger, een indicatie van in hoeverre de vragen van één subtoets hetzelfde meten. De uitkomsten van deze analyses laten zien dat alle subtoetsen in de gebruikte steekproef unidimensioneel zijn en dus dat we voor alle subtoetsen die in het onderzoek zijn opgenomen, somscores mogen berekenen.

Om na te gaan in welke mate verschillende subtoetsen dezelfde dan wel verschillende vaardigheden meten, zijn vervolgens op de somscores van de subtoetsen exploratieve factoranalyses verricht. Uit deze analyses blijkt dat de verschillende subtoetsen een zeer goed inhoudelijk te verdedigen structuur vertonen. Op grond van de resultaten van de exploratieve factoranalyses en van theoretische noties is vervolgens een te toetsen factormodel ontwikkeld. In dit model onderscheiden we een woordkennisfactor, een fonologische factor, een technisch-lezenfactor en een rekenfactor. Op de woordkennisfactor laden de toetsen die bedoeld zijn om receptieve ofwel passieve woordkennis te meten: TvK, Peabody, het onderdeel passieve woordenschat van de TAK en van de OBIS de onderdelen passieve woordenschat, rekenjargon en kennis van vormen. Op de fonologische factor laden de toetsen klankarticulatie en klankonderscheiding van zowel de OBIS als de

TAK (dus vier toetsen in totaal). Op de technisch-lezenfactor laden de OBIS-toetsen aanvankelijk schrijven, leesbegrip, kennis van letters, leesvaardigheid woorden en kennis van getallen. Op de rekenfactor tenslotte laden de OBIS-toetsen geheugen, sommen A, hoofdrekenen en sommen B.
Een toetsende factoranalyse laat zien dat dit model redelijk past. Dit resultaat ondersteunt de constructvaliditeit van de in het factormodel opgenomen subtoetsen en laat zien dat de vaardigheden die de subtoetsen meten zijn te groeperen in vier vaardigheden. Ook is door dit resultaat duidelijk dat we de scores op verschillende subtoetsen die bij één factor horen met behulp van item-responsmodellen op één en dezelfde schaal kunnen zetten.

Vervolgens is nagegaan of de toetsen meetinvariant ofwel onpartijdig zijn ten aanzien van enkele groepen of subpopulaties in de steekproef. Deze analyses gaan na of de subtoetsen in de onderscheiden groepen op een gelijke schaal meten (krijgen leerlingen behorend bij verschillende groepen die even vaardig zijn dezelfde score, ofwel is een score van 10 in groep één vergelijkbaar met een score van 10 in groep 2) en of de factorstructuur van de gemeten vaardigheid in beide groepen gelijk is. Als dit laatste niet het geval blijkt, zijn in beide groepen ondanks dat er één toets wordt gebruikt toch verschillende aspecten van taalvaardigheid gemeten. Deze partijdigheids- of biasanalyses zijn verricht voor groepen gebaseerd op sekse (jongens versus meisjes), thuistaal (spreekt men thuis Nederlands of een vreemde taal), de oude en inmiddels herziene leerlinggewichten (de .00-leerlingen versus de overigen en de .90-leerlingen versus de overigen) en leeftijd (leerlingen van vier tot zes versus leerlingen van zes tot acht). Verder zijn deze analyses niet per toets verricht, maar voor alle taaltoetsen ineens.
Omdat het op deze wijze doen van onderzoek naar partijdigheid of bias van essentieel belang is dat bij aanvang van de analyses wordt uitgegaan van een factormodel dat echt goed past, zijn nog enkele kleine wijzigingen in het uitgangsmodel aangebracht. De rekenfactor is verwijderd, omdat het model met alleen drie taalfactoren veel beter past en omdat het onderzoek zich gegeven de onderzoeksvragen moet richten op het toetsen van taalvaardigheid. Verder zijn de kleine OBIS-subtoetsen verwijderd, omdat deze, mede door het geringe aantal items per subtest, onvoldoende betrouwbaar meten. In dit model wordt woordenschat gemeten met de passieve woordkennistoetsen TvK, Peabody, TAK-passieve woordkennis en OBIS-passieve woordkennis. De fonologische factor wordt ge-

vormd door de subtoetsen klankarticulatie en klankonderscheiding van zowel de TAK als de OBIS (vier in totaal dus). De technisch-lezenfactor tenslotte wordt gevormd door de OBIS-subtoetsen kennis van letters, leesvaardigheid woorden en kennis van getallen. Het factormodel met deze drie taalfactoren past bijzonder goed.

Uit de analyses naar partijdigheid blijkt dat de toetsen niet partijdig zijn voor jongens of meisjes, maar wel voor alle andere onderscheiden groepsindelingen (thuistaal, leerlinggewicht en leeftijd). Deze partijdigheid betreft zowel de schaal waarop gemeten wordt (thuistaal en leeftijd) als de factorstructuur (thuistaal, leerlinggewicht en leeftijd). Binnen de onderscheiden groepen meten de toetsen taalvaardigheid dus niet op dezelfde schaal en, wat nog veel vervelender is, niet op een vergelijkbare wijze. Dit is een probleem als men de toetsen wil gebruiken om individuele leerlingen uit de voornoemde groepen op een eerlijke wijze te oormerken als zijnde zwak of juist niet zwak taalvaardig.

Om een indicatie te krijgen van de ernst van de gevonden partijdigheid, is voor de gevonden schaalverschil-partijdigheid nagegaan hoe groot de onterechte verschillen tussen scores van leerlingen in verschillende groepen zijn als we ze uitdrukken in standaarddeviaties. Schaalpartijdigheid werd gevonden voor de groepen gebaseerd op de variabelen thuistaal en leeftijd en de verschillen variëren over toetsen voor thuistaal van een derde tot één en een kwart standaarddeviatie en voor leeftijd een halve tot één standaarddeviatie. Deze verschillen zijn dus enorm groot. Uiteraard is het werkelijke invaliderende effect van de gevonden partijdigheid veel groter dan in deze effectmaat wordt uitgedrukt. Immers, we meten niet alleen op een verschillende schaal, we meten ook verschillende vaardigheden in de onderscheiden groepen (thuistaal, leerlinggewicht en leeftijd). De mate waarin toetsen verschillende vaardigheden meten, zijn niet in effectgroottes uit te drukken, daar we dan verschillende eenheden zouden vergelijken.

Met name de partijdigheid naar leeftijd blijkt groot. Als we bij het toetsen of er sprake is van meetinvariantie niet uitgaan van de strengste norm (het verschil in χ^2), maar van een veel coulanter criterium (RDR), dan blijkt volgens dit coulante criterium alleen de meetinvariantie naar leeftijd verworpen te moeten worden. De andere hierboven aangetoonde afwezigheid van meetinvariantie (naar leerlinggewicht en thuistaal) valt volgens dit coulante criterium dus nog mee. Het gebrek aan meetinvariantie naar leeftijd is dus het ernstigst. Bij leerlingen jonger dan zes meten de taaltoetsen echt iets anders dan bij leerlingen ouder dan zes. Om deze

reden is apart nagegaan hoe het gehanteerde factormodel past bij respectievelijk leerlingen jonger en leerlingen ouder dan zes jaar. Uit deze analyses blijkt dat het model voor de leerlingen ouder dan zes jaar heel erg goed past, zelfs exacte fit wordt niet verworpen, terwijl voor leerlingen jonger dan zes het model relatief slecht blijkt te passen. Ook is nagegaan of verschil in gehanteerde meetmethode (papier versus p.c.) verantwoordelijk zou kunnen zijn voor (een deel van) de gevonden partijdigheid, maar dit bleek niet het geval te zijn.

Hoewel de resultaten van het partijdigheidonderzoek dusdanig zijn dat men zich terecht af kan vragen of een toetsing van de taalvaardigheid bij de jonge leerlingen in onze steekproef op een deugdelijke wijze kan geschieden aan de hand van deze toetsen, is toch nog verder gekeken naar hoe de toetsen zich gedragen. Nagegaan is in hoeverre de scores op de subtoetsen samenhangen met de oude gewichtenregeling die onderscheid maakt tussen .00-, .25-, .40-, .70- en .90-leerlingen. In onze data bevinden zich overigens vrijwel uitsluitend .00-, .25- en .90-leerlingen, zodat de analyses zich tot deze drie groepen beperken. Daarnaast is nagegaan in hoeverre de scores op de toetsen samenhangen met de taal die thuis gesproken wordt. Deze variabele onderscheidt leerlingen die thuis Nederlands spreken, leerlingen die thuis een Nederlands dialect of Fries spreken en leerlingen die thuis een vreemde taal spreken. Uiteraard zijn de samenhangen met deze twee groepsvariabelen bestudeerd na controle voor verschil in vaardigheid veroorzaakt door leeftijdsverschillen tussen de leerlingen. Tegelijk is nagegaan of de toetsen voldoende spreiding vertonen bij de onderzochte groepen leerlingen.
Uit de analyses blijkt dat de scores van de meeste subtoetsen redelijk tot goed spreiden. Ook blijkt dat de toetsscores niet erg hoog samenhangen met de groepsindelingen naar leerlinggewicht of thuistaal. Hieruit volgt dat als we leerlingen op grond van de toetsscores selecteren, we heel andere groepen zullen vinden dan wanneer we selecteren op thuistaal of leerlinggewicht.

Een ander belangrijk aspect is of de verschillende subtoetsen overeenstemmen bij het oormerken van de 25% qua taal slechtst presterende leerlingen. Het antwoord op deze vraag betreft de convergente validiteit van de verschillende (sub)toetsen. Ook voor de validiteitscoëfficiënt geldt een minimumnorm van .80 indien sprake is van scores die consequenties hebben op het individuele niveau. Uit analyses blijkt dat het vrij veel uitmaakt met welk van de toetsen men de selectie maakt.

De mate van overeenstemming tussen verschillende taaltoetsen is ten hoogste .77 (kappa). De eerder genoemde minimumnorm van .80 wordt dus door geen van de toetscombinaties gehaald. Kijken we bijvoorbeeld naar de mate van overeenstemming tussen de vier toetsen voor het meten van receptieve woordkennis, dan vinden we overeenstemmingsmaten (kappa's) die lopen van .49 (TAK-passieve woordkennis en OBIS-passieve woordkennis) tot .77 (OBIS-passieve woordkennis en Peabody). Duidelijk is dat de toetsen niet op gelijke wijze differentiëren tussen het qua prestatie laagste kwart en bovenste driekwart van de leerlingen. Ook is een overzicht gemaakt waarin per toets de leerlingkenmerken leerlinggewicht, sekse en thuistaal van het onderste kwartiel aan scores op de betreffende toets gegeven worden. Ook uit dit overzicht blijkt dat het veel uitmaakt met welke toets men de 25% laagst taalvaardige leerlingen selecteert.

De leerkrachtoordelen over de luister- en spreekvaardigheid Nederlands van de leerlingen blijken vrij hoog samen te hangen met de toetsscores. De correlaties tussen de leerkrachtoordelen over spreekvaardigheid Nederlands en de vier woordenschattoetsen lopen bijvoorbeeld van .53 (Peabody) tot .59 (TAK-passieve woordkennis). Voor luistervaardigheid lopen de correlaties met de passieve woordkennistoetsen van .51 (TvK) tot .56 (OBIS-passieve woordenschat). De onderlinge correlaties tussen de toetsscores op deze toetsen lopen van .86 (Peabody * TvK) tot .63 (OBIS-passieve woordkennis * Peabody). Als we bedenken dat de spreek- en luistervaardigheid niet identiek zijn aan passieve woordkennis, dan kunnen we concluderen dat de leerkrachtoordelen het verrassend goed doen. Zeker als we ook nog bedenken dat elke leerkracht alleen zijn eigen leerlingen als referentie kan gebruiken. Het is immers aannemelijk dat de oordelen die leerkrachten geven over de taalvaardigheid van hun leerlingen gekleurd worden door de gemiddelde taalvaardigheid van de leerlingpopulatie waar de leerkracht mee werkt. Deze gemiddelde taalvaardigheid, ofwel het aantal taalzwakke leerlingen, verschilt per leerkracht en dus kan men verwachten dat leerkrachtoordelen hierdoor een zwakkere relatie met objectiever toetsscores zouden vertonen dan we nu zien. Blijkbaar komen de referentiekaders van de verschillende leerkrachten vrij aardig overeen.

Om na te gaan of het verstandig zou zijn om leerlingen voorafgaand aan een eventuele taaltoetsing te selecteren op grond van leerlingkenmerken, om zo alleen

die leerlingen te toetsen waarvan men kan aannemen dat ze meer kans hebben om als taalzwak aangemerkt te worden, zijn logistische regressieanalyses verricht. De afhankelijke variabele betreft per (sub)toets de gedichotomiseerde toetsscore die aangeeft of een leerling ja dan nee tot het kwart slechtst presterende leerlingen op de toets behoort. De onafhankelijke variabelen zijn leerlinggewicht (.00; .25; .90), thuistaal (vreemde taal; dialect/Fries; Nederlands), sekse en de twee leerkrachtoordelen over respectievelijk de spreekvaardigheid en de luistervaardigheid Nederlands van de leerling. Telkens is eerst de variabele leeftijd in de logistische regressie ingevoerd, zodat voor vaardigheidsverschillen veroorzaakt door leeftijdsverschillen is gecorrigeerd. De percentages verklaarde variantie in de dichotome variabele die aangeeft of een leerling in het onderste kwartiel scoort, lopen na correctie voor de leeftijdsverschillen van 5% (OBIS-geheugen) tot 45% (OBIS-passieve woordkennis). Duidelijk is dat door een preselectie gebaseerd op de hier gebruikte leerlingkenmerken een aanzienlijk deel van de zwak taalvaardige leerlingen (95-55%) onterecht niet getoetst zal worden. Uiteraard zou men de selectie minder strict kunnen maken, zodat deze percentages onterecht niet getoetste leerlingen kleiner worden. Als we bijvoorbeeld een logistische regressie verrichten waarin we niet 25%, maar 50% van de minst taalvaardige leerlingen selecteren, om vervolgens deze 50% met behulp van de taaltoetsen te verdelen in een relatief taalzwakke en een taalvaardige helft, zal het percentage onterecht niet getoetste taalzwakke leerlingen afnemen. Kortom, hoe hoger het percentage taalzwakke leerlingen dat men met behulp van de logistische regressie selecteert voor taaltoetsing, hoe kleiner het opercentage onterecht niet getoetste leerlingen, maar ook hoe geringer de winst die preselectie oplevert.

Ook is gekeken of leerkrachten de toetsen even goed kunnen afnemen als ervaren externe testleiders. Deze twee groepen blijken elkaar in afnamekwaliteit niet veel te ontlopen. De conclusie is dan ook dat leerkrachten goed in staat zijn om deze taaltoetsen af te nemen.
Kijken we naar de commentaren van testleiders op het afnamegemak van de toetsen, dan blijken de toetsafnamen nauwelijks problemen te hebben opgeleverd. Een enquête onder schoolleiders om te peilen hoe zij denken over het toetsen van jonge leerlingen laat zien dat de schoolleiders die bij dit onderzoek betrokken waren (een kleine en niet-random steekproef) positief staan tegenover het toetsen van jonge leerlingen.

Voor het beantwoorden van de laatste twee onderzoeksvragen zijn kwalitatieve analyses verricht. De afname van de vier toetsen levert in de praktijk geen belemmeringen op, zo blijkt uit het onderzoek. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400. Daarnaast is inzet van personeel van de school nodig om de gegevens te verwerken en de resultaten te berekenen. Bij OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Bij de Peabody kost de verwerking van de gegevens weinig tijd, maar is de verkregen informatie ook summier.
De afnameduur varieert. Gemiddeld kan worden uitgegaan van een afnameduur van 15 à 20 minuten. Daarbij moet worden aangetekend dat van de TAK en TvK delen zijn geselecteerd voor dit onderzoek. Afname van de gehele toetsen kost (aanzienlijk) meer tijd. Ook bleek uit het onderzoek dat er een grote spreiding was in de benodigde afnametijd per leerling.

De derde onderzoeksvraag betreft de geschiktheid van de vier taaltoetsen bij gebruik voor diagnostische doeleinden door leerkrachten. De vier toetsen hebben als doel inzicht te geven in de mate van achterstand die leerlingen hebben op taalgebied. Bij de TAK en de OBIS kan informatie op subtoetsniveau worden verkregen. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht, het observatie- en hulpprogramma. Bij de Peabody is het geboden inzicht in de mate van taal-ontwikkeling beperkt en ontbreken suggesties voor verbetering. Deze toets biedt dus de minste mogelijkheden op diagnostisch gebied. Overigens wordt de effectiviteit van de gesuggereerde diagnostische interventies niet empirisch onderbouwd.

De belangrijkste conclusies van het onderzoek zijn:

- De toetsen blijken voor leerlingen van vier tot acht jaar oud over het algemeen betrouwbare metingen van taalvaardigheid te zijn (homogeniteit). Nadere analyses leren echter dat de betrouwbaarheid van de toetsen voor de leerlingen van 4.4 tot zes jaar minder goed is;
- Zowel wat de test-hertestbetrouwbaarheid over de gehele scorerange betreft als wat betreft de hertestbetrouwbaarheid bij het aanwijzen van de 25% slechtst presterende leerlingen, scoren de toetsen minder goed. Acht

van de 21 subtoetsen hebben test-hertestbetrouwbaarheden onder .80. De TvK en de OBIS doen het hier iets beter dan de TAK en de Peabody. De stabiliteit van de toetsen in het oormerken van de 25% minst taalvaardige leerlingen is alleen voor de TvK en voor de OBIS-toetsen passieve woordenschat, klankarticulatie, kennis van letters en kennis van getallen voldoende;

- De (sub)toetsen blijken voldoende unidimensioneel om somscores over de (sub)toetsen te rechtvaardigen;
- Kijken we naar de gestandaardiseerde factorladingen in het getoetste model, dan blijkt voor passieve ofwel receptieve woordenschat de TvK de hoogste lading te vertonen en volgens dit criterium dus de beste woordkennistoets. Op de fonologische factor hebben de toetsen klankonderscheiding en klankarticulatie van de OBIS de hoogste gestandaardiseerde ladingen. De technisch-lezenfactor bestaat uitsluitend uit toetsen van de OBIS;
- Als we nagaan in hoeverre verschillende toetsen overeenstemmen in het oormerken van de 25% slechtst presterende leerlingen, dan blijken de toetsen hierin flink te verschillen. Het maakt dus veel uit welke toets men kiest om de zwakst taalvaardige leerlingen te selecteren;
- Uit de samenhang tussen toetsscores en de oude indeling naar leerlinggewicht blijkt dat de toetsen niet erg hoog samenhangen met deze indeling. Ook de indeling van leerlingen naar moedertaal hangt niet erg hoog met de toetsscores samen. Duidelijk is dat een belangrijk deel van de .90-leerlingen niet de minst taalvaardige leerlingen zijn;
- Er is een forse partijdigheid of bias geconstateerd in de toetsscores, wat wijst op een validiteitprobleem dat nader onderzocht moet worden om zicht te krijgen op de vraag wat de consequenties zijn voor de inzetbaarheid van de toetsen bij jonge leerlingen, bij leerlingen die het Nederlands niet als moedertaal hebben en bij leerlingen uit verschillende sociaaleconomische milieus. Duidelijk is dat de toetsen in de subgroepen niet alleen op een verschillende schaal (thuistaal en leeftijd), maar zelfs ongelijke vaardigheden meten (thuistaal, leerlinggewicht en leeftijd). Deze partijdigheid betekent dat één en dezelfde toets in de onderscheiden groepen verschillende vaardigheden meet. Om een indicatie van de ernst van de gevonden partijdigheid te krijgen, is gekeken naar de grootte van de

schaalverschillen in de onderscheiden groepen. Deze onterechte schaalverschillen in de scores zijn groot en bedragen soms meer dan een standaarddeviatie. Overigens is het feit dat de toetsen in voornoemde groepen ongelijke vaardigheden meten nog veel ernstiger dan het verschil in schaal. Vooral de bias naar leeftijd blijkt fors. Gebruik van de toetsen bij voornoemde groepen zal leiden tot een grote mate van willekeur in de toetsuitslag en zeker in gevallen waarbij er individuele consequenties aan de toetsuitslag verbonden worden, moet het hanteren van deze toetsen over de genoemde groepen worden afgeraden. Afname bij leerlingen ouder dan zes is op grond van onze resultaten wel goed te verdedigen;

- Het vooraf selecteren van leerlingen op grond van leerlinggewicht, sekse, thuistaal en het oordeel van de leerkracht over de taalvaardigheid Nederlands van de leerling om slechts bij een deel van de leerlingen taaltoetsen af te nemen en zo op een efficiëntere wijze de 25% slechtst presterende leerlingen te detecteren, zal ertoe leiden dat een groot deel van de taalzwakke leerlingen niet getoetst wordt. De voornoemde indicatoren hangen onvoldoende samen met de toetsscores van de leerlingen om een efficiënte preselectie te verrichten;
- Om na te gaan hoe goed de voorspellende waarde van de toetsen is voor later presteren, is een meting van het taalvaardigheidniveau van dezelfde leerlingen in groep drie en vier nodig. Zou men besluiten tot deze longitudinale dataverzameling, dan kan tevens nagegaan worden in hoeverre de taalscores schools presteren voorspellen en of de partijdigheid naar leeftijd bij oudere leerlingen vermindert. Gegeven de uitstekende modelfit bij leerlingen van zes tot acht jaar lijkt dit voor leeftijdsbias waarschijnlijk. Of de andere gevonden bias leeftijdafhankelijk is, is een open vraag. Wel is gbleken dat deze bias minder ernstig is dan de leeftijdsbias;
- De toetsen blijken makkelijk af te nemen bij de jonge leerlingen en leraren doen de afnamen niet minder goed dan getrainde testleiders;
- De selecte steekproef van schoolleiders die deelnam aan het onderzoek staat over het algemeen positief tegenover het toetsen van leerlingen in groep één en twee van het basisonderwijs;

- Afname van de vier in het onderzoek gebruikte toetsen levert in de praktijk geen belemmeringen op. De materiële kosten variëren – uitgaand van toetsing van 50 leerlingen per school – van ongeveer € 200 tot ruim € 400, naast personele kosten gemaakt voor het verwerken van de gegevens. Alleen bij de OBIS is geen personele inzet nodig voor de verwerking van de gegevens. Bij de Peabody kost de verwerking van de gegevens weinig tijd, maar is de verkregen informatie ook summier;
- De gemiddelde afnameduur van de toetsen zoals in dit onderzoek afgenomen (TAK en TvK dus deels) varieert van 15 tot 20 minuten. Ook bleek dat er een grote spreiding was in de benodigde tijd per leerling. Bij de TvK kan door groepsgewijze afname worden bespaard op begeleidingstijd bij de afname. De in onze studie gepresenteerde resultaten zijn gebaseerd op individuele afnamen. Onduidelijk is of de betrouwbaarheid en validiteit van de TvK zoals hier gerapporteerd, gelijk blijven bij groepsgewijze afname;
- Drie van de vier toetsen bieden aanwijzingen voor diagnostisch gebruik door leerkrachten. De TAK en de OBIS bieden informatie op subtoetsniveau. De TAK voorziet bovendien in uitgebreide suggesties voor vormen van taalstimulering. Bij de TvK is het diagnostische deel, inclusief hulpprogramma, in een apart pakket ondergebracht, het observatie- en hulpprogramma. Bij de Peabody is het inzicht in de mate van taalontwikkeling beperkt en ontbreken suggesties voor verbetering.

Literatuur

Aarnoutse, C. A. J. (1988/1996). *Begrippentest-plaatjes, begrippentest-woorden en woordenschattest, groep 3 basisonderwijs*. Lisse: Swets.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.

Bishop, D. V. M. (2004). Specific language impairment: Diagnostic dilemmas. In L. Verhoeven & H. van Balkom (Eds.), *Classification of developmental language disorders* (p. 309-326). Londen: Lawrence Erlbaum Associates.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research*, *17*, 303-316.

Bon, W. H. J. van, (1982). *Taaltests voor kinderen, TvK*. Lisse: Swets.

Bosker, R. J. & Guldemond, H. (2004). *Een herijking van de gewichtenregeling.* Groningen: GION Instituut voor Onderzoek van Onderwijs, Opvoeding en Ontwikkeling.

Browne, M. W., & Du Toit, S. H. C. (1992). Automated fittings of nonstandard models. *Multivariate Behavioral Research, 27,* 269-300.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis.* (2nd ed) Hillsdale, NJ: Erlbaum.

Dudgeon, P. (2003). *Niesem. A computer program for calculating noncentral interval estimates (and power analysis) for structural equation modeling. Version: October 2003.* Melbourne: University of Melbourne.

Dunn, L. M., & Dunn, L. M., (2005). *Peabody Picture Vocabulary Test-III-NL. Nederlandse versie Liesbeth Schlichting. Handleiding*. Amsterdam: Harcourt.

Eldik, M. C. M. van, Schlichting, J. E. P. T., Lutje Spelberg, H. C., Meulen, B. F. van der, & Meulen, Sj. van der (1995). *Reynell Test voor taalbegrip, RTT*. Lisse: Swets.

Emmelot, Y., & Van schooten, E. van, (2005). *Eisen taaltoets in het kader van de gewichtenregeling.* Amsterdam: SCO-Kohnstamm Instituut.

Evers, A. Vliet-Mulder, J. C. van, Resing, W. C. M. , Starren, J. C. M. G., Alphen de Veer, R. J. van, & Boxtel, H. van (2002). *COTAN Testboek voor het onderwijs*. z. pl.: NDC-Boom.

Field, A., (2005). *Discovering statistics using spss (and sex, drugs and rock 'n' roll).* London: Sage.

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory. principles and applications.* Dordrecht: Kluwer.

Hays, W. L. (1981). *Statistics*. New York: Holt, Rinehar and Winston.

Hoeven, A. van der (z.j.) *OBIS voor kinderen van 4 tot 6 jaar*. Onderbouwinformatiesysteem. Nijmegen: ITS.

Hu, Li-tze, & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis' conventional criteria versus new alternatives. *Structural Equation Modeling, 6* (1), 1-55.

Koning, L. & Westra, E. (2000). *PRAVOO Leerlingvolg- en hulpsysteem voor groep 1, 2, (3) en SBO*. Lekkerkerk: PRAVOO.

Krom, R. S. H., Ouborg, M. J., & Kamphuis, F. H. (2001). *Wetenschappelijke verantwoording van de toetsseries Luisteren 1, 2 en 3*. Arnhem: Citogroep.

Kuyk, J. J. van, (1996). *Handleiding Taal voor kleuters, LVS*. Arnhem: Citogroep.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods. 1* (2), 130-149.

Meerling (1981). *Methoden en technieken van psychologisch onderzoek. Deel 1. Model, observatie en beslissing. (2e dr.).* Meppel: Boom.

Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, 29, 177-185.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543.

Muthén, L. K., & Muthén, B. O. (1999). *Mplus. The comprehensive modeling program for applied researchers. User's guide.* Los Angeles: Muthén & Muthén.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research, 75* (3), 417-.

Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Slichting, L. (1993; 2003). *TARSP, Taal Analyse Remediering en Screening Procedure.* Lisse: Swets & Zeitlinger.

Schlichting, J. E. P. T., Eldik, M. C. M., Lutje Spelberg, H. C., Meulen, Sj. van der, & Meulen, B. F. (1995). *Schlichting test voor taalproductie*. Lisse: Swets.

Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Veen, A., Roeleveld, J., & Leseman, P. (2000). *Evaluatie van Kaleidoscoop en Piramide. Eindrapportage*. Amsterdam: SCO-Kohnstamm Instituut.

Verhoeven, A., & Vermeer, A. (2001). *Taaltoets Alle Kinderen*. Arnhem: Citogroep.

APPENDIX

Tabel A1: Nagegaan of test- en hertestgemiddelden van elkaar verschillen (leereffect bij hertest), mbv Paired Samples T-Test op cases die de hertest hebben gedaan. (N=df+1)

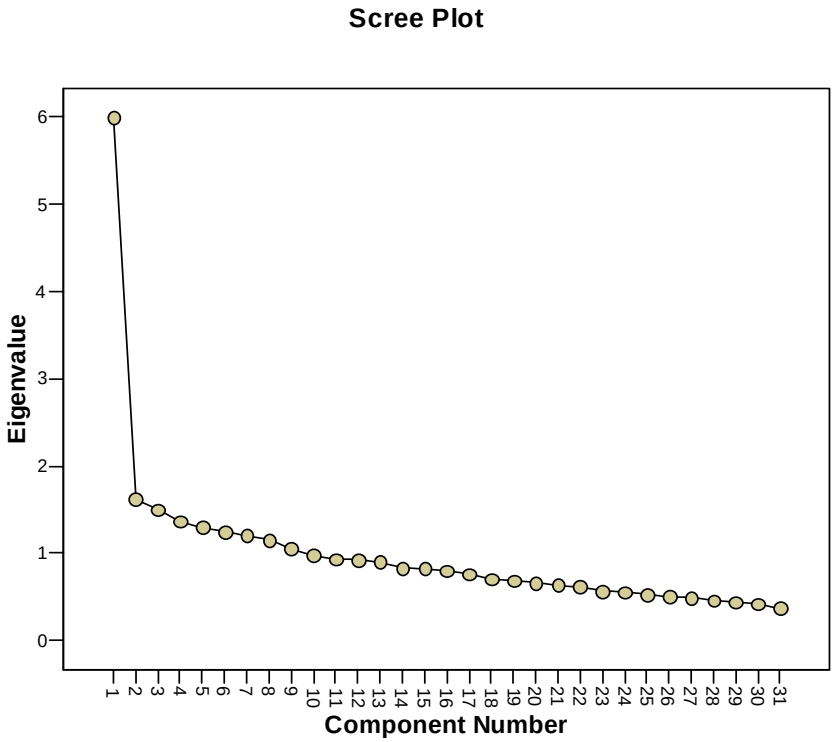
Subtoets	gem. ver-schil	S.D. gemiddeld verschil	S.E. gemiddeld verschil	95% C.I. verschil		t	df	p (2-zijdig)
Taal voor kleuters	-,27	2,09	,22	-,70	,16	-1,25	91	,216
Peabody	-3,57	9,82	1,00	-5,56	-1,58	-3,56	95	,001
tak klankonderscheiding	-4,63	7,90	,81	-6,25	-3,01	-5,68	93	,000
tak klankarticulatie	-1,10	3,84	,40	-1,88	-,31	-2,76	93	,007
tak passieve woordkennis	-11,53	21,05	2,17	-15,84	-7,22	-5,31	93	,000
obis, aanvankelijk schrijven	,122	1,01	,11	-,10	,34	1,09	81	,278
obis passieve woordenschat	-1,98	3,17	,35	-2,67	-1,28	-5,64	81	,000
obis, beginnende geletterdheid, leesbegrip	-1,17	1,87	,21	-1,58	-,76	-5,67	81	,000
obis nazeggen	-,42	1,21	,13	-,69	-,15	-3,11	81	,003
obis klankonderscheiding rijmwoorden	-,61	1,24	,14	-,88	-,34	-4,44	81	,000
obis somscore letterkennis	-,40	2,47	,27	-,95	,14	-1,47	81	,145
obis leesvaardigheid woorden	-,20	1,99	,22	-,63	,24	-,89	81	,377
obis leesvaardigheid zinnen	-,79	3,11	,34	-1,48	-,11	-2,31	81	,023
obis close of lange verhaaltjes juiste woord kiezen	-,02	,50	,05	-,13	,08	-,45	81	,657
obis rekenbegrip	-,28	,85	,09	-,47	-,09	-2,99	81	,004
obis geheugen	,01	,87	,10	-,18	,20	,13	81	,899
obis sommen a	-,60	1,28	,14	-,88	-,32	-4,21	81	,000
obis getallen	-,24	1,67	,18	-,61	,12	-1,33	81	,189
obis vormen	-,21	1,07	,12	-,44	,03	-1,75	81	,084
obis hoofdrekenen	-,21	1,16	,13	-,46	,05	-1,62	81	,110
obis sommen B	-,43	1,38	,15	-,73	-,12	-2,80	81	,006

Tabellen A2 t/m A21: Resultaten homogeniteitsanalyses (Cronbach's alfa) en Scree plots van de PCA's per (sub)toets.

Tabel A2: Taal voor Kleuters, 32 items, receptieve woordkennis (N= 341)

Taal voor kleuters	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
k101a v101a p1	23,48	24,333	,360	,829
k102a v102a p2	23,55	23,771	,413	,827
k103a v103a p3	23,49	24,427	,313	,830
k104a v104a p4	23,60	24,048	,287	,831
k105a v105a p5	23,55	23,760	,426	,827
k106a v106a p6	23,85	23,475	,322	,830
k107a v107a p7	23,49	24,468	,283	,831
k108a v108a p8	23,83	23,308	,361	,828
k109a v109a p9	23,58	23,503	,456	,825
k110a v110a p10	23,74	22,895	,484	,824
k111a v111a p11	23,48	24,403	,348	,830
k112a v112a p12	23,61	24,303	,209	,833
k113a v113a p13	23,60	23,588	,408	,827
k114a v114a p14	23,67	23,581	,356	,828
k115a v115a p15	24,33	25,694	-,178	,841
k116a v116a p16	23,86	24,396	,127	,838
k117a v117a p17	23,76	23,923	,241	,833
k118a v118a p18	23,69	22,908	,506	,823
k119a v119a p19	23,69	23,508	,361	,828
k120a v120a p20	23,62	23,331	,463	,825
k121a v121a p21	23,58	24,256	,241	,832
k122a v122a p22	23,50	24,104	,411	,828
k123a v123a p23	23,72	22,841	,502	,823
k124a v124a p24	23,73	22,956	,472	,824
k125a v125a p25	23,49	24,392	,313	,830
k126a v126a p26	23,49	24,256	,383	,829
k127a v127a p27	23,57	24,375	,217	,833
k128a v128a p28	23,72	24,479	,127	,837
k129a v129a p29	23,77	22,669	,519	,822
k130a v130a p30	23,54	23,790	,432	,827
k131a v131a p31	23,87	23,399	,337	,829
k132a v132a p32	23,55	23,754	,428	,827

Scree plot pca van Taal voor Kleuters, 31 items



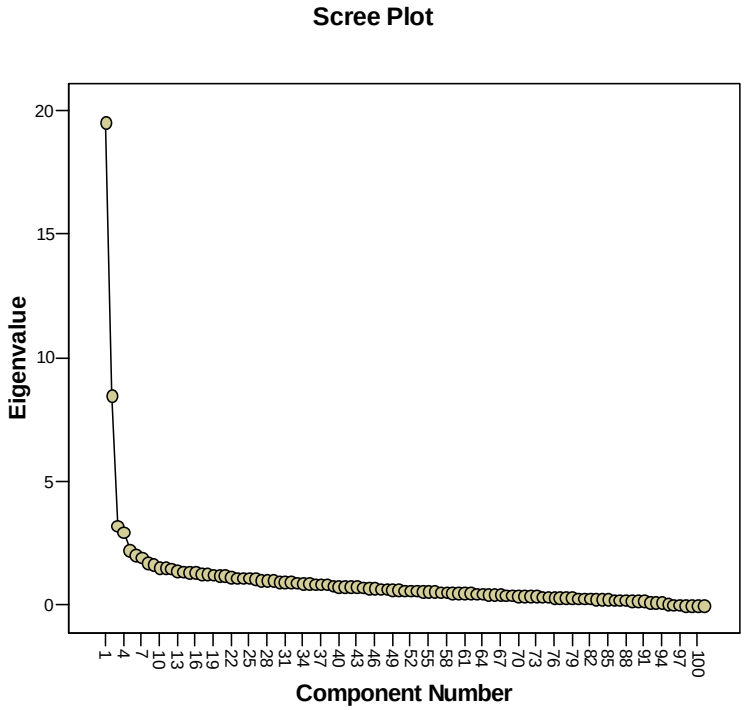
Tabel A3: Peabody, 108 items, receptieve woordkennis (N= 349)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
p1a v1a poes	70.05	244.549	.168	.953
p2a v2a hand	70.05	244.834	.000	.953
p3a v3a schaar	70.06	244.083	.257	.953
p4a v4a oog	70.05	244.538	.175	.953
p5a v5a baby	70.05	244.834	.000	.953
p6a v6a broek	70.05	244.247	.246	.953
p7a v7a drinken	70.05	244.834	.000	.953
p8a v8a vliegtuig	70.05	244.834	.000	.953
p9a v9a lopen	70.05	244.716	.069	.953
p10a v10a schildpad	70.06	243.720	.296	.953
p11a v11a schommelen	70.06	243.821	.301	.953
p12a v12a schep	70.05	244.276	.234	.953
p13a v13a cadeautje	70.05	244.834	.000	.953
p14a v14a springen	70.05	244.834	.000	.953
p15a v15a vlieg	70.05	244.276	.234	.953
p16a v16a hek	70.08	242.778	.389	.953
p17a v17a naar boven	70.06	243.962	.299	.953
p18a v18a bank	70.05	244.549	.168	.953
p19a v19a timmeren	70.08	242.761	.393	.953
p20a v20a computer	70.05	244.834	.000	.953
p21a v21a plant	70.07	242.963	.396	.953
p22a v22a trekken	70.06	243.821	.301	.953
p23a v23a emmer	70.08	242.641	.416	.953
p24a v24a koe	70.06	244.112	.247	.953
p25a v25a trommel	70.07	243.682	.279	.953
p26a v26a kruk	70.11	241.561	.445	.952
p27a v27a pijl	70.19	240.683	.373	.952
p28a v28a ziek	70.08	243.709	.192	.953
p29a v29a blikje	70.16	241.915	.291	.953
p30a v30a varen	70.13	241.747	.350	.953
p31a v31a geld	70.06	244.399	.147	.953
p32a v32a inschenken	70.26	237.981	.529	.952
p33a v33a post	70.07	243.500	.264	.953
p34a v34a slopen	70.39	240.279	.294	.953
p35a v35a kangoeroe	70.20	239.217	.498	.952

p36a	v36a kist	70.08	242.881	.337	.953
p37a	v37a fruit	70.18	241.365	.322	.953
p38a	v38a vaas	70.20	239.787	.442	.952
p39a	v39a handschoen	70.07	243.366	.357	.953
p40a	v40a blij	70.17	241.116	.349	.952
p41a	v41a verrekijker	70.16	240.384	.440	.952
p42a	v42a vuilnis	70.18	239.162	.525	.952
p43a	v43a onder de tafel	70.11	242.140	.348	.953
p44a	v44a cactus	70.21	240.260	.387	.952
p45a	v45a trompet	70.17	239.806	.478	.952
p46a	v46a haai	70.13	241.438	.400	.952
p47a	v47a boren	70.14	241.834	.324	.953
p48a	v48a kin	70.14	241.158	.395	.952
p49a	v49a vitamine	70.17	240.482	.416	.952
p50a	v50a stopcontact	70.32	240.718	.284	.953
p51a	v51a trainen	70.23	240.442	.355	.952
p52a	v52a drieling	70.22	238.906	.499	.952
p53a	v53a groepje	70.20	239.556	.467	.952
p54a	v54a bouwvakker	70.35	237.606	.497	.952
p55a	v55a knagen	70.28	239.500	.396	.952
p56a	v56a vlot	70.47	238.560	.394	.952
p57a	v57a mikken	70.38	238.000	.455	.952
p58a	v58a ambulance	70.23	240.357	.360	.952
p59a	v59a vierkant	70.20	240.337	.395	.952
p60a	v60a tot ziens	70.36	240.335	.298	.953
p61a	v61a ventilator	70.40	237.148	.507	.952
p62a	v62a hurken	70.46	236.778	.516	.952
p63a	v63a sieraad	70.38	237.442	.495	.952
p64a	v64a schoffelen	70.61	238.297	.410	.952
p65a	v65a rimpels	70.48	237.417	.469	.952
p66a	v66a welkom	70.31	238.410	.459	.952
p67a	v67a wortels	70.48	236.319	.542	.952
p68a	v68a afleveren	70.64	238.892	.374	.953
p69a	v69a graan	70.48	238.503	.397	.952
p70a	v70a voetganger	70.48	238.084	.425	.952
p71a	v71a repareren	70.36	238.409	.434	.952
p72a	v72a eiland	70.45	237.697	.455	.952
p73a	v73a hoef	70.64	236.232	.552	.952
p74a	v74a schuin	70.50	236.274	.542	.952

p75a	v75a strompelen	70.63	237.710	.452	.952
p76a	v76a groente	70.63	236.854	.509	.952
p77a	v77a finish	70.62	236.892	.504	.952
p78a	v78a ovaal	70.63	237.119	.491	.952
p79a	v79a huwen	70.83	242.426	.175	.953
p80a	v80a sip	70.70	236.881	.528	.952
p81a	v81a sorteren	70.61	237.617	.455	.952
p82a	v82a prehistorisch	70.79	241.442	.235	.953
p83a	v83a halfvol	70.69	236.513	.548	.952
p84a	v84a kluis	70.60	235.683	.582	.952
p85a	v85a venster	70.79	237.768	.511	.952
p86a	v86a ploegen	70.69	235.934	.587	.952
p87a	v87a van leer	70.86	238.112	.539	.952
p88a	v88a dam	70.79	238.062	.487	.952
p89a	v89a omhelzen	70.88	239.736	.430	.952
p90a	v90a vitrine	70.74	237.968	.467	.952
p91a	v91a woud	70.75	237.331	.517	.952
p92a	v92a geketend	70.76	237.258	.527	.952
p93a	v93a autoriteit	70.85	239.729	.396	.952
p94a	v94a haspel	70.79	237.742	.509	.952
p95a	v95a schuren	70.77	237.864	.486	.952
p96a	v96a prooi	70.81	238.219	.487	.952
p97a	v97a voertuig	70.93	240.062	.472	.952
p98a	v98a onverwacht	70.93	240.613	.416	.952
p99a	v99a burcht	70.92	240.850	.372	.952
p100a	v100a vergiet	70.92	240.425	.413	.952
p101a	v101a dakkapel	70.89	239.785	.438	.952
p102a	v102a vergezellen	70.90	239.669	.460	.952
p103a	v103a verstelbaar	70.93	240.453	.433	.952
p104a	v104a pelikaan	70.83	237.570	.551	.952
p105a	v105a klarinet	70.90	239.599	.459	.952
p106a	v106a pedaal	70.94	240.390	.449	.952
p107a	v107a bankbiljet	70.92	239.827	.471	.952
p108a	v108a hiel	70.87	238.277	.546	.952

Scree plot pca van Peabody, 101 items (7 items verwijderd vanwege plafondeffect; sd=0)

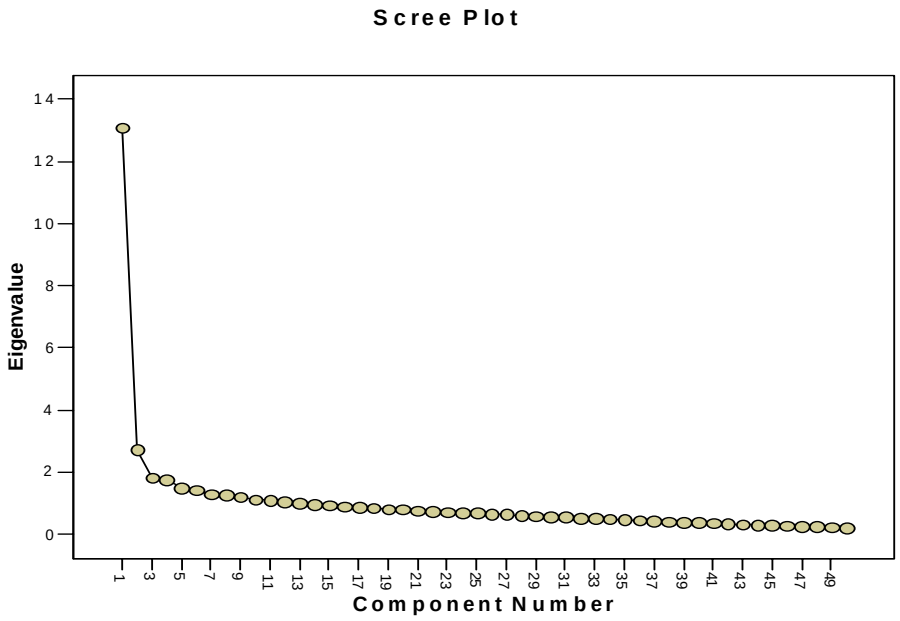


Tabel A4: Taaltoets alle Kinderen, klankonderscheiding, 50 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v101a v101a voer	40.84	69.937	.565	.932
v102a v102a weg	40.86	70.120	.466	.932
v103a v103a pan	40.78	72.302	.106	.934
v104a v104a boos	41.20	69.339	.365	.934
v105a v105a krans	41.00	68.467	.556	.932
v106a v106a beek	40.96	68.779	.548	.932
v107a v107a boer	40.81	71.680	.224	.934
v108a v108a das	41.04	68.811	.482	.932
v109a v109a mier	40.83	70.261	.520	.932
v110a v110a reuk	40.86	69.674	.548	.932
v111a v111a wip	40.80	72.116	.133	.934
v112a v112a zoon	40.90	69.495	.505	.932
v113a v113a hak	40.90	69.118	.580	.931
v114a v114a lijm	40.79	72.018	.182	.934
v115a v115a ton	41.36	69.872	.306	.934
v116a v116a buik	40.87	69.660	.526	.932
v117a v117a vals	40.86	69.241	.633	.931
v118a v118a maag	40.81	72.022	.136	.934
v119a v119a stop	40.88	69.380	.560	.932
v120a v120a goud	40.93	69.339	.499	.932
v121a v121a poes	40.78	72.050	.189	.934
v122a v122a maan	40.95	68.328	.632	.931
v123a v123a groep	41.12	68.825	.444	.933
v124a v124a blok	40.91	68.917	.600	.931
v125a v125a bus	40.80	71.962	.178	.934
v126a v126a pin	40.95	68.736	.564	.931
v127a v127a rus	40.97	68.407	.589	.931
v128a v128a hoed	40.95	68.787	.556	.932
v129a v129a pet	40.79	71.777	.250	.934
v130a v130a dorp	40.91	69.099	.568	.932
v131a v131a koets	41.02	68.518	.535	.932
v132a v132a huis	40.80	72.186	.117	.934
v133a v133a veer	40.93	69.001	.543	.932
v134a v134a praat	40.91	68.775	.624	.931
v135a v135a teen	40.86	69.400	.602	.931

v136a	v136a jas	40.79	72.250	.102	.934
v137a	v137a heg	40.90	68.856	.626	.931
v138a	v138a rok	40.87	68.838	.697	.931
v139a	v139a muis	40.85	69.336	.656	.931
v140a	v140a kool	40.86	71.931	.117	.934
v141a	v141a rok	40.89	68.853	.651	.931
v142a	v142a pijn	40.88	69.226	.589	.931
v143a	v143a mos	40.88	71.777	.131	.935
v144a	v144a trein	41.19	68.798	.433	.933
v145a	v145a beuk	40.91	68.665	.628	.931
v146a	v146a zeur	41.04	68.967	.459	.932
v147a	v147a wit	40.82	71.923	.157	.934
v148a	v148a rem	41.11	68.978	.430	.933
v149a	v149a bleven	41.10	68.451	.499	.932
v150a	v150a bang	41.00	68.328	.574	.931

Scree plot pca van Taaltoets alle kinderen, 50 items

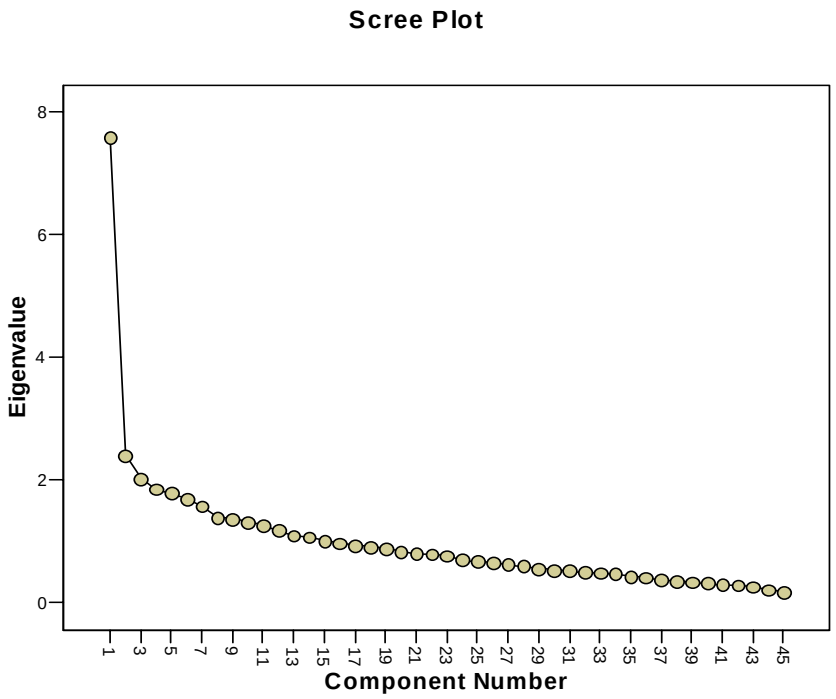


Tabel A5: Taaltoets alle Kinderen, klankarticulatie, 45 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v201a v201a gum	41.17	15.114	.394	.856
v202a v202a lief	41.15	15.355	.320	.857
v203a v203a bok	41.15	15.576	.022	.860
v204a v204a nek	41.15	15.558	.051	.859
v205a v205a fijn	41.16	15.460	.148	.859
v206a v206a hooi	41.16	15.344	.230	.858
v207a v207a soep	41.17	15.091	.414	.855
v208a v208a pit	41.16	15.414	.146	.859
v209a v209a deuk	41.15	15.435	.208	.858
v210a v210a neef	41.17	15.262	.266	.857
v211a v211a zout	41.16	15.407	.168	.858
v212a v212a muur	41.17	15.126	.385	.856
v213a v213a kuit	41.20	15.290	.138	.860
v214a v214a tang	41.23	14.931	.266	.858
v215a v215a jaar	41.20	15.229	.184	.859
v216a v216a spin	41.17	14.997	.468	.854
v217a v217a knoop	41.22	14.775	.373	.855
v218a v218a schram	41.32	14.174	.440	.854
v219a v219a grot	41.25	14.581	.392	.855
v220a v220a strak	41.22	14.327	.593	.850
v221a v221a slaap	41.18	14.820	.530	.853
v222a v222a vlug	41.18	14.938	.465	.854
v223a v223a zwijn	41.19	14.933	.397	.855
v224a v224a kring	41.21	14.622	.481	.853
v225a v225a bleek	41.19	15.235	.201	.858
v226a v226a spijt	41.20	14.836	.418	.854
v227a v227a drop	41.20	14.667	.506	.853
v228a v228a schrik	41.26	14.085	.583	.849
v229a v229a spreuk	41.22	14.372	.570	.850
v230a v230a stout	41.18	14.947	.419	.855
v231a v231a vasts	41.18	15.331	.171	.859
v232a v232a rups	41.20	14.882	.390	.855
v233a v233a herfst	41.33	14.376	.359	.856
v234a v234a arts	41.46	14.312	.302	.860

v235a	v235a hemd	41.19	15.539	.007	.862
v236a	v236a wolk	41.20	14.701	.485	.853
v237a	v237a barst	41.27	14.648	.330	.856
v238a	v238a kind	41.16	15.386	.237	.858
v239a	v239a beurt	41.19	15.114	.260	.857
v240a	v240a fiets	41.18	14.962	.426	.855
v241a	v241a merk	41.22	15.020	.249	.858
v242a	v242a ernst	41.40	14.264	.344	.858
v243a	v243a paars	41.23	14.812	.321	.856
v244a	v244a angst	41.24	14.721	.341	.856
v245a	v245a dans	41.18	15.084	.324	.856

Scree plot pca van Taaltoets alle kinderen, klankarticulatie, 45 items



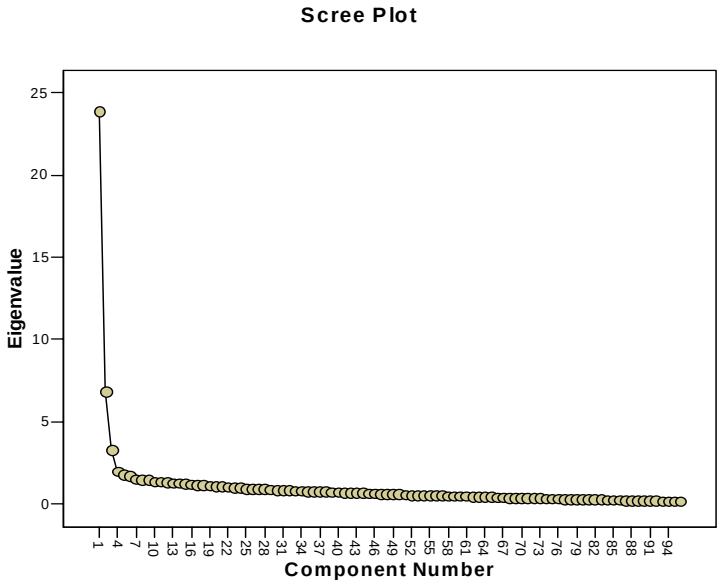
Tabel A6: Taaltoets alle Kinderen, passieve woordenschat, 96 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v301a v301a touw	51.69	356.585	.275	.967
v302a v302a baard	51.72	356.063	.250	.967
v303a v303a helikopter	51.68	357.437	.175	.967
v304a v304a uil	51.71	355.443	.353	.967
v305a v305a pleister	51.69	356.151	.325	.967
v306a v306a knie	51.68	356.616	.312	.967
v307a v307a fles	51.70	356.529	.245	.967
v308a v308a schrijven	51.70	356.741	.226	.967
v309a v309a pet	51.68	356.530	.310	.967
v310a v310a strik	51.78	352.771	.457	.967
v311a v311a zagen	51.72	354.246	.442	.967
v312a v312a veter	51.70	356.303	.293	.967
v313a v313a peer	51.70	356.393	.271	.967
v314a v314a raket	51.75	353.219	.474	.967
v315a v315a schelp	51.70	355.551	.395	.967
v316a v316a hamer	51.73	353.877	.455	.967
v317a v317a branden	51.76	354.575	.332	.967
v318a v318a inkt	51.84	350.833	.514	.966
v319a v319a elleboog	51.80	352.222	.465	.967
v320a v320a bijl	51.92	350.871	.450	.967
v321a v321a gieter	51.84	350.577	.541	.966
v322a v322a spier	51.74	354.674	.368	.967
v323a v323a druppelen	51.89	350.527	.488	.966
v324a v324a nest	51.88	349.532	.566	.966
v325a v325a wimper	51.90	351.387	.428	.967
v326a v326a rots	51.84	350.322	.552	.966
v327a v327a vol	51.97	351.799	.374	.967
v328a v328a spons	51.82	350.556	.570	.966
v329a v329a ladder	51.80	350.851	.575	.966
v330a v330a krans	52.03	349.116	.506	.966
v331a v331a schild	51.95	349.884	.491	.966
v332a v332a rem	51.93	350.029	.498	.966
v333a v333a schroeven	51.94	349.515	.520	.966
v334a v334a fontein	51.95	349.014	.547	.966
v335a v335a kneden	51.92	349.224	.551	.966

v336a	v336a envelop	51.93	348.932	.563	.966
v337a	v337a snoer	52.12	349.501	.469	.967
v338a	v338a gewei	52.28	352.654	.309	.967
v339a	v339a steil	52.09	350.622	.411	.967
v340a	v340a strak	52.03	351.247	.387	.967
v341a	v341a balkon	52.10	350.065	.440	.967
v342a	v342a breekbaar	52.00	349.011	.523	.966
v343a	v343a treuren	52.13	349.665	.460	.967
v344a	v344a apparaat	51.84	349.118	.644	.966
v345a	v345a wenkbrauw	51.95	348.028	.606	.966
v346a	v346a rivier	51.91	347.998	.631	.966
v347a	v347a uitgeput	51.94	347.959	.614	.966
v348a	v348a temmen	52.15	350.706	.403	.967
v349a	v349a eetbaar	52.03	348.156	.559	.966
v350a	v350a wijzer	52.04	346.654	.641	.966
v351a	v351a vacht	52.16	347.758	.562	.966
v352a	v352a korst	51.94	346.825	.682	.966
v353a	v353a leuning	52.07	347.360	.593	.966
v354a	v354a breed	52.39	351.048	.439	.967
v355a	v355a heffen	52.19	350.749	.401	.967
v356a	v356a schaven	52.19	347.226	.592	.966
v357a	v357a kudde	52.14	347.345	.585	.966
v358a	v358a wapen	52.18	347.895	.556	.966
v359a	v359a krom	52.13	346.093	.654	.966
v360a	v360a liniaal	52.18	346.645	.624	.966
v361a	v361a overval	52.09	346.849	.617	.966
v362a	v362a spaak	52.33	349.998	.470	.967
v363a	v363a mengen	52.13	346.976	.605	.966
v364a	v364a pilaar	52.33	349.914	.476	.967
v365a	v365a oceaan	52.21	348.258	.538	.966
v366a	v366a overstroming	52.20	346.289	.645	.966
v367a	v367a operatie	52.20	346.602	.627	.966
v368a	v368a lijst	52.46	351.953	.426	.967
v369a	v369a luik	52.27	346.629	.640	.966
v370a	v370a toespraak	52.23	347.772	.567	.966
v371a	v371a vijl	52.31	348.465	.552	.966
v372a	v372a grendel	52.48	352.797	.388	.967
v373a	v373a zool	52.44	349.979	.542	.966

v374a	v374a verplegen	52.28	347.820	.577	.966
v375a	v375a kabel	52.32	348.895	.529	.966
v376a	v376a kade	52.45	351.303	.465	.967
v377a	v377a beeldhouwen	52.52	353.851	.351	.967
v378a	v378a oever	52.51	352.758	.428	.967
v379a	v379a kuit	52.53	353.686	.371	.967
v380a	v380a klokhuis	52.47	350.541	.540	.966
v381a	v381a geraamte	52.51	353.111	.394	.967
v382a	v382a trechter	52.49	351.253	.507	.966
v383a	v383a loket	52.50	352.501	.433	.967
v384a	v384a greppel	52.53	353.372	.399	.967
v385a	v385a grazen	52.52	352.575	.446	.967
v386a	v386a demonstratie	52.46	350.249	.544	.966
v387a	v387a roer	52.41	349.582	.545	.966
v388a	v388a ventiel	52.52	352.815	.428	.967
v389a	v389a manen	52.53	352.626	.455	.967
v390a	v390a scharnier	52.54	353.155	.430	.967
v391a	v391a kajuit	52.49	351.385	.505	.966
v392a	v392a stronk	52.54	352.864	.464	.967
v393a	v393a dar	52.59	355.457	.307	.967
v394a	v394a lectuur	52.61	356.574	.247	.967
v395a	v395a chirurg	52.58	354.751	.367	.967
v396a	v396a hengsel	52.61	356.181	.299	.967

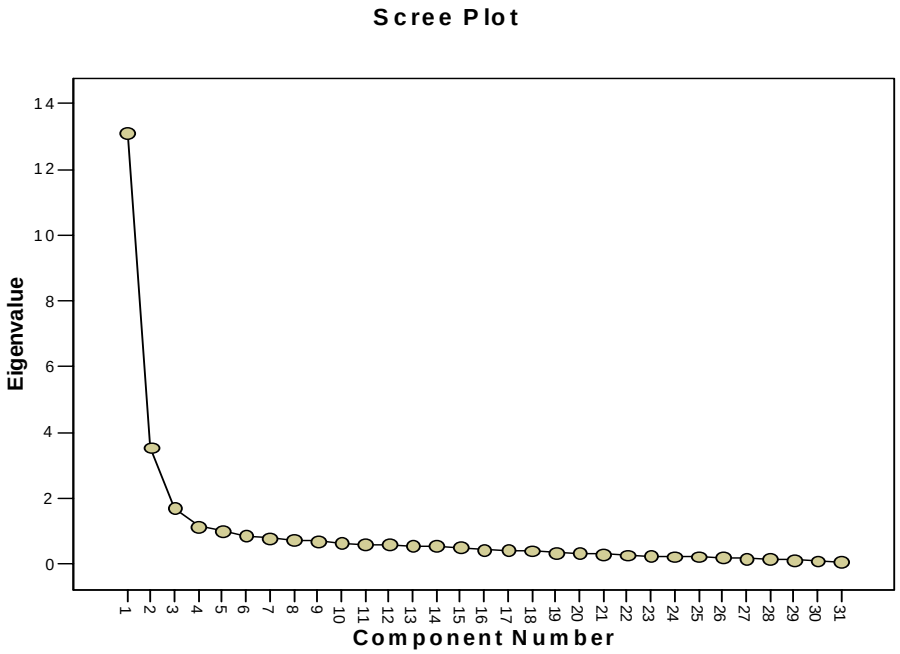
Scree plot pca van Taaltoets alle kinderen, passieve woordenschat, 96 items



Tabel A7: OBIS, passieve woordenschat, 31 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q7PV0_hair	23.51	47.551	.437	.947
q8PV0_eyes	23.48	48.210	.304	.948
q9PV0_mouth	23.48	48.157	.331	.948
q10PV0_shoes	23.47	48.279	.359	.948
q11PV0_arms	23.49	47.866	.449	.947
q12PV0_dog	23.49	47.953	.363	.947
q13PV0_tree	23.53	47.061	.507	.946
q14PV0_bird	23.51	47.434	.450	.947
q16PV1_carrots	23.53	47.119	.479	.947
q17PV1_knife	23.56	46.883	.460	.947
q18PV1_fork	23.59	45.567	.705	.945
q19PV1_cupboard	23.50	47.382	.536	.946
q20PV1_cherries	23.65	45.207	.657	.945
q21PV1_pan	23.58	45.847	.647	.945
q22PV1_bowl	23.80	45.024	.563	.946
q25PV2_butterfly	23.60	45.209	.757	.944
q26PV2_kite	23.62	44.907	.776	.944
q27PV2_castle	23.58	45.242	.806	.944
q28PV2_wasp	23.71	44.096	.792	.943
q29PV2_pigeon	23.71	44.666	.685	.945
q30PV2_windmill	23.67	44.185	.822	.943
q31PV2_turtle	23.65	44.192	.854	.943
q32PV2_violin	23.83	44.582	.620	.946
q33PV2_padlock	23.73	43.970	.793	.943
q34PV2_toadstool	23.68	43.891	.865	.943
q36PV3_yacht	23.88	44.540	.612	.946
q37PV3_cash	23.72	43.798	.830	.943
q38PV3_microscope	24.12	45.903	.420	.948
q39PV3_jewellery	24.07	45.293	.500	.947
q40PV3_saxophone	24.11	45.529	.477	.947
q41PV3_cosmetics	23.99	44.496	.612	.946

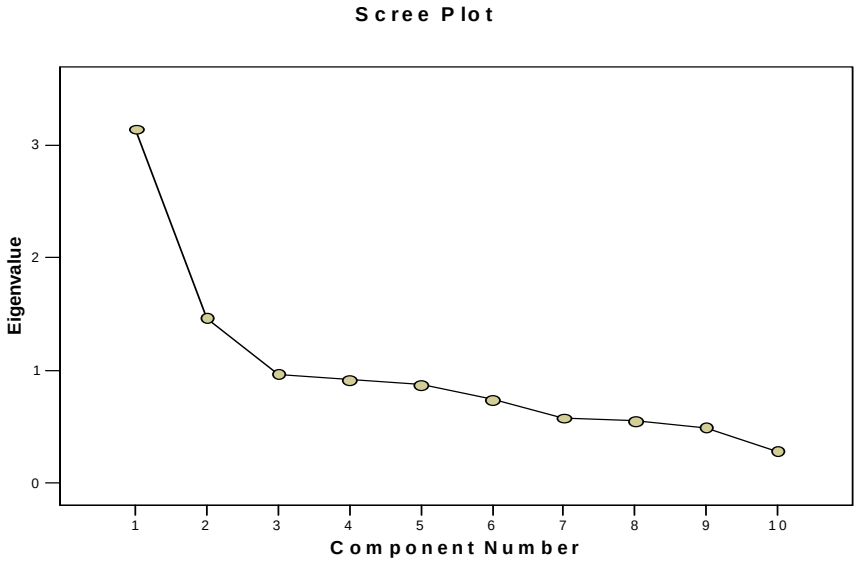
Scree plot pca van OBIS, onderdeel passieve woordenschat, 31 items



Tabel A8: OBIS, beginnende geletterdheid, leesbegrip 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q45IAR1_someone_writing	4.40	4.754	.167	.738
q46IAR1_someone_reading	4.37	4.787	.269	.733
q47IAR1_writing	4.73	4.360	.184	.751
q49IAR2_word	4.70	3.971	.405	.712
q50IAR2_Letter	4.60	3.885	.521	.691
q51IAR2_start_of_story	4.87	3.640	.568	.680
q52IAR2_capital_Letter	5.21	4.545	.211	.736
q53IAR2_full_stop	5.04	3.937	.452	.703
q54IAR2_sentence_start	5.08	3.779	.577	.681
q55IAR2_sentence_finish	5.11	3.819	.587	.681

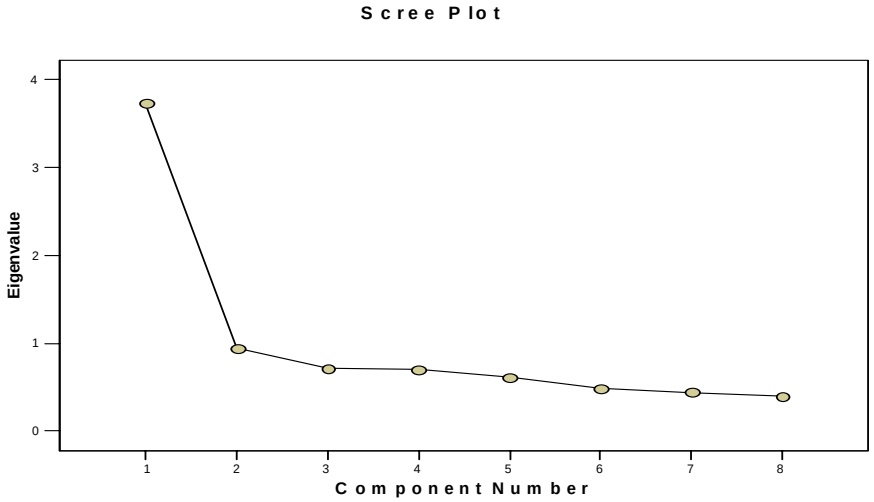
Scree plot pca van OBIS, onderdeel beginnende geletterdheid, leesbegrip 10 items



Tabel A9: OBIS, beginnende geletterdheid, klankarticulatie (onzinwoorden nazeggen) 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q59Rep_stop	5.51	4.070	.240	.840
q60Rep_mantle	5.63	3.400	.549	.811
q61Rep_frigglejang	5.75	3.210	.530	.815
q62Rep_riotous	5.63	3.355	.601	.805
q63Rep_denalty	5.69	3.157	.632	.799
q64Rep_juxtapose	5.64	3.216	.681	.793
q65Rep_enterprising	5.70	3.097	.672	.793
q66Rep_observatory	5.85	3.091	.541	.816

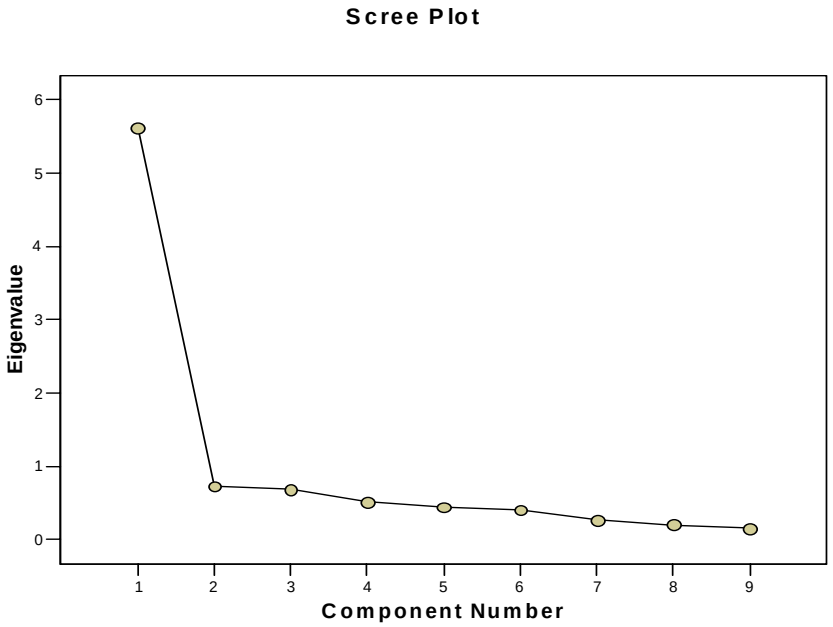
Scree plot pca van OBIS, onderdeel klankarticulatie (onzinwoorden nazeggen) 8 items



Tabel A10: OBIS, beginnende geletterdheid, klankonderscheiding (rijmwoorden herkennen) 9 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q81Rhy_hat	5.61	8.152	.636	.918
q82Rhy_mouse	5.65	8.030	.633	.918
q83Rhy_toes	5.64	8.127	.606	.920
q84Rhy_cherries	5.65	7.848	.713	.913
q85Rhy_sun	5.73	7.486	.796	.908
q86Rhy_bin	5.71	7.395	.855	.904
q87Rhy_drum	5.71	7.409	.855	.904
q88Rhy_dish	5.73	7.393	.838	.905
q89Rhy_pan	5.97	7.874	.567	.924

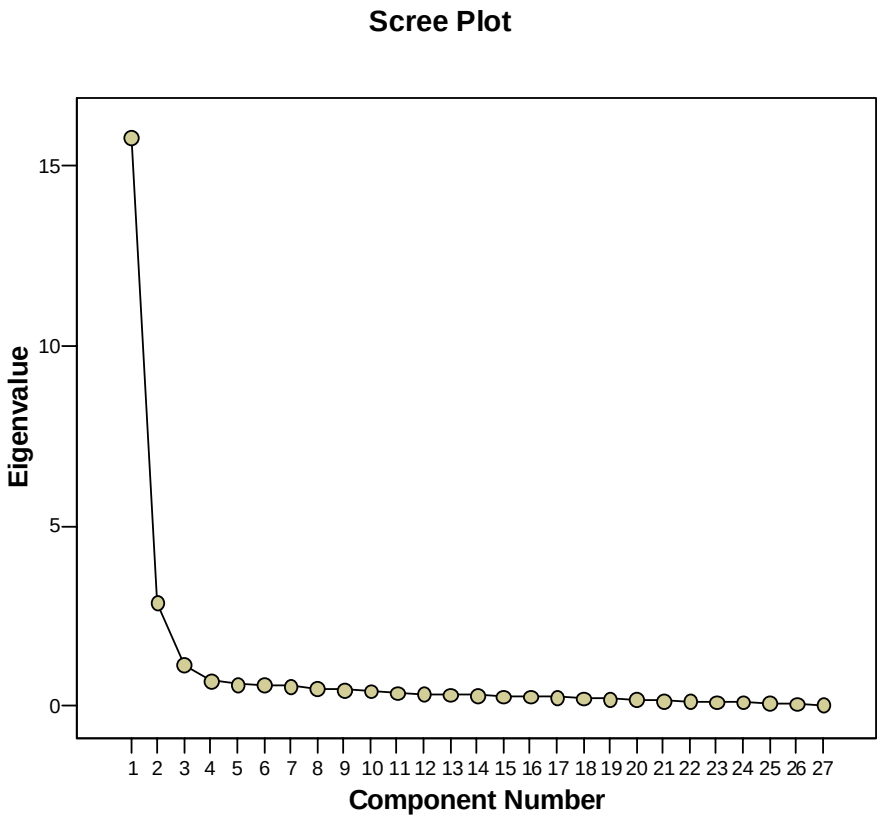
Scree plot pca van OBIS, beginnende geletterdheid, onderdeel klankonderscheiding (rijmwoorden herkennen) 9 items



Tabel A11: OBIS, beginnende geletterdheid, kennis van letters/ het alfabet, 27 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q94Name_Letter	5.94	59.545	.454	.968
q95Lett_S	6.04	59.610	.420	.969
q96Lett_o	6.04	58.398	.586	.967
q97Lett_A	6.21	58.135	.627	.967
q98Lett_c	6.38	58.377	.706	.966
q99Lett_K	6.13	58.009	.632	.967
q100Lett_t	6.37	58.240	.715	.966
q101Lett_B	6.29	57.637	.738	.966
q102Lett_z	6.31	57.213	.817	.965
q103Lett_m	6.32	57.838	.725	.966
q104Lett_x	6.38	57.659	.816	.965
q105Lett_H	6.41	57.817	.836	.965
q106Lett_e	6.40	57.867	.810	.965
q107Lett_r	6.38	57.584	.836	.965
q108Lett_L	6.42	58.203	.784	.965
q109Lett_F	6.44	58.165	.828	.965
q110Lett_w	6.45	58.272	.834	.965
q111Lett_D	6.47	58.599	.806	.965
q112Lett_n	6.46	58.325	.837	.965
q113Lett_V	6.46	58.202	.860	.965
q114Lett_u	6.49	59.160	.772	.966
q115Lett_J	6.51	59.574	.721	.966
q116Lett_Y	6.53	60.366	.631	.967
q117Lett_p	6.48	58.635	.829	.965
q118Lett_G	6.50	59.347	.760	.966
q119Lett_Q	6.55	60.878	.573	.967
q120Lett_i	6.51	59.627	.728	.966

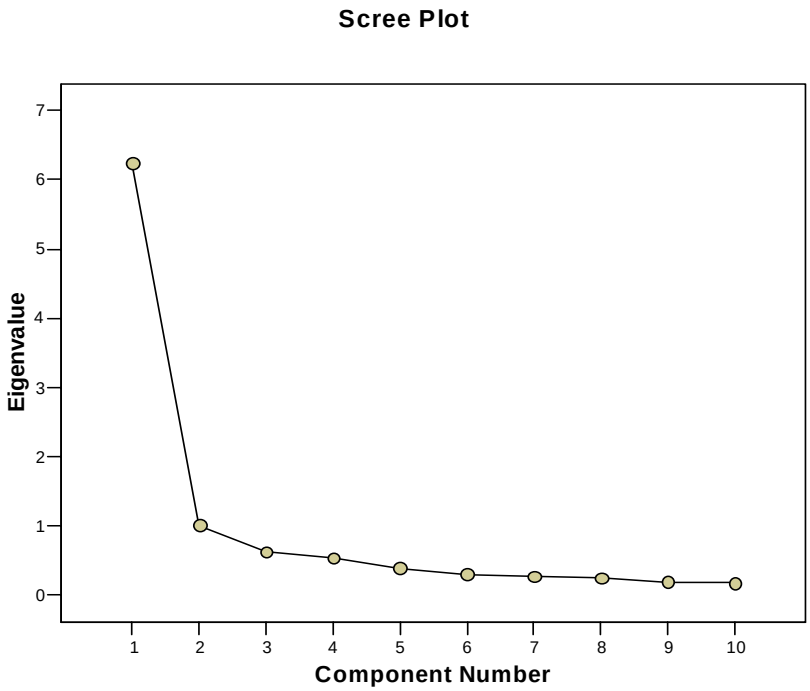
Scree plot pca van OBIS, beginnende geletterdheid, onderdeel kennis van letters/ het alfabet, 27 items



Tabel A12: OBIS, beginnende geletterdheid, leesvaardigheid woorden, 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q179W1_cat	1.97	8.358	.706	.925
q180W1_rabbit	1.97	8.323	.723	.924
q181W1_house	2.03	8.372	.742	.923
q182W1_dog	2.09	8.514	.747	.922
q183W1_duck	2.06	8.306	.800	.920
q185W2_car	2.05	8.149	.856	.916
q186W2_butterfly	2.17	8.834	.729	.924
q187W2_ball	2.19	9.160	.623	.929
q188W2_tree	2.19	8.985	.715	.925
q189W2_flower	2.21	9.086	.693	.926

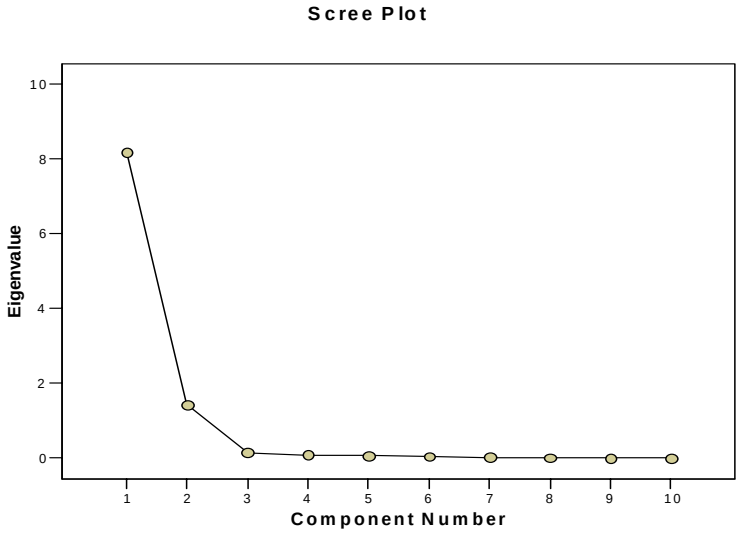
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid woorden, 10 items



Tabel A13: OBIS, beginnende geletterdheid, leesvaardigheid zinnen, 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item -Total Correlation	Cronbach's Alfa if Item Deleted
q204Story_Ben_1	2.37	114.001	.819	.939
q205Story_Ben_2	2.30	105.767	.858	.936
q206Story_Ben_3	2.44	113.331	.877	.937
q208Story_Cat_1	2.10	81.063	.952	.952
q209Story_Cat_2	2.46	101.124	.912	.934
q210Story_Cat_3	2.67	121.902	.885	.942
q212Story_Dog	2.66	119.910	.861	.941
q213Story_Boy	2.67	121.969	.865	.942
q214Story_Sheep	2.63	115.354	.859	.938
q215Story_Cat	2.69	123.627	.869	.944

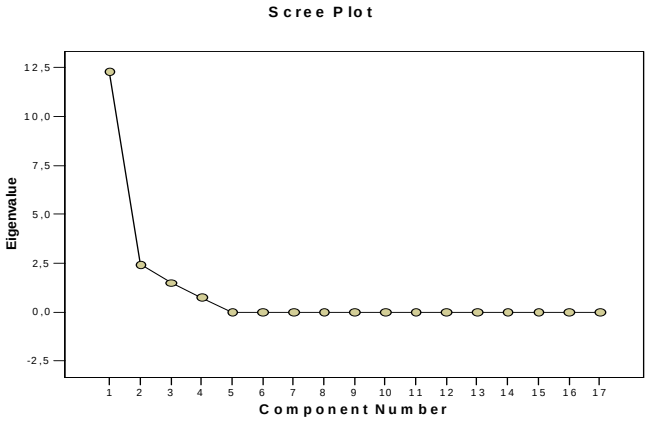
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid zinnen, 10 items



Tabel A14: OBIS, beginnende geletterdheid, leesvaardigheid close meerkeuze, 17 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q219WTS_your	.13	1.505	.585	.977
q220WTS_on	.12	1.367	.964	.972
q221WTS_walk	.12	1.367	.964	.972
q222WTS_two	.12	1.367	.964	.972
q223WTS_leave	.13	1.482	.715	.975
q224WTS_was	.12	1.367	.964	.972
q225WTS_looked	.13	1.416	.885	.973
q226WTS_were	.13	1.482	.715	.975
q227WTS_they	.13	1.416	.885	.973
q228WTS_wearing	.13	1.416	.885	.973
q229WTS_thought	.13	1.416	.885	.973
q230WTS_they2	.13	1.416	.885	.973
q231WTS_for	.13	1.416	.885	.973
q232WTS_buy	.13	1.465	.815	.974
q233WTS_at	.13	1.531	.650	.976
q234WTS_to	.13	1.531	.650	.976
q235WTS_carried	.13	1.482	.715	.975

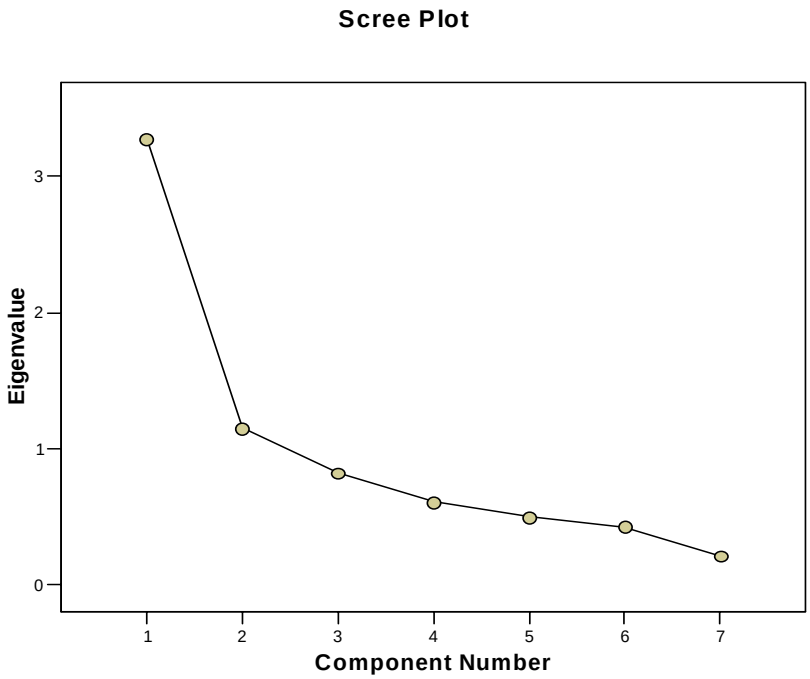
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid close meerkeuze, 17 items



Tabel A15: OBIS, aanvankelijk rekenen, rekenbegrip (kennis rekenjargon), 7 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q258IAM_biggest_cat	5.59	.866	.391	.707
q259IAM_smallest_cat	5.59	.872	.418	.706
q260IAM_more_balloons	5.60	.806	.512	.682
q261IAM_most_water	5.71	.579	.526	.673
q262IAM_least_water	5.76	.576	.423	.733
q263IAM_tallest	5.60	.770	.612	.663
q264IAM_shortest	5.61	.739	.616	.653

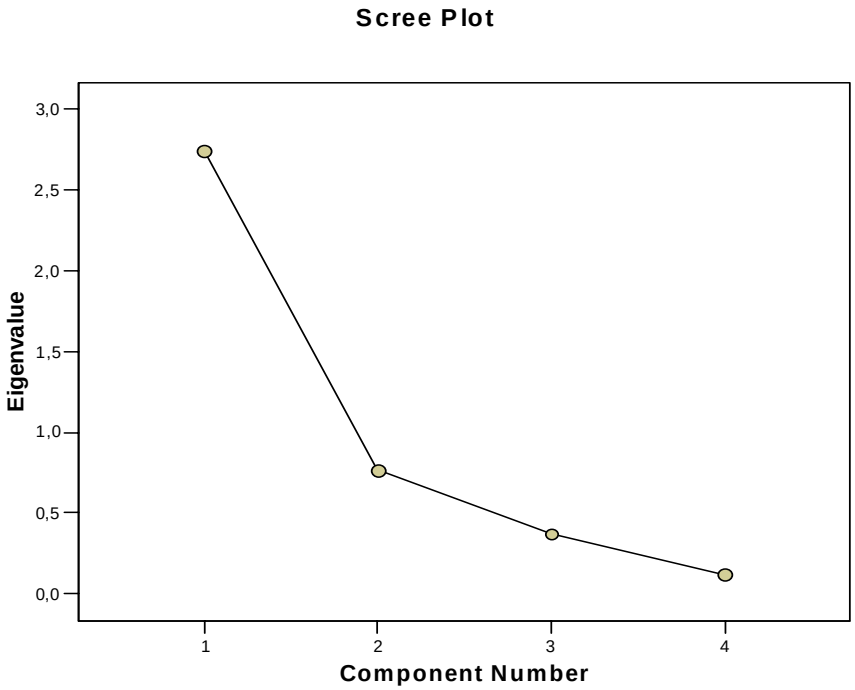
Scree plot pca van OBIS, aanvankelijk rekenen, rekenbegrip (kennis rekenjargon), 7 items



Tabel A16: OBIS, aanvankelijk rekenen, tellen (geheugen voor aantallen), 4 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q267Count_apples_here	2.37	1.080	.612	.828
q268Count_apples_there	2.43	.934	.651	.799
q269Count_fish_here	2.51	.769	.735	.759
q270Count_fish_there	2.60	.684	.753	.758

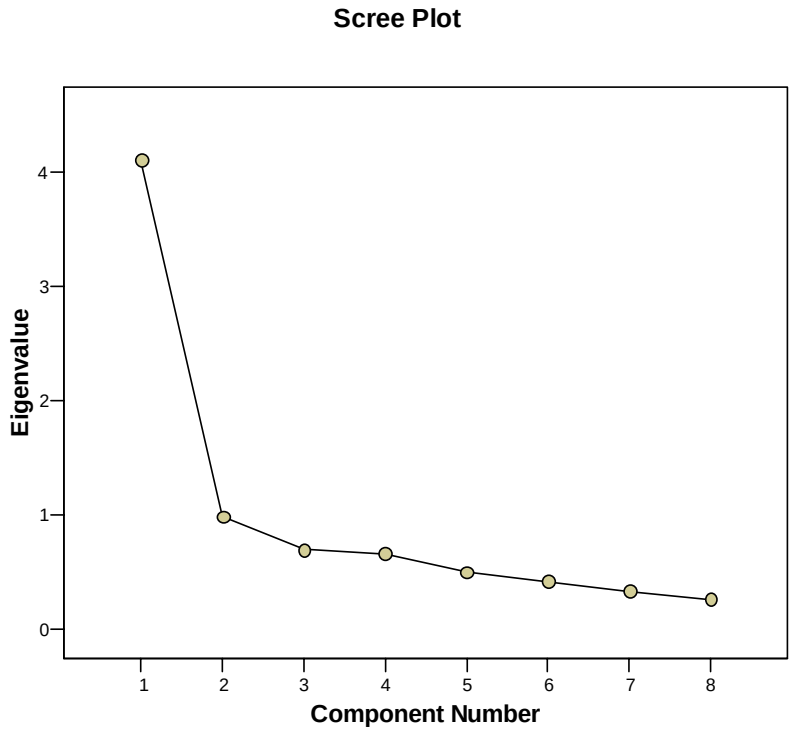
Scree plot pca van OBIS, aanvankelijk rekenen, tellen (geheugen voor aantallen), 4 items



Tabel A17: OBIS, aanvankelijk rekenen, sommen A, 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q272Sums_A_31_balls	4.64	5.135	.506	.849
q273Sums_A_41_cars	4.67	4.863	.626	.837
q274Sums_A_63_ice_cream	4.87	4.664	.498	.850
q275Sums_A_21_rabbits	4.82	4.424	.676	.828
q276Sums_A_31_bikes	4.79	4.364	.744	.820
q277Sums_A_22_puppies	4.85	4.315	.712	.823
q278Sums_A_32_pips	4.99	4.350	.621	.835
q279Sums_A_43_rockets	5.20	4.690	.469	.854

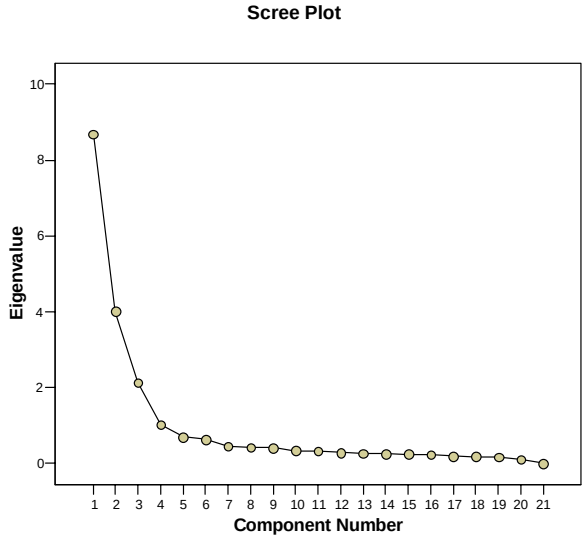
Scree plot pca van OBIS, aanvankelijk rekenen, rekenen sommen A, 8 items



Tabel A18: OBIS, aanvankelijk rekenen, getallenkennis receptief, 21 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q282Num_4	8.29	22.353	.701	.922
q283Num_1	8.18	23.526	.538	.925
q284Num_3	8.23	22.807	.668	.923
q285Num_2	8.23	22.841	.663	.923
q286Num_5	8.28	22.260	.742	.921
q287Num_7	8.38	21.589	.805	.919
q288Num_6	8.41	21.577	.789	.920
q289Num_9	8.46	21.538	.770	.920
q290Num_8	8.42	21.440	.813	.919
q291Num_0	8.38	21.508	.823	.919
q292Num_teen1	8.68	21.950	.671	.923
q293Num_teen2	8.72	22.035	.665	.923
q294Num_teen3	8.72	22.076	.655	.923
q295Num_2dig1	8.93	23.573	.466	.926
q296Num_2dig2	8.92	23.396	.501	.926
q297Num_2dig3	8.93	23.450	.494	.926
q298Num_3dig1	9.06	24.933	.239	.929
q299Num_3dig2	9.07	25.130	.167	.929
q300Num_3dig3	9.06	25.037	.199	.929
q301Num_3dig4	9.06	24.965	.237	.929
q302Num_3dig5	9.06	24.965	.237	.929

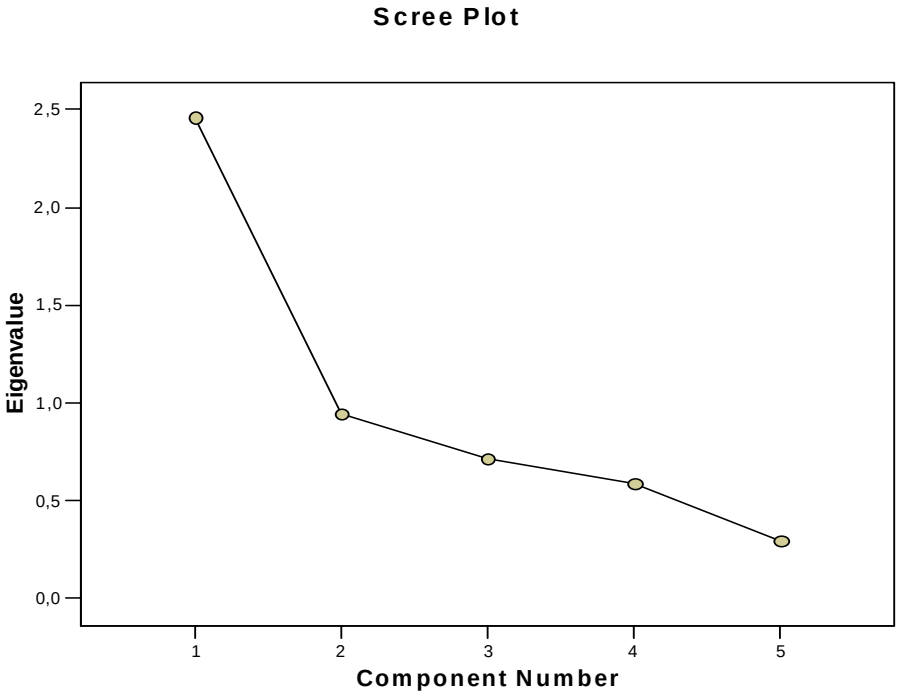
Scree plot pca van OBIS, aanvankelijk rekenen, getallenkennis receptief, 21 items



Tabel A19: OBIS, aanvankelijk rekenen, kennis vormen receptief, 5 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q321Sh_star	3.22	1.451	.201	.767
q322Sh_circle	3.36	1.042	.477	.692
q323Sh_square	3.39	.926	.608	.638
q324Sh_triangle	3.38	.884	.696	.599
q325Sh_hexagon	3.50	.869	.524	.684

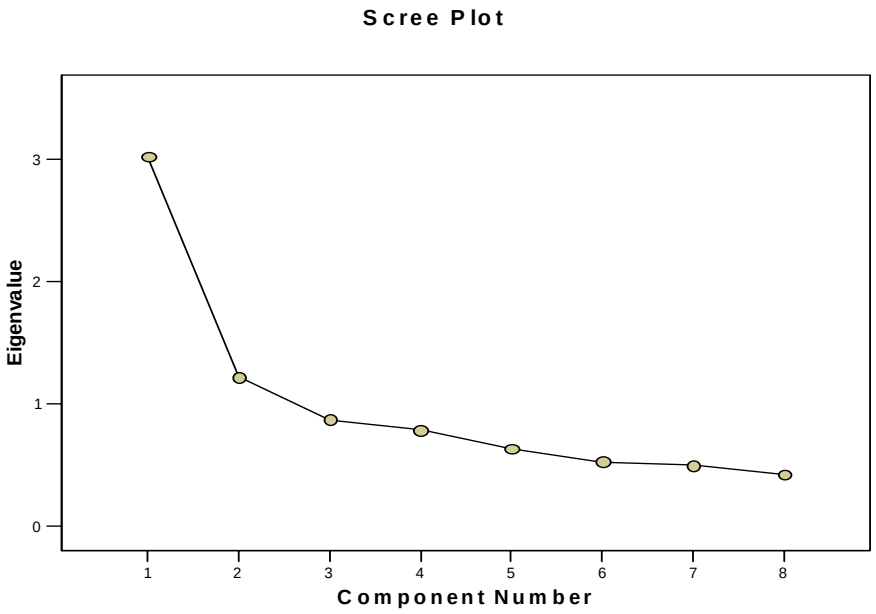
Scree plot pca van OBIS, aanvankelijk rekenen, kennis vormen receptief, 5 items



Tabel A20: OBIS, aanvankelijk rekenen, hoofdrekenen, 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q328M_1_more_than_5	1.11	2.083	.492	.720
q329M_3_less_than_7	1.36	1.928	.609	.689
q330M_2_more_than_6	1.43	1.989	.592	.693
q331M_3_more_than_8	1.56	2.154	.583	.696
q332M_6_less_than_15	1.66	2.483	.450	.725
q333M_8_more_than_13	1.73	2.838	.265	.753
q334M_10_less_than_25	1.69	2.587	.442	.730
q335M_21_more_than_32	1.75	2.940	.190	.761

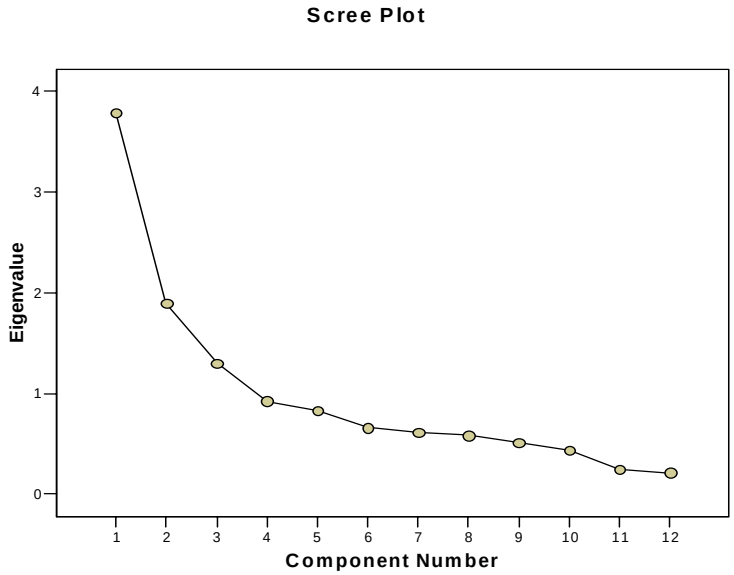
Scree plot pca van OBIS, aanvankelijk rekenen, hoofdrekenen, 8 items



Tabel A21: OBIS, aanvankelijk rekenen, sommen B, 12 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cron-bach's Alfa if Item Deleted
q338Sums_B_three_quarter_circle	1.53	2.454	.307	.715
q339Sums_B_5p_apple_coins	1.38	2.354	.456	.676
q340Sums_B_half_the_bikes	1.82	2.451	.360	.698
q341Sums_B_7+3=	1.96	2.485	.484	.670
q342Sums_B_sequence_10,20,30,40,?	1.96	2.467	.501	.667
q343Sums_B_sequence_2,4,6,8,?,12	2.03	2.710	.441	.681
q344Sums_B_half_of_6	2.06	2.824	.415	.688
q345Sums_B_4+11=	2.09	2.934	.420	.695
q346Sums_B_9-6=	2.10	2.979	.392	.699
q347Sums_B_12p_orange_coin	2.09	2.959	.400	.697
q348Sums_B_15-4=	2.11	3.093	.278	.709
q349Sums_B_15+21=	2.11	3.131	.211	.713

Scree plot pca van OBIS, aanvankelijk rekenen, sommen B, 12 items



APPENDIX

Tabel A1: Nagegaan of test- en hertestgemiddelden van elkaar verschillen (leereffect bij hertest), mbv Paired Samples T-Test op cases die de hertest hebben gedaan. (N=df+1)

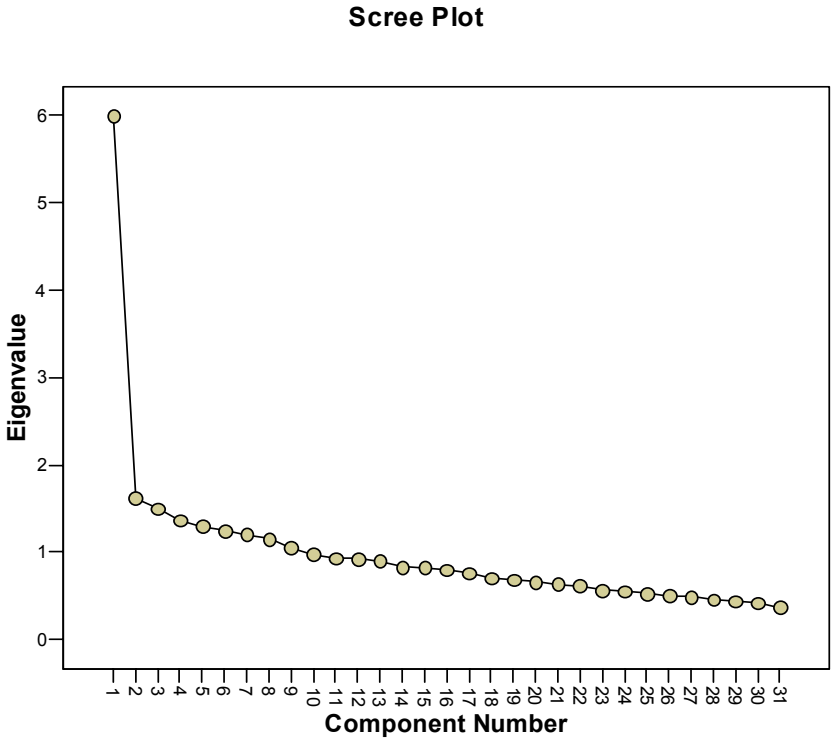
Subtoets	gem. ver-schil	S.D. gemiddeld verschil	S.E. gemiddeld verschil	95% C.I. verschil		t	df	p (2-zijdig)
Taal voor kleuters	-,27	2,09	,22	-,70	,16	-1,25	91	,216
Peabody	-3,57	9,82	1,00	-5,56	-1,58	-3,56	95	,001
tak klankonderscheiding	-4,63	7,90	,81	-6,25	-3,01	-5,68	93	,000
tak klankarticulatie	-1,10	3,84	,40	-1,88	-,31	-2,76	93	,007
tak passieve woordkennis	-11,53	21,05	2,17	-15,84	-7,22	-5,31	93	,000
obis, aanvankelijk schrijven	,122	1,01	,11	-,10	,34	1,09	81	,278
obis passieve woordenschat	-1,98	3,17	,35	-2,67	-1,28	-5,64	81	,000
obis, beginnende geletterdheid, leesbegrip	-1,17	1,87	,21	-1,58	-,76	-5,67	81	,000
obis nazeggen	-,42	1,21	,13	-,69	-,15	-3,11	81	,003
obis klankonderscheiding rijmwoorden	-,61	1,24	,14	-,88	-,34	-4,44	81	,000
obis somscore letterkennis	-,40	2,47	,27	-,95	,14	-1,47	81	,145
obis leesvaardigheid woorden	-,20	1,99	,22	-,63	,24	-,89	81	,377
obis leesvaardigheid zinnen	-,79	3,11	,34	-1,48	-,11	-2,31	81	,023
obis close of lange verhaaltjes juiste woord kiezen	-,02	,50	,05	-,13	,08	-,45	81	,657
obis rekenbegrip	-,28	,85	,09	-,47	-,09	-2,99	81	,004
obis geheugen	,01	,87	,10	-,18	,20	,13	81	,899
obis sommen a	-,60	1,28	,14	-,88	-,32	-4,21	81	,000
obis getallen	-,24	1,67	,18	-,61	,12	-1,33	81	,189
obis vormen	-,21	1,07	,12	-,44	,03	-1,75	81	,084
obis hoofdrekenen	-,21	1,16	,13	-,46	,05	-1,62	81	,110
obis sommen B	-,43	1,38	,15	-,73	-,12	-2,80	81	,006

Tabellen A2 t/m A21: Resultaten homogeniteitsanalyses (Cronbach's alfa) en Scree plots van de PCA's per (sub)toets.

Tabel A2: Taal voor Kleuters, 32 items, receptieve woordkennis (N= 341)

Taal voor kleuters	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
k101a v101a p1	23,48	24,333	,360	,829
k102a v102a p2	23,55	23,771	,413	,827
k103a v103a p3	23,49	24,427	,313	,830
k104a v104a p4	23,60	24,048	,287	,831
k105a v105a p5	23,55	23,760	,426	,827
k106a v106a p6	23,85	23,475	,322	,830
k107a v107a p7	23,49	24,468	,283	,831
k108a v108a p8	23,83	23,308	,361	,828
k109a v109a p9	23,58	23,503	,456	,825
k110a v110a p10	23,74	22,895	,484	,824
k111a v111a p11	23,48	24,403	,348	,830
k112a v112a p12	23,61	24,303	,209	,833
k113a v113a p13	23,60	23,588	,408	,827
k114a v114a p14	23,67	23,581	,356	,828
k115a v115a p15	24,33	25,694	-,178	,841
k116a v116a p16	23,86	24,396	,127	,838
k117a v117a p17	23,76	23,923	,241	,833
k118a v118a p18	23,69	22,908	,506	,823
k119a v119a p19	23,69	23,508	,361	,828
k120a v120a p20	23,62	23,331	,463	,825
k121a v121a p21	23,58	24,256	,241	,832
k122a v122a p22	23,50	24,104	,411	,828
k123a v123a p23	23,72	22,841	,502	,823
k124a v124a p24	23,73	22,956	,472	,824
k125a v125a p25	23,49	24,392	,313	,830
k126a v126a p26	23,49	24,256	,383	,829
k127a v127a p27	23,57	24,375	,217	,833
k128a v128a p28	23,72	24,479	,127	,837
k129a v129a p29	23,77	22,669	,519	,822
k130a v130a p30	23,54	23,790	,432	,827
k131a v131a p31	23,87	23,399	,337	,829
k132a v132a p32	23,55	23,754	,428	,827

Scree plot pca van Taal voor Kleuters, 31 items



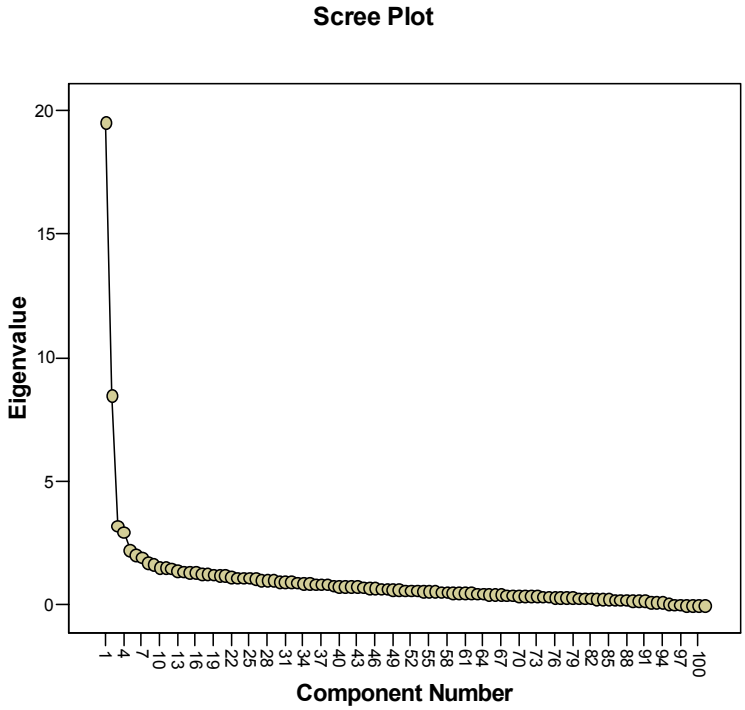
Tabel A3: Peabody, 108 items, receptieve woordkennis (N= 349)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
p1a v1a poes	70.05	244.549	.168	.953
p2a v2a hand	70.05	244.834	.000	.953
p3a v3a schaar	70.06	244.083	.257	.953
p4a v4a oog	70.05	244.538	.175	.953
p5a v5a baby	70.05	244.834	.000	.953
p6a v6a broek	70.05	244.247	.246	.953
p7a v7a drinken	70.05	244.834	.000	.953
p8a v8a vliegtuig	70.05	244.834	.000	.953
p9a v9a lopen	70.05	244.716	.069	.953
p10a v10a schildpad	70.06	243.720	.296	.953
p11a v11a schommelen	70.06	243.821	.301	.953
p12a v12a schep	70.05	244.276	.234	.953
p13a v13a cadeautje	70.05	244.834	.000	.953
p14a v14a springen	70.05	244.834	.000	.953
p15a v15a vlieg	70.05	244.276	.234	.953
p16a v16a hek	70.08	242.778	.389	.953
p17a v17a naar boven	70.06	243.962	.299	.953
p18a v18a bank	70.05	244.549	.168	.953
p19a v19a timmeren	70.08	242.761	.393	.953
p20a v20a computer	70.05	244.834	.000	.953
p21a v21a plant	70.07	242.963	.396	.953
p22a v22a trekken	70.06	243.821	.301	.953
p23a v23a emmer	70.08	242.641	.416	.953
p24a v24a koe	70.06	244.112	.247	.953
p25a v25a trommel	70.07	243.682	.279	.953
p26a v26a kruk	70.11	241.561	.445	.952
p27a v27a pijl	70.19	240.683	.373	.952
p28a v28a ziek	70.08	243.709	.192	.953
p29a v29a blikje	70.16	241.915	.291	.953
p30a v30a varen	70.13	241.747	.350	.953
p31a v31a geld	70.06	244.399	.147	.953
p32a v32a inschenken	70.26	237.981	.529	.952
p33a v33a post	70.07	243.500	.264	.953
p34a v34a slopen	70.39	240.279	.294	.953
p35a v35a kangoeroe	70.20	239.217	.498	.952

p36a	v36a kist	70.08	242.881	.337	.953
p37a	v37a fruit	70.18	241.365	.322	.953
p38a	v38a vaas	70.20	239.787	.442	.952
p39a	v39a handschoen	70.07	243.366	.357	.953
p40a	v40a blij	70.17	241.116	.349	.952
p41a	v41a verrekijker	70.16	240.384	.440	.952
p42a	v42a vuilnis	70.18	239.162	.525	.952
p43a	v43a onder de tafel	70.11	242.140	.348	.953
p44a	v44a cactus	70.21	240.260	.387	.952
p45a	v45a trompet	70.17	239.806	.478	.952
p46a	v46a haai	70.13	241.438	.400	.952
p47a	v47a boren	70.14	241.834	.324	.953
p48a	v48a kin	70.14	241.158	.395	.952
p49a	v49a vitamine	70.17	240.482	.416	.952
p50a	v50a stopcontact	70.32	240.718	.284	.953
p51a	v51a trainen	70.23	240.442	.355	.952
p52a	v52a drieling	70.22	238.906	.499	.952
p53a	v53a groepje	70.20	239.556	.467	.952
p54a	v54a bouwvakker	70.35	237.606	.497	.952
p55a	v55a knagen	70.28	239.500	.396	.952
p56a	v56a vlot	70.47	238.560	.394	.952
p57a	v57a mikken	70.38	238.000	.455	.952
p58a	v58a ambulance	70.23	240.357	.360	.952
p59a	v59a vierkant	70.20	240.337	.395	.952
p60a	v60a tot ziens	70.36	240.335	.298	.953
p61a	v61a ventilator	70.40	237.148	.507	.952
p62a	v62a hurken	70.46	236.778	.516	.952
p63a	v63a sieraad	70.38	237.442	.495	.952
p64a	v64a schoffelen	70.61	238.297	.410	.952
p65a	v65a rimpels	70.48	237.417	.469	.952
p66a	v66a welkom	70.31	238.410	.459	.952
p67a	v67a wortels	70.48	236.319	.542	.952
p68a	v68a afleveren	70.64	238.892	.374	.953
p69a	v69a graan	70.48	238.503	.397	.952
p70a	v70a voetganger	70.48	238.084	.425	.952
p71a	v71a repareren	70.36	238.409	.434	.952
p72a	v72a eiland	70.45	237.697	.455	.952
p73a	v73a hoef	70.64	236.232	.552	.952
p74a	v74a schuin	70.50	236.274	.542	.952

p75a	v75a strompelen	70.63	237.710	.452	.952
p76a	v76a groente	70.63	236.854	.509	.952
p77a	v77a finish	70.62	236.892	.504	.952
p78a	v78a ovaal	70.63	237.119	.491	.952
p79a	v79a huwen	70.83	242.426	.175	.953
p80a	v80a sip	70.70	236.881	.528	.952
p81a	v81a sorteren	70.61	237.617	.455	.952
p82a	v82a prehistorisch	70.79	241.442	.235	.953
p83a	v83a halfvol	70.69	236.513	.548	.952
p84a	v84a kluis	70.60	235.683	.582	.952
p85a	v85a venster	70.79	237.768	.511	.952
p86a	v86a ploegen	70.69	235.934	.587	.952
p87a	v87a van leer	70.86	238.112	.539	.952
p88a	v88a dam	70.79	238.062	.487	.952
p89a	v89a omhelzen	70.88	239.736	.430	.952
p90a	v90a vitrine	70.74	237.968	.467	.952
p91a	v91a woud	70.75	237.331	.517	.952
p92a	v92a geketend	70.76	237.258	.527	.952
p93a	v93a autoriteit	70.85	239.729	.396	.952
p94a	v94a haspel	70.79	237.742	.509	.952
p95a	v95a schuren	70.77	237.864	.486	.952
p96a	v96a prooi	70.81	238.219	.487	.952
p97a	v97a voertuig	70.93	240.062	.472	.952
p98a	v98a onverwacht	70.93	240.613	.416	.952
p99a	v99a burcht	70.92	240.850	.372	.952
p100a	v100a vergiet	70.92	240.425	.413	.952
p101a	v101a dakkapel	70.89	239.785	.438	.952
p102a	v102a vergezellen	70.90	239.669	.460	.952
p103a	v103a verstelbaar	70.93	240.453	.433	.952
p104a	v104a pelikaan	70.83	237.570	.551	.952
p105a	v105a klarinet	70.90	239.599	.459	.952
p106a	v106a pedaal	70.94	240.390	.449	.952
p107a	v107a bankbiljet	70.92	239.827	.471	.952
p108a	v108a hiel	70.87	238.277	.546	.952

Scree plot pca van Peabody, 101 items (7 items verwijderd vanwege plafondeffect; sd=0)

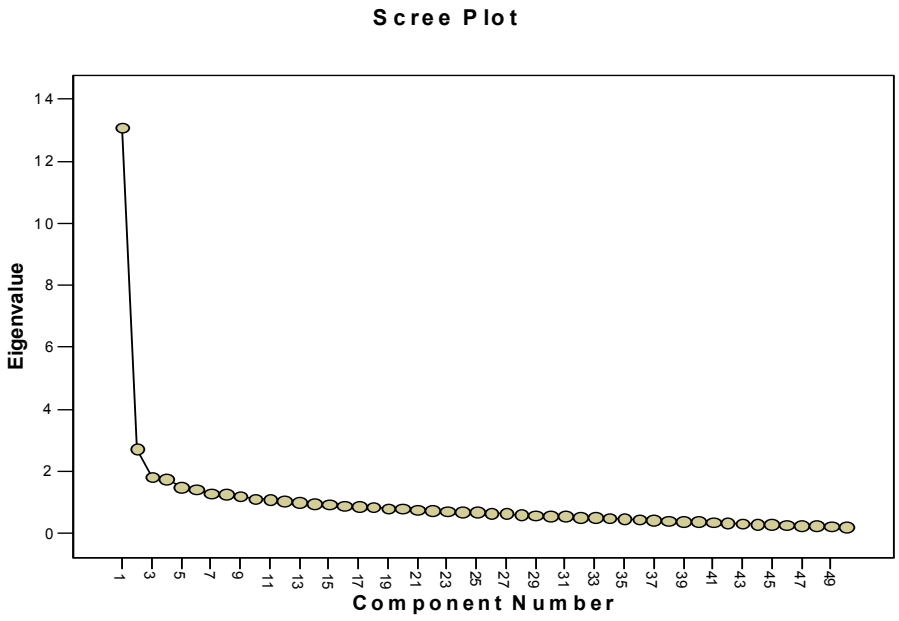


Tabel A4: Taaltoets alle Kinderen, klankonderscheiding, 50 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v101a v101a voer	40.84	69.937	.565	.932
v102a v102a weg	40.86	70.120	.466	.932
v103a v103a pan	40.78	72.302	.106	.934
v104a v104a boos	41.20	69.339	.365	.934
v105a v105a krans	41.00	68.467	.556	.932
v106a v106a beek	40.96	68.779	.548	.932
v107a v107a boer	40.81	71.680	.224	.934
v108a v108a das	41.04	68.811	.482	.932
v109a v109a mier	40.83	70.261	.520	.932
v110a v110a reuk	40.86	69.674	.548	.932
v111a v111a wip	40.80	72.116	.133	.934
v112a v112a zoon	40.90	69.495	.505	.932
v113a v113a hak	40.90	69.118	.580	.931
v114a v114a lijm	40.79	72.018	.182	.934
v115a v115a ton	41.36	69.872	.306	.934
v116a v116a buik	40.87	69.660	.526	.932
v117a v117a vals	40.86	69.241	.633	.931
v118a v118a maag	40.81	72.022	.136	.934
v119a v119a stop	40.88	69.380	.560	.932
v120a v120a goud	40.93	69.339	.499	.932
v121a v121a poes	40.78	72.050	.189	.934
v122a v122a maan	40.95	68.328	.632	.931
v123a v123a groep	41.12	68.825	.444	.933
v124a v124a blok	40.91	68.917	.600	.931
v125a v125a bus	40.80	71.962	.178	.934
v126a v126a pin	40.95	68.736	.564	.931
v127a v127a rus	40.97	68.407	.589	.931
v128a v128a hoed	40.95	68.787	.556	.932
v129a v129a pet	40.79	71.777	.250	.934
v130a v130a dorp	40.91	69.099	.568	.932
v131a v131a koets	41.02	68.518	.535	.932
v132a v132a huis	40.80	72.186	.117	.934
v133a v133a veer	40.93	69.001	.543	.932
v134a v134a praat	40.91	68.775	.624	.931
v135a v135a teen	40.86	69.400	.602	.931

v136a	v136a jas	40.79	72.250	.102	.934
v137a	v137a heg	40.90	68.856	.626	.931
v138a	v138a rok	40.87	68.838	.697	.931
v139a	v139a muis	40.85	69.336	.656	.931
v140a	v140a kool	40.86	71.931	.117	.934
v141a	v141a rok	40.89	68.853	.651	.931
v142a	v142a pijn	40.88	69.226	.589	.931
v143a	v143a mos	40.88	71.777	.131	.935
v144a	v144a trein	41.19	68.798	.433	.933
v145a	v145a beuk	40.91	68.665	.628	.931
v146a	v146a zeur	41.04	68.967	.459	.932
v147a	v147a wit	40.82	71.923	.157	.934
v148a	v148a rem	41.11	68.978	.430	.933
v149a	v149a bleven	41.10	68.451	.499	.932
v150a	v150a bang	41.00	68.328	.574	.931

Scree plot pca van Taaltoets alle kinderen, 50 items

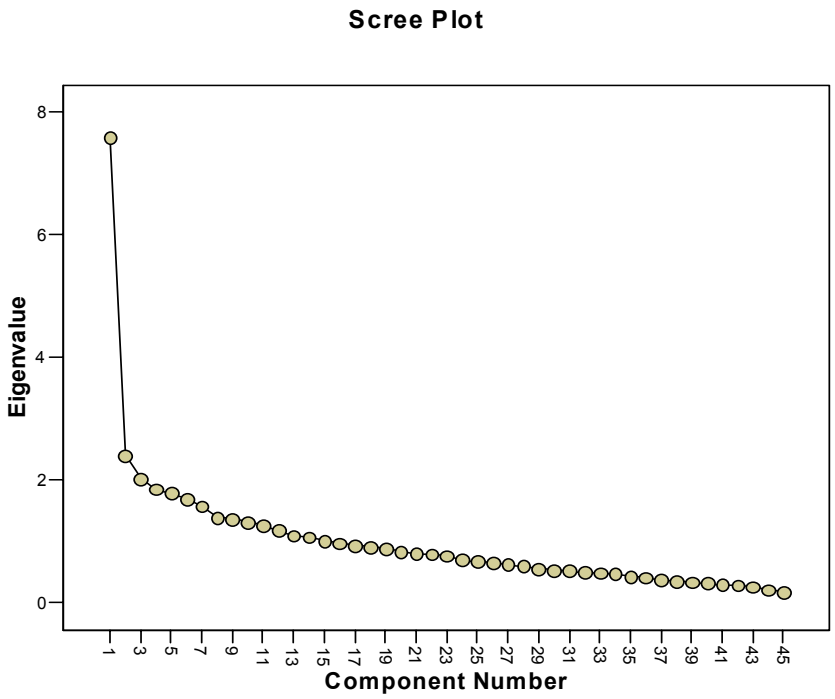


Tabel A5: Taaltoets alle Kinderen, klankarticulatie, 45 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v201a v201a gum	41.17	15.114	.394	.856
v202a v202a lief	41.15	15.355	.320	.857
v203a v203a bok	41.15	15.576	.022	.860
v204a v204a nek	41.15	15.558	.051	.859
v205a v205a fijn	41.16	15.460	.148	.859
v206a v206a hooi	41.16	15.344	.230	.858
v207a v207a soep	41.17	15.091	.414	.855
v208a v208a pit	41.16	15.414	.146	.859
v209a v209a deuk	41.15	15.435	.208	.858
v210a v210a neef	41.17	15.262	.266	.857
v211a v211a zout	41.16	15.407	.168	.858
v212a v212a muur	41.17	15.126	.385	.856
v213a v213a kuit	41.20	15.290	.138	.860
v214a v214a tang	41.23	14.931	.266	.858
v215a v215a jaar	41.20	15.229	.184	.859
v216a v216a spin	41.17	14.997	.468	.854
v217a v217a knoop	41.22	14.775	.373	.855
v218a v218a schram	41.32	14.174	.440	.854
v219a v219a grot	41.25	14.581	.392	.855
v220a v220a strak	41.22	14.327	.593	.850
v221a v221a slaap	41.18	14.820	.530	.853
v222a v222a vlug	41.18	14.938	.465	.854
v223a v223a zwijn	41.19	14.933	.397	.855
v224a v224a kring	41.21	14.622	.481	.853
v225a v225a bleek	41.19	15.235	.201	.858
v226a v226a spijt	41.20	14.836	.418	.854
v227a v227a drop	41.20	14.667	.506	.853
v228a v228a schrik	41.26	14.085	.583	.849
v229a v229a spreuk	41.22	14.372	.570	.850
v230a v230a stout	41.18	14.947	.419	.855
v231a v231a vasts	41.18	15.331	.171	.859
v232a v232a rups	41.20	14.882	.390	.855
v233a v233a herfst	41.33	14.376	.359	.856
v234a v234a arts	41.46	14.312	.302	.860

v235a	v235a hemd	41.19	15.539	.007	.862
v236a	v236a wolk	41.20	14.701	.485	.853
v237a	v237a barst	41.27	14.648	.330	.856
v238a	v238a kind	41.16	15.386	.237	.858
v239a	v239a beurt	41.19	15.114	.260	.857
v240a	v240a fiets	41.18	14.962	.426	.855
v241a	v241a merk	41.22	15.020	.249	.858
v242a	v242a ernst	41.40	14.264	.344	.858
v243a	v243a paars	41.23	14.812	.321	.856
v244a	v244a angst	41.24	14.721	.341	.856
v245a	v245a dans	41.18	15.084	.324	.856

Scree plot pca van Taaltoets alle kinderen, klankarticulatie, 45 items



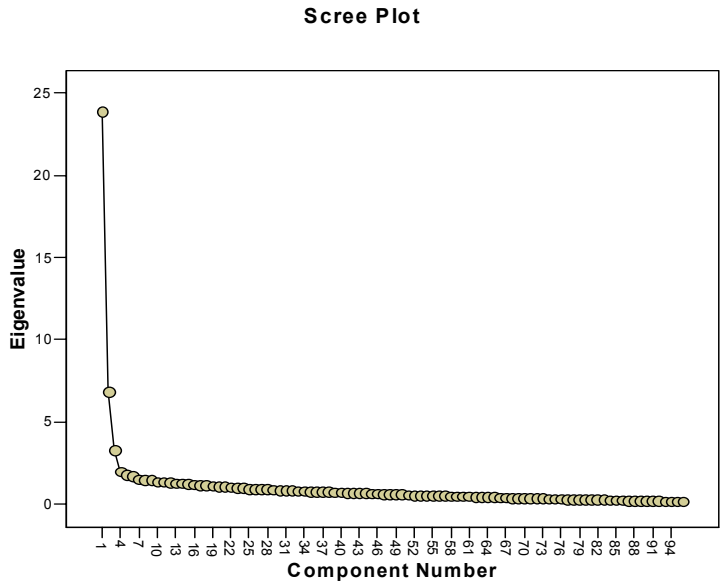
Tabel A6: Taaltoets alle Kinderen, passieve woordenschat, 96 items (N= 352)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
v301a v301a touw	51.69	356.585	.275	.967
v302a v302a baard	51.72	356.063	.250	.967
v303a v303a helikopter	51.68	357.437	.175	.967
v304a v304a uil	51.71	355.443	.353	.967
v305a v305a pleister	51.69	356.151	.325	.967
v306a v306a knie	51.68	356.616	.312	.967
v307a v307a fles	51.70	356.529	.245	.967
v308a v308a schrijven	51.70	356.741	.226	.967
v309a v309a pet	51.68	356.530	.310	.967
v310a v310a strik	51.78	352.771	.457	.967
v311a v311a zagen	51.72	354.246	.442	.967
v312a v312a veter	51.70	356.303	.293	.967
v313a v313a peer	51.70	356.393	.271	.967
v314a v314a raket	51.75	353.219	.474	.967
v315a v315a schelp	51.70	355.551	.395	.967
v316a v316a hamer	51.73	353.877	.455	.967
v317a v317a branden	51.76	354.575	.332	.967
v318a v318a inkt	51.84	350.833	.514	.966
v319a v319a elleboog	51.80	352.222	.465	.967
v320a v320a bijl	51.92	350.871	.450	.967
v321a v321a gieter	51.84	350.577	.541	.966
v322a v322a spier	51.74	354.674	.368	.967
v323a v323a druppelen	51.89	350.527	.488	.966
v324a v324a nest	51.88	349.532	.566	.966
v325a v325a wimper	51.90	351.387	.428	.967
v326a v326a rots	51.84	350.322	.552	.966
v327a v327a vol	51.97	351.799	.374	.967
v328a v328a spons	51.82	350.556	.570	.966
v329a v329a ladder	51.80	350.851	.575	.966
v330a v330a krans	52.03	349.116	.506	.966
v331a v331a schild	51.95	349.884	.491	.966
v332a v332a rem	51.93	350.029	.498	.966
v333a v333a schroeven	51.94	349.515	.520	.966
v334a v334a fontein	51.95	349.014	.547	.966
v335a v335a kneden	51.92	349.224	.551	.966

v336a	v336a envelop	51.93	348.932	.563	.966
v337a	v337a snoer	52.12	349.501	.469	.967
v338a	v338a gewei	52.28	352.654	.309	.967
v339a	v339a steil	52.09	350.622	.411	.967
v340a	v340a strak	52.03	351.247	.387	.967
v341a	v341a balkon	52.10	350.065	.440	.967
v342a	v342a breekbaar	52.00	349.011	.523	.966
v343a	v343a treuren	52.13	349.665	.460	.967
v344a	v344a apparaat	51.84	349.118	.644	.966
v345a	v345a wenkbrauw	51.95	348.028	.606	.966
v346a	v346a rivier	51.91	347.998	.631	.966
v347a	v347a uitgeput	51.94	347.959	.614	.966
v348a	v348a temmen	52.15	350.706	.403	.967
v349a	v349a eetbaar	52.03	348.156	.559	.966
v350a	v350a wijzer	52.04	346.654	.641	.966
v351a	v351a vacht	52.16	347.758	.562	.966
v352a	v352a korst	51.94	346.825	.682	.966
v353a	v353a leuning	52.07	347.360	.593	.966
v354a	v354a breed	52.39	351.048	.439	.967
v355a	v355a heffen	52.19	350.749	.401	.967
v356a	v356a schaven	52.19	347.226	.592	.966
v357a	v357a kudde	52.14	347.345	.585	.966
v358a	v358a wapen	52.18	347.895	.556	.966
v359a	v359a krom	52.13	346.093	.654	.966
v360a	v360a liniaal	52.18	346.645	.624	.966
v361a	v361a overval	52.09	346.849	.617	.966
v362a	v362a spaak	52.33	349.998	.470	.967
v363a	v363a mengen	52.13	346.976	.605	.966
v364a	v364a pilaar	52.33	349.914	.476	.967
v365a	v365a oceaan	52.21	348.258	.538	.966
v366a	v366a overstroming	52.20	346.289	.645	.966
v367a	v367a operatie	52.20	346.602	.627	.966
v368a	v368a lijst	52.46	351.953	.426	.967
v369a	v369a luik	52.27	346.629	.640	.966
v370a	v370a toespraak	52.23	347.772	.567	.966
v371a	v371a vijl	52.31	348.465	.552	.966
v372a	v372a grendel	52.48	352.797	.388	.967
v373a	v373a zool	52.44	349.979	.542	.966

v374a	v374a verplegen	52.28	347.820	.577	.966
v375a	v375a kabel	52.32	348.895	.529	.966
v376a	v376a kade	52.45	351.303	.465	.967
v377a	v377a beeldhouwen	52.52	353.851	.351	.967
v378a	v378a oever	52.51	352.758	.428	.967
v379a	v379a kuit	52.53	353.686	.371	.967
v380a	v380a klokhuis	52.47	350.541	.540	.966
v381a	v381a geraamte	52.51	353.111	.394	.967
v382a	v382a trechter	52.49	351.253	.507	.966
v383a	v383a loket	52.50	352.501	.433	.967
v384a	v384a greppel	52.53	353.372	.399	.967
v385a	v385a grazen	52.52	352.575	.446	.967
v386a	v386a demonstratie	52.46	350.249	.544	.966
v387a	v387a roer	52.41	349.582	.545	.966
v388a	v388a ventiel	52.52	352.815	.428	.967
v389a	v389a manen	52.53	352.626	.455	.967
v390a	v390a scharnier	52.54	353.155	.430	.967
v391a	v391a kajuit	52.49	351.385	.505	.966
v392a	v392a stronk	52.54	352.864	.464	.967
v393a	v393a dar	52.59	355.457	.307	.967
v394a	v394a lectuur	52.61	356.574	.247	.967
v395a	v395a chirurg	52.58	354.751	.367	.967
v396a	v396a hengsel	52.61	356.181	.299	.967

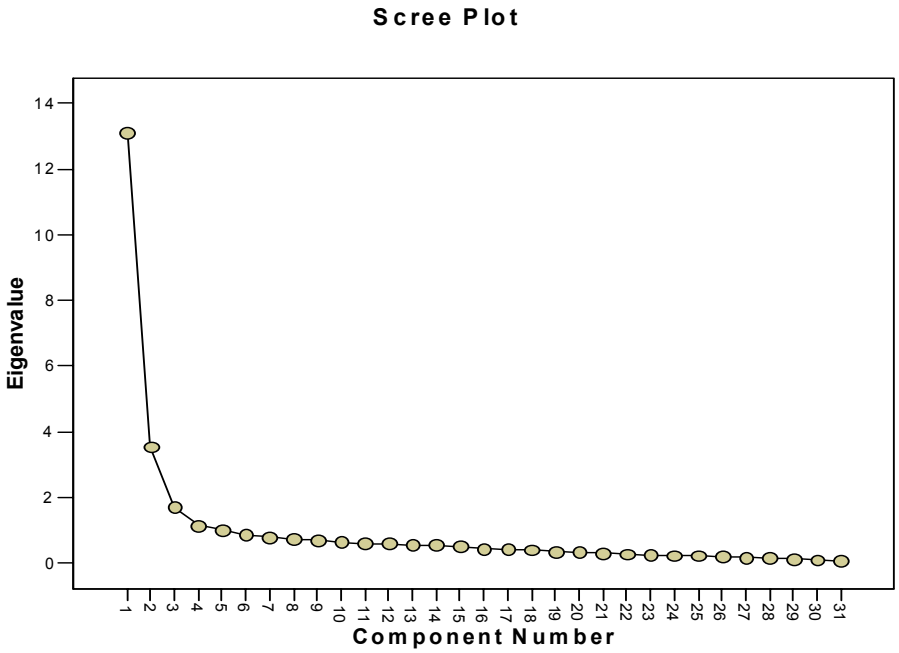
Scree plot pca van Taaltoets alle kinderen, passieve woordenschat, 96 items



Tabel A7: OBIS, passieve woordenschat, 31 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q7PV0_hair	23.51	47.551	.437	.947
q8PV0_eyes	23.48	48.210	.304	.948
q9PV0_mouth	23.48	48.157	.331	.948
q10PV0_shoes	23.47	48.279	.359	.948
q11PV0_arms	23.49	47.866	.449	.947
q12PV0_dog	23.49	47.953	.363	.947
q13PV0_tree	23.53	47.061	.507	.946
q14PV0_bird	23.51	47.434	.450	.947
q16PV1_carrots	23.53	47.119	.479	.947
q17PV1_knife	23.56	46.883	.460	.947
q18PV1_fork	23.59	45.567	.705	.945
q19PV1_cupboard	23.50	47.382	.536	.946
q20PV1_cherries	23.65	45.207	.657	.945
q21PV1_pan	23.58	45.847	.647	.945
q22PV1_bowl	23.80	45.024	.563	.946
q25PV2_butterfly	23.60	45.209	.757	.944
q26PV2_kite	23.62	44.907	.776	.944
q27PV2_castle	23.58	45.242	.806	.944
q28PV2_wasp	23.71	44.096	.792	.943
q29PV2_pigeon	23.71	44.666	.685	.945
q30PV2_windmill	23.67	44.185	.822	.943
q31PV2_turtle	23.65	44.192	.854	.943
q32PV2_violin	23.83	44.582	.620	.946
q33PV2_padlock	23.73	43.970	.793	.943
q34PV2_toadstool	23.68	43.891	.865	.943
q36PV3_yacht	23.88	44.540	.612	.946
q37PV3_cash	23.72	43.798	.830	.943
q38PV3_microscope	24.12	45.903	.420	.948
q39PV3_jewellery	24.07	45.293	.500	.947
q40PV3_saxophone	24.11	45.529	.477	.947
q41PV3_cosmetics	23.99	44.496	.612	.946

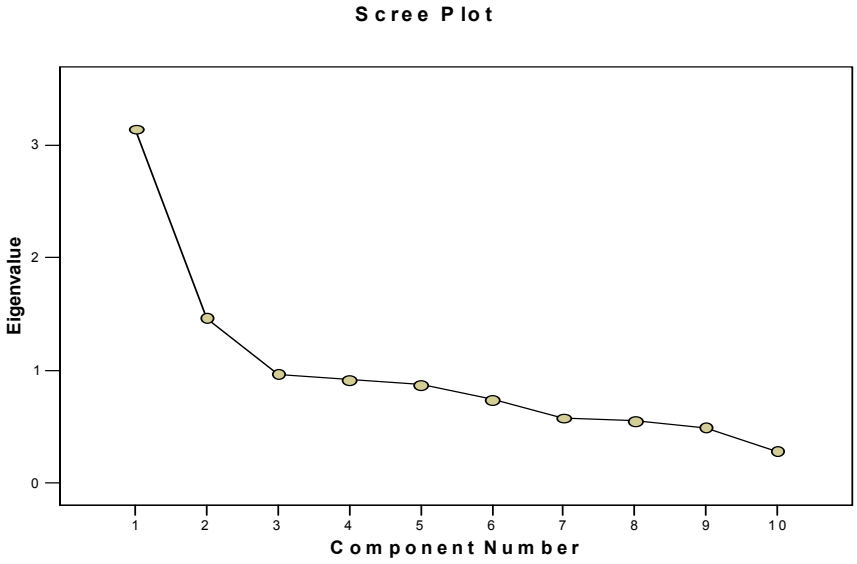
Scree plot pca van OBIS, onderdeel passieve woordenschat, 31 items



Tabel A8: OBIS, beginnende geletterdheid, leesbegrip 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q45IAR1_someone_writing	4.40	4.754	.167	.738
q46IAR1_someone_reading	4.37	4.787	.269	.733
q47IAR1_writing	4.73	4.360	.184	.751
q49IAR2_word	4.70	3.971	.405	.712
q50IAR2_Letter	4.60	3.885	.521	.691
q51IAR2_start_of_story	4.87	3.640	.568	.680
q52IAR2_capital_Letter	5.21	4.545	.211	.736
q53IAR2_full_stop	5.04	3.937	.452	.703
q54IAR2_sentence_start	5.08	3.779	.577	.681
q55IAR2_sentence_finish	5.11	3.819	.587	.681

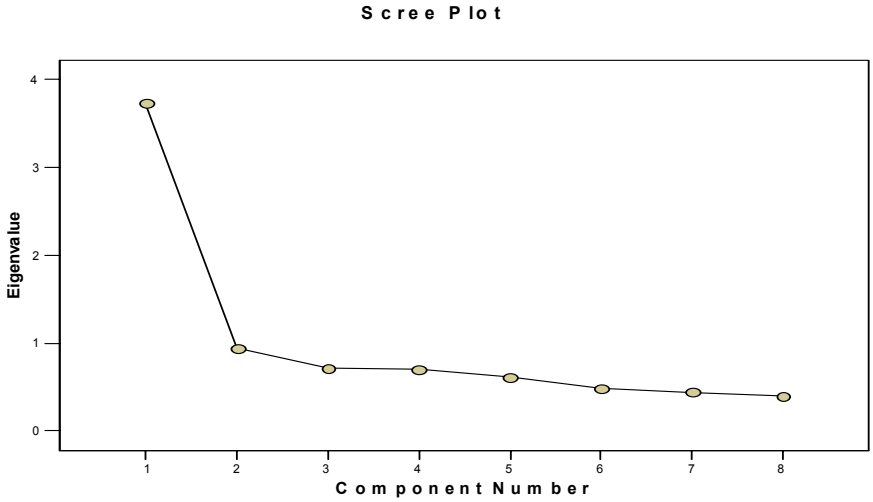
Scree plot pca van OBIS, onderdeel beginnende geletterdheid, leesbegrip 10 items



Tabel A9: OBIS, beginnende geletterdheid, klankarticulatie (onzinwoorden nazeggen) 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q59Rep_stop	5.51	4.070	.240	.840
q60Rep_mantle	5.63	3.400	.549	.811
q61Rep_frigglejang	5.75	3.210	.530	.815
q62Rep_riotous	5.63	3.355	.601	.805
q63Rep_denalty	5.69	3.157	.632	.799
q64Rep_juxtapose	5.64	3.216	.681	.793
q65Rep_enterprising	5.70	3.097	.672	.793
q66Rep_observatory	5.85	3.091	.541	.816

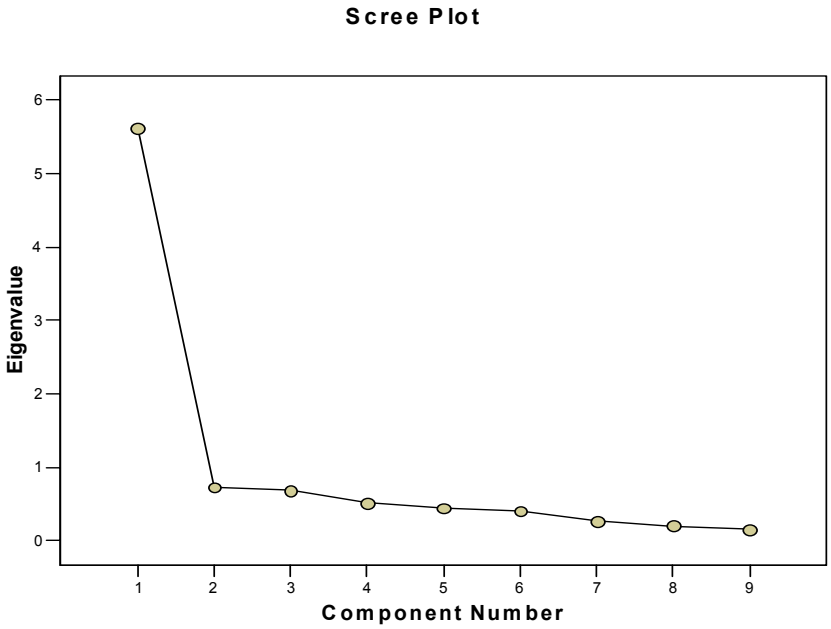
Scree plot pca van OBIS, onderdeel klankarticulatie (onzinwoorden nazeggen) 8 items



Tabel A10: OBIS, beginnende geletterdheid, klankonderscheiding (rijmwoorden herkennen) 9 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q81Rhy_hat	5.61	8.152	.636	.918
q82Rhy_mouse	5.65	8.030	.633	.918
q83Rhy_toes	5.64	8.127	.606	.920
q84Rhy_cherries	5.65	7.848	.713	.913
q85Rhy_sun	5.73	7.486	.796	.908
q86Rhy_bin	5.71	7.395	.855	.904
q87Rhy_drum	5.71	7.409	.855	.904
q88Rhy_dish	5.73	7.393	.838	.905
q89Rhy_pan	5.97	7.874	.567	.924

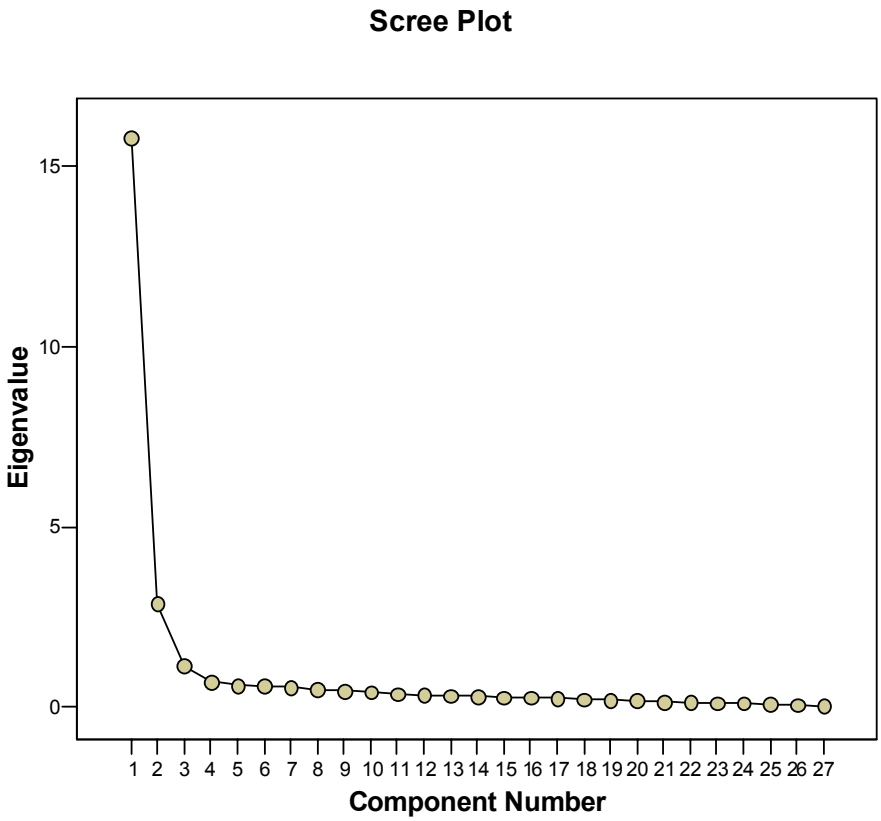
Scree plot pca van OBIS, beginnende geletterdheid, onderdeel klankonderscheiding (rijmwoorden herkennen) 9 items



Tabel A11: OBIS, beginnende geletterdheid, kennis van letters/ het alfabet, 27 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q94Name_Letter	5.94	59.545	.454	.968
q95Lett_S	6.04	59.610	.420	.969
q96Lett_o	6.04	58.398	.586	.967
q97Lett_A	6.21	58.135	.627	.967
q98Lett_c	6.38	58.377	.706	.966
q99Lett_K	6.13	58.009	.632	.967
q100Lett_t	6.37	58.240	.715	.966
q101Lett_B	6.29	57.637	.738	.966
q102Lett_z	6.31	57.213	.817	.965
q103Lett_m	6.32	57.838	.725	.966
q104Lett_x	6.38	57.659	.816	.965
q105Lett_H	6.41	57.817	.836	.965
q106Lett_e	6.40	57.867	.810	.965
q107Lett_r	6.38	57.584	.836	.965
q108Lett_L	6.42	58.203	.784	.965
q109Lett_F	6.44	58.165	.828	.965
q110Lett_w	6.45	58.272	.834	.965
q111Lett_D	6.47	58.599	.806	.965
q112Lett_n	6.46	58.325	.837	.965
q113Lett_V	6.46	58.202	.860	.965
q114Lett_u	6.49	59.160	.772	.966
q115Lett_J	6.51	59.574	.721	.966
q116Lett_Y	6.53	60.366	.631	.967
q117Lett_p	6.48	58.635	.829	.965
q118Lett_G	6.50	59.347	.760	.966
q119Lett_Q	6.55	60.878	.573	.967
q120Lett_i	6.51	59.627	.728	.966

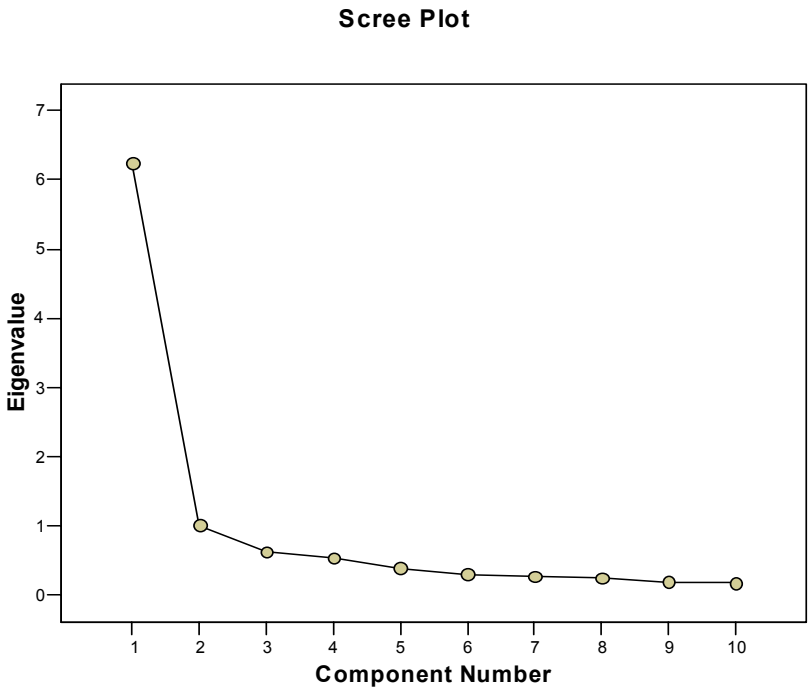
Scree plot pca van OBIS, beginnende geletterdheid, onderdeel kennis van letters/ het alfabet, 27 items



Tabel A12: OBIS, beginnende geletterdheid, leesvaardigheid woorden, 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q179W1_cat	1.97	8.358	.706	.925
q180W1_rabbit	1.97	8.323	.723	.924
q181W1_house	2.03	8.372	.742	.923
q182W1_dog	2.09	8.514	.747	.922
q183W1_duck	2.06	8.306	.800	.920
q185W2_car	2.05	8.149	.856	.916
q186W2_butterfly	2.17	8.834	.729	.924
q187W2_ball	2.19	9.160	.623	.929
q188W2_tree	2.19	8.985	.715	.925
q189W2_flower	2.21	9.086	.693	.926

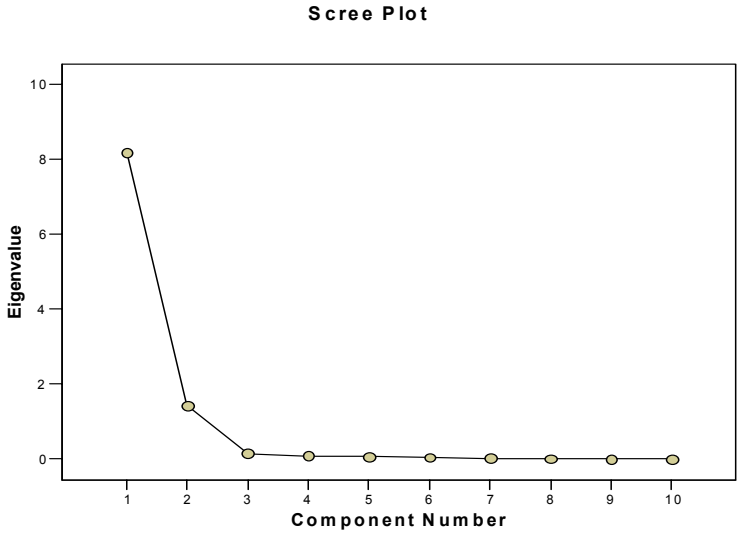
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid woorden, 10 items



Tabel A13: OBIS, beginnende geletterdheid, leesvaardigheid zinnen, 10 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item -Total Correlation	Cronbach's Alfa if Item Deleted
q204Story_Ben_1	2.37	114.001	.819	.939
q205Story_Ben_2	2.30	105.767	.858	.936
q206Story_Ben_3	2.44	113.331	.877	.937
q208Story_Cat_1	2.10	81.063	.952	.952
q209Story_Cat_2	2.46	101.124	.912	.934
q210Story_Cat_3	2.67	121.902	.885	.942
q212Story_Dog	2.66	119.910	.861	.941
q213Story_Boy	2.67	121.969	.865	.942
q214Story_Sheep	2.63	115.354	.859	.938
q215Story_Cat	2.69	123.627	.869	.944

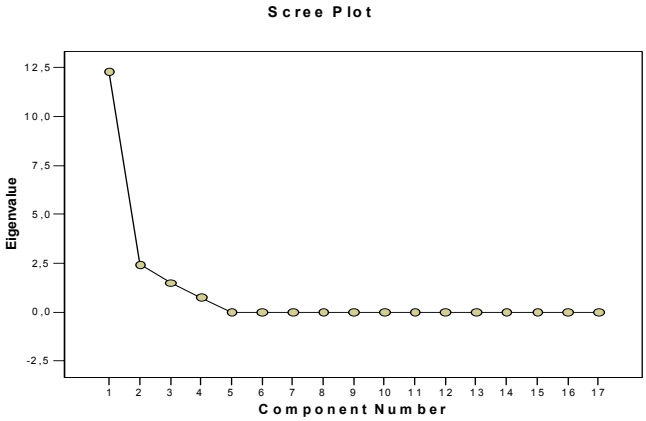
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid zinnen, 10 items



Tabel A14: OBIS, beginnende geletterdheid, leesvaardigheid close meerkeuze, 17 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q219WTS_your	.13	1.505	.585	.977
q220WTS_on	.12	1.367	.964	.972
q221WTS_walk	.12	1.367	.964	.972
q222WTS_two	.12	1.367	.964	.972
q223WTS_leave	.13	1.482	.715	.975
q224WTS_was	.12	1.367	.964	.972
q225WTS_looked	.13	1.416	.885	.973
q226WTS_were	.13	1.482	.715	.975
q227WTS_they	.13	1.416	.885	.973
q228WTS_wearing	.13	1.416	.885	.973
q229WTS_thought	.13	1.416	.885	.973
q230WTS_they2	.13	1.416	.885	.973
q231WTS_for	.13	1.416	.885	.973
q232WTS_buy	.13	1.465	.815	.974
q233WTS_at	.13	1.531	.650	.976
q234WTS_to	.13	1.531	.650	.976
q235WTS_carried	.13	1.482	.715	.975

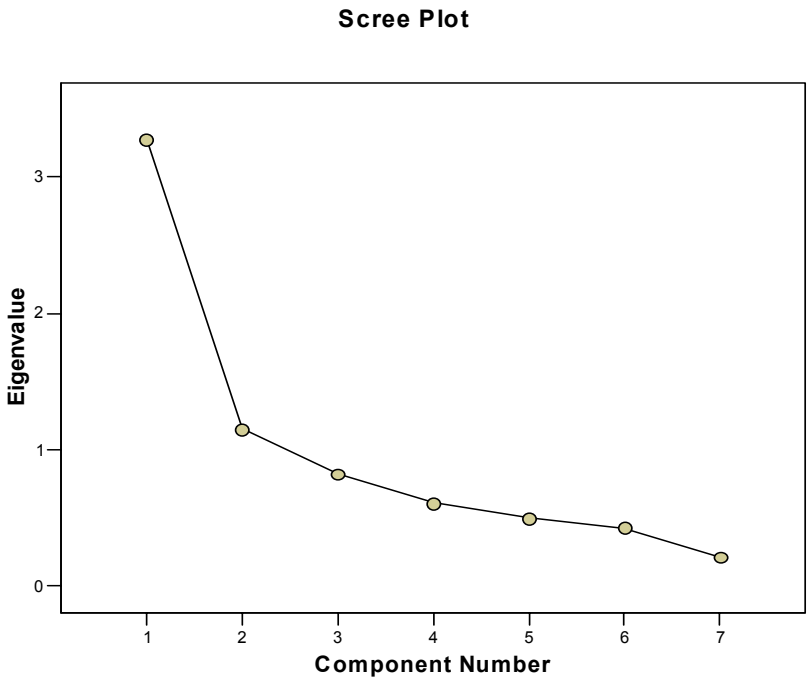
Scree plot pca van OBIS, beginnende geletterdheid, leesvaardigheid close meerkeuze, 17 items



Tabel A15: OBIS, aanvankelijk rekenen, rekenbegrip (kennis rekenjargon), 7 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q258IAM_biggest_cat	5.59	.866	.391	.707
q259IAM_smallest_cat	5.59	.872	.418	.706
q260IAM_more_balloons	5.60	.806	.512	.682
q261IAM_most_water	5.71	.579	.526	.673
q262IAM_least_water	5.76	.576	.423	.733
q263IAM_tallest	5.60	.770	.612	.663
q264IAM_shortest	5.61	.739	.616	.653

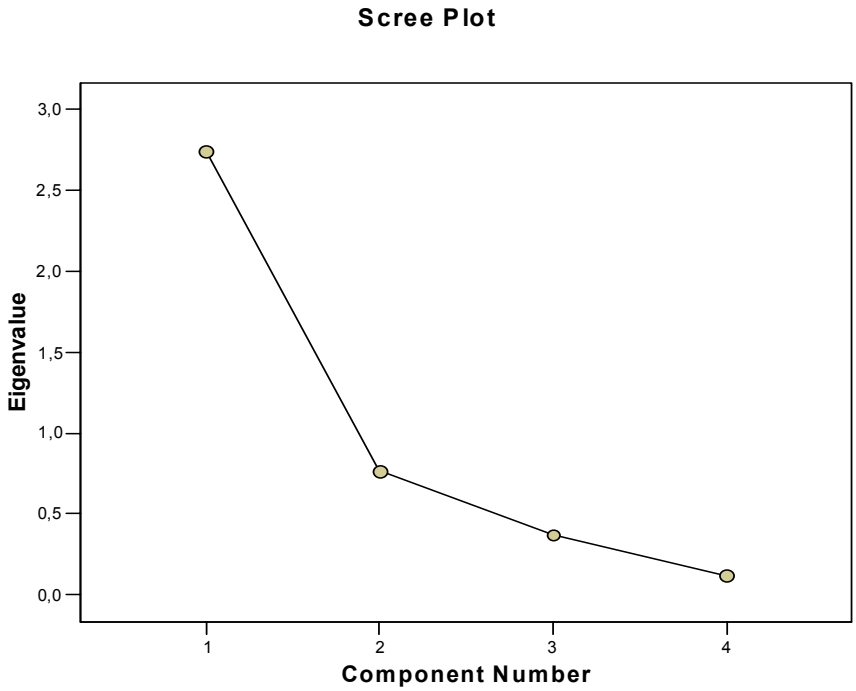
Scree plot pca van OBIS, aanvankelijk rekenen, rekenbegrip (kennis rekenjargon), 7 items



Tabel A16: OBIS, aanvankelijk rekenen, tellen (geheugen voor aantallen), 4 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q267Count_apples_here	2.37	1.080	.612	.828
q268Count_apples_there	2.43	.934	.651	.799
q269Count_fish_here	2.51	.769	.735	.759
q270Count_fish_there	2.60	.684	.753	.758

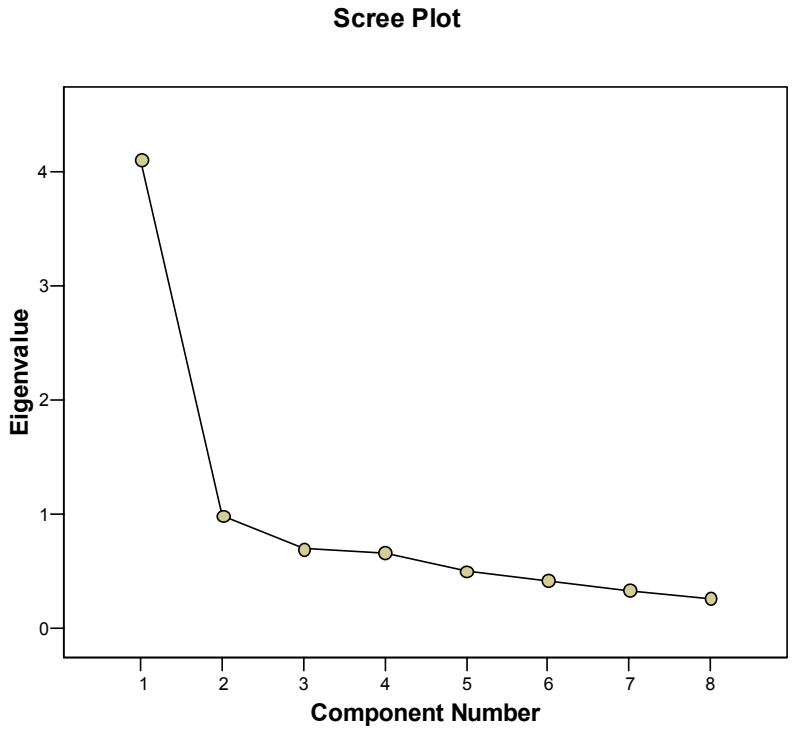
Scree plot pca van OBIS, aanvankelijk rekenen, tellen (geheugen voor aantallen), 4 items



Tabel A17: OBIS, aanvankelijk rekenen, sommen A, 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q272Sums_A_31_balls	4.64	5.135	.506	.849
q273Sums_A_41_cars	4.67	4.863	.626	.837
q274Sums_A_63_ice_cream	4.87	4.664	.498	.850
q275Sums_A_21_rabbits	4.82	4.424	.676	.828
q276Sums_A_31_bikes	4.79	4.364	.744	.820
q277Sums_A_22_puppies	4.85	4.315	.712	.823
q278Sums_A_32_pips	4.99	4.350	.621	.835
q279Sums_A_43_rockets	5.20	4.690	.469	.854

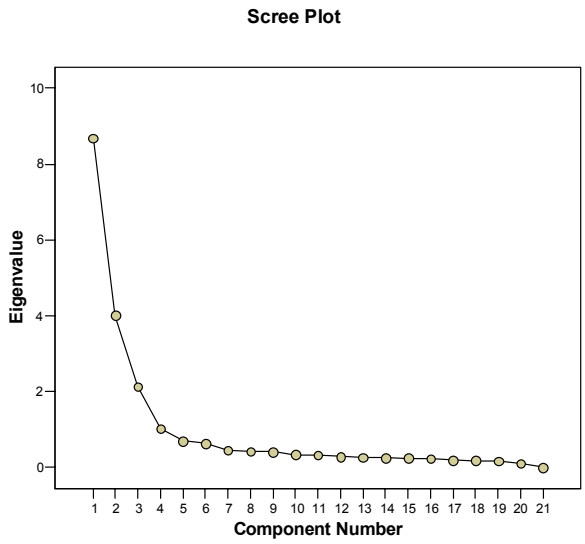
Scree plot pca van OBIS, aanvankelijk rekenen, rekenen sommen A, 8 items



Tabel A18: OBIS, aanvankelijk rekenen, getallenkennis receptief, 21 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q282Num_4	8.29	22.353	.701	.922
q283Num_1	8.18	23.526	.538	.925
q284Num_3	8.23	22.807	.668	.923
q285Num_2	8.23	22.841	.663	.923
q286Num_5	8.28	22.260	.742	.921
q287Num_7	8.38	21.589	.805	.919
q288Num_6	8.41	21.577	.789	.920
q289Num_9	8.46	21.538	.770	.920
q290Num_8	8.42	21.440	.813	.919
q291Num_0	8.38	21.508	.823	.919
q292Num_teen1	8.68	21.950	.671	.923
q293Num_teen2	8.72	22.035	.665	.923
q294Num_teen3	8.72	22.076	.655	.923
q295Num_2dig1	8.93	23.573	.466	.926
q296Num_2dig2	8.92	23.396	.501	.926
q297Num_2dig3	8.93	23.450	.494	.926
q298Num_3dig1	9.06	24.933	.239	.929
q299Num_3dig2	9.07	25.130	.167	.929
q300Num_3dig3	9.06	25.037	.199	.929
q301Num_3dig4	9.06	24.965	.237	.929
q302Num_3dig5	9.06	24.965	.237	.929

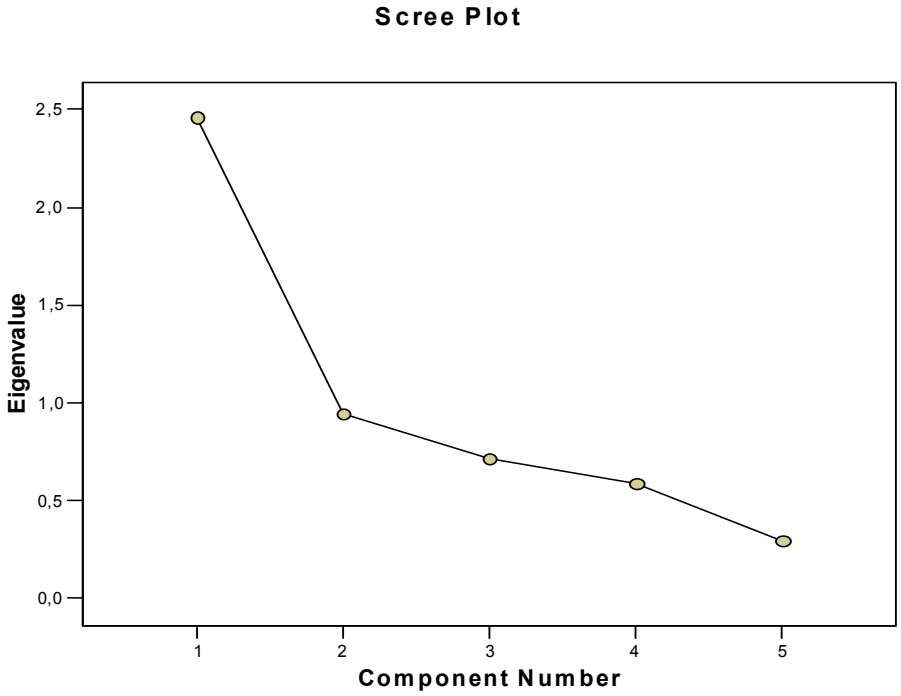
Scree plot pca van OBIS, aanvankelijk rekenen, getallenkennis receptief, 21 items



Tabel A19: OBIS, aanvankelijk rekenen, kennis vormen receptief, 5 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q321Sh_star	3.22	1.451	.201	.767
q322Sh_circle	3.36	1.042	.477	.692
q323Sh_square	3.39	.926	.608	.638
q324Sh_triangle	3.38	.884	.696	.599
q325Sh_hexagon	3.50	.869	.524	.684

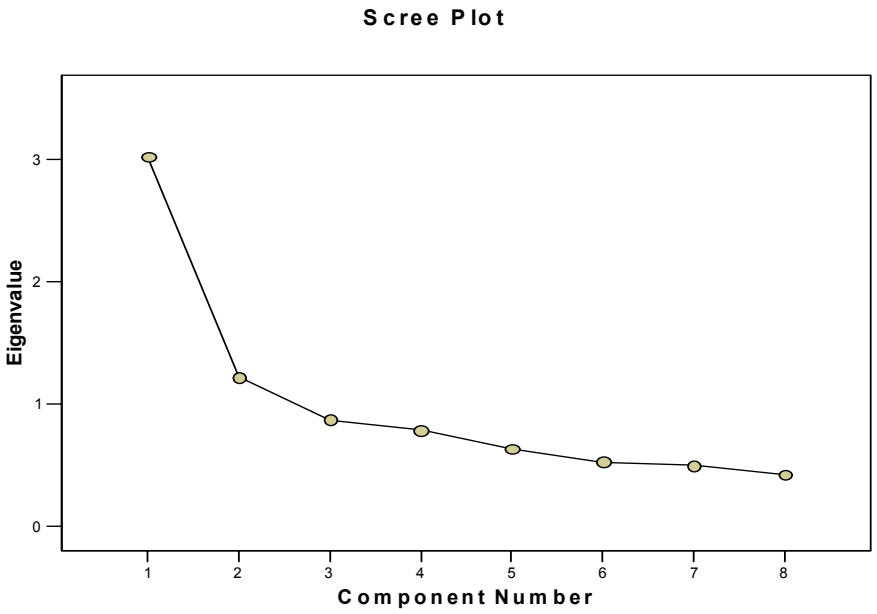
Scree plot pca van OBIS, aanvankelijk rekenen, kennis vormen receptief, 5 items



Tabel A20: OBIS, aanvankelijk rekenen, hoofdrekenen, 8 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alfa if Item Deleted
q328M_1_more_than_5	1.11	2.083	.492	.720
q329M_3_less_than_7	1.36	1.928	.609	.689
q330M_2_more_than_6	1.43	1.989	.592	.693
q331M_3_more_than_8	1.56	2.154	.583	.696
q332M_6_less_than_15	1.66	2.483	.450	.725
q333M_8_more_than_13	1.73	2.838	.265	.753
q334M_10_less_than_25	1.69	2.587	.442	.730
q335M_21_more_than_32	1.75	2.940	.190	.761

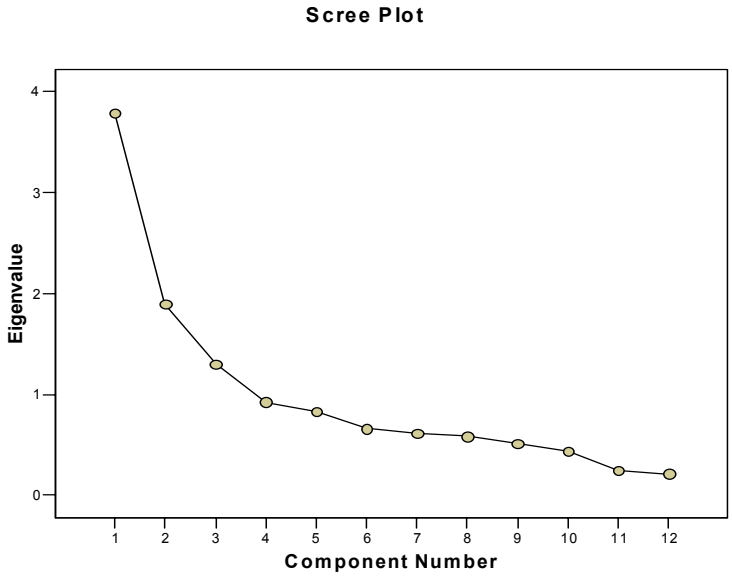
Scree plot pca van OBIS, aanvankelijk rekenen, hoofdrekenen, 8 items



Tabel A21: OBIS, aanvankelijk rekenen, sommen B, 12 items (N= 344)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cron-bach's Alfa if Item Deleted
q338Sums_B_three_quarter_circle	1.53	2.454	.307	.715
q339Sums_B_5p_apple_coins	1.38	2.354	.456	.676
q340Sums_B_half_the_bikes	1.82	2.451	.360	.698
q341Sums_B_7+3=	1.96	2.485	.484	.670
q342Sums_B_sequence_10,20,30,40,?	1.96	2.467	.501	.667
q343Sums_B_sequence_2,4,6,8,?,12	2.03	2.710	.441	.681
q344Sums_B_half_of_6	2.06	2.824	.415	.688
q345Sums_B_4+11=	2.09	2.934	.420	.695
q346Sums_B_9-6=	2.10	2.979	.392	.699
q347Sums_B_12p_orange_coin	2.09	2.959	.400	.697
q348Sums_B_15-4=	2.11	3.093	.278	.709
q349Sums_B_15+21=	2.11	3.131	.211	.713

Scree plot pca van OBIS, aanvankelijk rekenen, sommen B, 12 items



APPENDIX
KRUISTABELLEN ONDERSTE KWARTIEL PER SUBTOETS, ONGECORRIGEERD VOOR LEEFTIJD. (1=onderste kwartiel; 2=bovenste 75%)

OBIS-schrijven + TvK

R=.27 Kappa=.27		TvK		
		1,00	2,00	Total
OBIS schrijven	1,00	7	10	17
	2,00	9	51	60
Total		16	61	77

OBIS-schrijven + peabody

R=.39 Kappa=.39		Peabody		
		1,00	2,00	Total
OBIS schrijven	1,00	15	13	28
	2,00	9	46	55
Total		24	59	83

OBIS-schrijven + TAK-klankonderscheiding

R=.27 Kappa=.27		TAK-klankonderscheiding		
		1,00	2,00	Total
OBIS schrijven	1,00	8	14	22
	2,00	8	60	68
Total		16	74	90

OBIS-schrijven + TAK-klankarticulatie

R=Kappa=.45		TAK-klankarticulatie		
		1,00	2,00	Total
OBIS schrijven	1,00	11	11	22
	2,00	6	62	68
Total		17	73	90

OBIS-schrijven + TAK-passieve woordkennis

R= .40 Kappa=.38		TAK-passieve woordkennis 1,00	2,00	Total
OBIS schrijven	1,00	9	13	22
	2,00	5	63	68
Total		14	76	90

OBIS-schrijven + OBIS-passieve woordkennis

R=Kappa=.43		OBIS-passieve woordkennis 1,00	2,00	Total
OBIS schrijven	1,00	49	38	87
	2,00	36	221	257
Total		85	259	344

OBIS-schrijven +OBIS-leesbegrip

R=Kappa= .33		OBIS-leesbegrip 1,00	2,00	Total
OBIS schrijven	1,00	39	48	87
	2,00	35	222	257
Total		74	270	344

OBIS-schrijven +OBIS-klankarticulatie

R=Kappa= .42		OBIS-klankarticulatie 1,00	2,00	Total
OBIS schrijven	1,00	43	44	87
	2,00	27	230	257
Total		70	274	344

OBIS-schrijven +OBIS-klankonderscheiding

R=Kappa= .47		OBIS-klankonderscheiding 1,00	2,00	Total
OBIS schrijven	1,00	53	34	87
	2,00	35	222	257
Total		88	256	344

OBIS-schrijven +OBIS-letterkennis

R=Kappa=.51		OBIS-letterkennis 1,00	2,00	Total
OBIS schrijven	1,00	51	36	87
	2,00	25	232	257
Total		76	268	344

OBIS-schrijven +OBIS-rekenbegrip

R=Kappa=.17		OBIS-rekenbegrip 1,00	2,00	Total
OBIS schrijven	1,00	31	56	87
	2,00	49	208	257
Total		80	264	344

OBIS-schrijven +OBIS-geheugen

R=Kappa= .36		OBIS-geheugen 1,00	2,00	Total
OBIS schrijven	1,00	45	42	87
	2,00	40	217	257
Total		85	259	344

OBIS-schrijven +OBIS-sommen A

R=Kappa=.53		OBIS-sommen A 1,00	2,00	Total
OBIS schrijven	1,00	59	28	87
	2,00	34	223	257
Total		93	251	344

OBIS-schrijven +OBIS-getallenkennis

R=Kappa=.53		OBIS-getallenkennis 1,00	2,00	Total
OBIS schrijven	1,00	59	28	87
	2,00	34	223	257
Total		93	251	344

OBIS-schrijven +OBIS-vormenkennis

R=Kappa= .34		OBIS-vormenkennis 1,00	2,00	Total
OBIS schrijven	1,00	39	48	87
	2,00	33	224	257
Total		72	272	344

OBIS-schrijven +OBIS-sommen B

R=Kappa= .54		OBIS-sommen B 1,00	2,00	Total
OBIS schrijven	1,00	52	35	87
	2,00	22	235	257
Total		74	270	344

TvK + Peabody

R=Kappa= .68		Peabody 1,00	Peabody 2,00	Total
TvK	1,00	18	4	22
	2,00	7	58	65
Total		25	62	87

TvK + TAK-klankonderscheiding

R=Kappa= .37		TAK-klankonderscheiding 1,00	TAK-klankonderscheiding 2,00	Total
TvK	1,00	10	12	22
	2,00	7	55	62
Total		17	67	84

TvK + TAK-klankarticulatie

R=Kappa= .22		TAK-klankarticulatie 1,00	TAK-klankarticulatie 2,00	Total
TvK	1,00	8	14	22
	2,00	10	52	62
Total		18	66	84

TvK + TAK-passieve woordkennis

R=Kappa= .56		TAK-passieve woordkennis 1,00	TAK-passieve woordkennis 2,00	Total
TvK	1,00	14	8	22
	2,00	6	56	62
Total		20	64	84

TvK + OBIS-passieve woordkennis

R=Kappa=.53 .52		OBIS-passieve woordkennis 1,00	2,00	Total
TvK	1,00	12	4	16
	2,00	10	51	61
Total		22	55	77

TvK + OBIS-leesbegrip

R=Kappa=.31		OBIS-leesbegrip 1,00	2,00	Total
TvK	1,00	7	9	16
	2,00	8	53	61
Total		15	62	77

TvK + OBIS-klankarticulatie

R=Kappa=.45		OBIS-klankarticulatie 1,00	2,00	Total
TvK	1,00	8	8	16
	2,00	5	56	61
Total		13	64	77

TvK + OBIS-klankonderscheiding

R=Kappa=.52		OBIS-klankonderscheiding 1,00	2,00	Total
TvK	1,00	12	4	16
	2,00	10	51	61
Total		22	55	77

TvK + OBIS-letterkennis

R=Kappa=.21		OBIS-letterkennis 1,00	2,00	Total
TvK	1,00	6	10	16
	2,00	10	51	61
Total		16	61	77

TvK + OBIS-rekenbegrip

R=Kappa=.17		OBIS-rekenbegrip 1,00	2,00	Total
TvK	1,00	6	10	16
	2,00	12	49	61
Total		18	59	77

TvK + OBIS-geheugen

R=Kappa=.17		OBIS-geheugen 1,00	2,00	Total
TvK	1,00	5	11	16
	2,00	9	52	61
Total		14	63	77

TvK + OBIS-sommen A

R=Kappa=.35		OBIS-sommen A 1,00	2,00	Total
TvK	1,00	8	8	16
	2,00	9	52	61
Total		17	60	77

TvK + OBIS-getallenkennis

R=Kappa=.17		OBIS-getallenkennis 1,00	2,00	Total
TvK	1,00	6	10	16
	2,00	12	49	61
Total		18	59	77

TvK + OBIS-vormenkennis

R=Kappa=.28		OBIS-vormenkennis 1,00	2,00	Total
TvK	1,00	8	8	16
	2,00	12	49	61
Total		20	57	77

TvK + OBIS-sommen B

R=Kappa=.25 .24		OBIS-sommen B 1,00	2,00	Total
TvK	1,00	5	11	16
	2,00	6	55	61
Total		11	66	77

Peabody + TAK-klankonderscheiding

R= .42 Kappa=.41		TAK-klankonderscheiding 1,00	2,00	Total
Peabody	1,00	11	12	23
	2,00	6	53	59
Total		17	65	82

Peabody + TAK-klankarticulatie

R=.48 Kappa=.46		TAK-klankarticulatie 1,00	2,00	Total
Peabody	1,00	11	12	23
	2,00	4	55	59
Total		15	67	82

Peabody + TAK-passieve woordkennis

R= .62 Kappa=.61		TAK-passieve woordkennis 1,00	2,00	Total
Peabody	1,00	14	9	23
	2,00	3	56	59
Total		17	65	82

Peabody + OBIS-passieve woordkennis

R=Kappa= .77		OBIS-passieve woordkennis 1,00	2,00	Total
Peabody	1,00	21	3	24
	2,00	5	54	59
Total		26	57	83

Peabody + OBIS-leesbegrip

R=Kappa=.58		OBIS-leesbegrip 1,00	2,00	Total
Peabody	1,00	18	6	24
	2,00	9	50	59
Total		27	56	83

Peabody + OBIS-klankarticulatie

R=Kappa=.33		OBIS-klankarticulatie 1,00	2,00	Total
Peabody	1,00	13	11	24
	2,00	12	47	59
Total		25	58	83

Peabody + OBIS-klankonderscheiding

R=Kappa=.71		OBIS-klankonderscheiding 1,00	2,00	Total
Peabody	1,00	19	5	24
	2,00	5	54	59
Total		24	59	83

Peabody + OBIS-letterkennis

R=Kappa=.62		OBIS-letterkennis 1,00	2,00	Total
Peabody	1,00	18	6	24
	2,00	7	52	59
Total		25	58	83

Peabody + OBIS-rekenbegrip

R=Kappa=.38		OBIS-rekenbegrip 1,00	2,00	Total
Peabody	1,00	13	11	24
	2,00	10	49	59
Total		23	60	83

Peabody + OBIS-geheugen

R=Kappa=.45		OBIS-geheugen 1,00	2,00	Total
Peabody	1,00	15	9	24
	2,00	10	49	59
Total		25	58	83

Peabody + OBIS-sommen A

R=Kappa=.69		OBIS-sommen A 1,00	2,00	Total
Peabody	1,00	20	4	24
	2,00	7	52	59
Total		27	56	83

Peabody + OBIS-getallenkennis

R=Kappa=.45		OBIS-getallenkennis 1,00	2,00	Total
Peabody	1,00	15	9	24
	2,00	10	49	59
Total		25	58	83

Peabody + OBIS-vormenkennis

R =.49 Kappa=.47		OBIS-vormenkennis 1,00	2,00	Total
Peabody	1,00	11	13	24
	2,00	3	56	59
Total		14	69	83

Peabody + OBIS-sommen B

R=Kappa=.74		OBIS-sommen B		
		1,00	2,00	Total
Peabody	1,00	20	4	24
	2,00	5	54	59
Total		25	58	83

TAK-klankonderscheiding + TAK-klankarticulatie

R=Kappa= .37		TAK-klankarticulatie		
		1,00	2,00	Total
TAK-klankonderscheiding	1,00	47	46	93
	2,00	38	221	259
Total		85	267	352

TAK-klankonderscheiding + TAK-passieve woordkennis

R=Kappa= .49		TAK-passieve woordkennis		
		1,00	2,00	Total
TAK-klankonderscheiding	1,00	56	37	93
	2,00	32	227	259
Total		88	264	352

TAK-klankonderscheiding + OBIS-passieve woordkennis

R=Kappa= .22		OBIS-passieve woordkennis		
		1,00	2,00	Total
TAK-klankonderscheiding	1,00	6	10	16
	2,00	11	63	74
Total		17	73	90

TAK-klankonderscheiding + OBIS-leesbegrip

R=Kappa=.26		OBIS-leesbegrip 1,00	2,00	Total
TAK-klankonderscheiding	1,00	6	10	16
	2,00	9	65	74
Total		15	75	90

TAK-klankonderscheiding + OBIS-klankarticulatie

R=Kappa=.34		OBIS-klankarticulatie 1,00	2,00	Total
TAK-klankonderscheiding	1,00	7	9	16
	2,00	8	66	74
Total		15	75	90

TAK-klankonderscheiding + OBIS-klankonderscheiding

R=Kappa= .34		OBIS-klankonderscheiding 1,00	2,00	Total
TAK-klankonderscheiding	1,00	9	7	16
	2,00	13	61	74
Total		22	68	90

TAK-klankonderscheiding + OBIS-letterkennis

R=Kappa= .20		OBIS-letterkennis 1,00	2,00	Total
TAK-klankonderscheiding	1,00	5	11	16
	2,00	9	65	74
Total		14	76	90

TAK-klankonderscheiding + OBIS-rekenbegrip

R=Kappa= .13		OBIS-rekenbegrip 1,00	2,00	Total
TAK-klankonderscheiding	1,00	5	11	16
	2,00	13	61	74
Total		18	72	90

TAK-klankonderscheiding + OBIS-geheugen

R= .23 Kappa=.22		OBIS-geheugen 1,00	2,00	Total
TAK-klankonderscheiding	1,00	8	8	16
	2,00	17	57	74
Total		25	65	90

TAK-klankonderscheiding + OBIS-sommen A

R= .57 Kappa=.56		OBIS-sommen A 1,00	2,00	Total
TAK-klankonderscheiding	1,00	12	4	16
	2,00	9	65	74
Total		21	69	90

TAK-klankonderscheiding + OBIS-getallenkennis

R= .43 Kappa=.42		OBIS-getallenkennis 1,00	2,00	Total
TAK-klankonderscheiding	1,00	10	6	16
	2,00	11	63	74
Total		21	69	90

TAK-klankonderscheiding + OBIS-vormenkennis

R=Kappa= .54		OBIS-vormenkennis 1,00	2,00	Total
TAK-klankonderscheiding	1,00	11	5	16
	2,00	8	66	74
Total		19	71	90

TAK-klankonderscheiding + OBIS-sommen B

R=Kappa=.36		OBIS-sommen B 1,00	2,00	Total
TAK-klankonderscheiding	1,00	7	9	16
	2,00	7	67	74
Total		14	76	90

TAK-klankarticulatie + TAK-passieve woordkennis

R=Kappa=.36		TAK-passieve woordkennis 1,00	2,00	Total
TAK-klankarticulatie	1,00	45	40	85
	2,00	43	224	267
Total		88	264	352

TAK-klankarticulatie + OBIS-passieve woordkennis

R=Kappa=.28		OBIS-passieve woordkennis 1,00	2,00	Total
TAK-klankarticulatie	1,00	7	10	17
	2,00	10	63	73
Total		17	73	90

TAK-klankarticulatie + OBIS-leesbegrip

R=Kappa=.17		OBIS-leesbegrip 1,00	2,00	Total
TAK-klankarticulatie	1,00	5	12	17
	2,00	10	63	73
Total		15	75	90

TAK-klankarticulatie + OBIS-klankarticulatie

R=Kappa=.47		OBIS-klankarticulatie 1,00	2,00	Total
TAK-klankarticulatie	1,00	9	8	17
	2,00	6	67	73
Total		15	75	90

TAK-klankarticulatie + OBIS-klankonderscheiding

R=Kappa=.32		OBIS-klankonderscheiding 1,00	2,00	Total
TAK-klankarticulatie	1,00	9	8	17
	2,00	13	60	73
Total		22	68	90

TAK-klankarticulatie + OBIS-letterkennis

R=Kappa=.26		OBIS-letterkennis 1,00	2,00	Total
TAK-klankarticulatie	1,00	6	11	17
	2,00	8	65	73
Total		14	76	90

TAK-klankarticulatie + OBIS-rekenbegrip

R=Kappa=.04		OBIS-rekenbegrip 1,00	2,00	Total
TAK-klankarticulatie	1,00	4	13	17
	2,00	14	59	73
Total		18	72	90

TAK-klankarticulatie + OBIS-geheugen

R =.33 Kappa=.32		OBIS-geheugen 1,00	2,00	Total
TAK-klankarticulatie	1,00	10	7	17
	2,00	15	58	73
Total		25	65	90

TAK-klankarticulatie + OBIS-sommen A

R=Kappa= .47		OBIS-sommen A 1,00	2,00	Total
TAK-klankarticulatie	1,00	11	6	17
	2,00	10	63	73
Total		21	69	90

TAK-klankarticulatie + OBIS-getallenkennis

R =.54 Kappa=.53		OBIS-getallenkennis 1,00	2,00	Total
TAK-klankarticulatie	1,00	12	5	17
	2,00	9	64	73
Total		21	69	90

TAK-klankarticulatie + OBIS-vormenkennis

R=Kappa= .38		OBIS-vormenkennis 1,00	2,00	Total
TAK-klankarticulatie	1,00	9	8	17
	2,00	10	63	73
Total		19	71	90

TAK-klankarticulatie + OBIS-sommen B

R=.50 Kappa=.49		OBIS-sommen B 1,00	2,00	Total
TAK-klankarticulatie	1,00	9	8	17
	2,00	5	68	73
Total		14	76	90

TAK-passieve woordenschat + OBIS-passieve woordkennis

R =.50 Kappa=.49		OBIS-passieve woordkennis 1,00	2,00	Total
TAK-passieve woordenschat	1,00	9	5	14
	2,00	8	68	76
Total		17	73	90

TAK-passieve woordenschat + OBIS-leesbegrip

R=Kappa= .47		OBIS-leesbegrip 1,00	2,00	Total
TAK-passieve woordenschat	1,00	8	6	14
	2,00	7	69	76
Total		15	75	90

TAK-passieve woordenschat + OBIS-klankarticulatie

R=Kappa= .38		OBIS-klankarticulatie 1,00	2,00	Total
TAK-passieve woordenschat	1,00	7	7	14
	2,00	8	68	76
Total		15	75	90

TAK-passieve woordenschat + OBIS-klankonderscheiding

R=Kappa=.18		OBIS-klankonderscheiding 1,00	2,00	Total
TAK-passieve woordenschat	1,00	6	8	14
	2,00	16	60	76
Total		22	68	90

TAK-passieve woordenschat + OBIS-letterkennis

R=Kappa=.32		OBIS-letterkennis 1,00	2,00	Total
TAK-passieve woordenschat	1,00	6	8	14
	2,00	8	68	76
Total		14	76	90

TAK-passieve woordenschat + OBIS-rekenbegrip

R=Kappa=-.06		OBIS-rekenbegrip 1,00	2,00	Total
TAK-passieve woordenschat	1,00	2	12	14
	2,00	16	60	76
Total		18	72	90

TAK-passieve woordenschat + OBIS-geheugen

R =.35 Kappa=.33		OBIS-geheugen 1,00	2,00	Total
TAK-passieve woordenschat	1,00	9	5	14
	2,00	16	60	76
Total		25	65	90

TAK-passieve woordenschat + OBIS-sommen A

R= .42 Kappa=.40		OBIS-sommen A 1,00	2,00	Total
TAK-passieve woordenschat	1,00	9	5	14
	2,00	12	64	76
Total		21	69	90

TAK-passieve woordenschat + OBIS-getallenkennis

R= .42 Kappa=.40		OBIS-getallenkennis 1,00	2,00	Total
TAK-passieve woordenschat	1,00	9	5	14
	2,00	12	64	76
Total		21	69	90

TAK-passieve woordenschat + OBIS-vormenkennis

R=Kappa= .23		OBIS-vormenkennis 1,00	2,00	Total
TAK-passieve woordenschat	1,00	6	8	14
	2,00	13	63	76
Total		19	71	90

TAK-passieve woordenschat + OBIS-sommen B

R=Kappa= .49		Obis-sommen B 1,00	Obis-sommen B 2,00	Total
TAK-passieve woordenschat	1,00	8	6	14
	2,00	6	70	76
Total		14	76	90

OBIS-passieve woordenschat + OBIS-leesbegrip

R=Kappa= .49		OBIS-leesbegrip 1,00	OBIS-leesbegrip 2,00	Total
OBIS-passieve woordenschat	1,00	48	37	85
	2,00	26	233	259
Total		74	270	344

OBIS-passieve woordenschat + OBIS-klankarticulatie

R=Kappa= .43		OBIS-klankarticulatie 1,00	OBIS-klankarticulatie 2,00	Total
OBIS-passieve woordenschat	1,00	43	42	85
	2,00	27	232	259
Total		70	274	344

OBIS-passieve woordenschat + OBIS-klankonderscheiding

R=Kappa= .53		OBIS-klankonderscheiding 1,00	OBIS-klankonderscheiding 2,00	Total
OBIS-passieve woordenschat	1,00	56	29	85
	2,00	32	227	259
Total		88	256	344

OBIS-passieve woordenschat + OBIS-letterkennis

R=Kappa= .43		OBIS-letterkennis 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	45	40	85
	2,00	31	228	259
Total		76	268	344

OBIS-passieve woordenschat + OBIS-rekenbegrip

R=Kappa= .32		OBIS-rekenbegrip 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	40	45	85
	2,00	40	219	259
Total		80	264	344

OBIS-passieve woordenschat + OBIS-geheugen

R=Kappa= .36		OBIS-geheugen 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	44	41	85
	2,00	41	218	259
Total		85	259	344

OBIS-passieve woordenschat + OBIS-sommen A

R=Kappa= .53		OBIS-sommen A 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	58	27	85
	2,00	35	224	259
Total		93	251	344

OBIS-passieve woordenschat + OBIS-getallen

R=Kappa= .38		OBIS-getallen 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	48	37	85
	2,00	45	214	259
Total		93	251	344

OBIS-passieve woordenschat + OBIS-vormenkennis

R= .39 Kappa=.38		OBIS-vormenkennis 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	41	44	85
	2,00	31	228	259
Total		72	272	344

OBIS-passieve woordenschat + OBIS-sommen B

R=Kappa= .57		OBIS-sommen B 1,00	2,00	Total
OBIS-passieve woordenschat	1,00	53	32	85
	2,00	21	238	259
Total		74	270	344

OBIS-leesbegrip + OBIS-klankarticulatie

R=Kappa= .32		OBIS-klankarticulatie 1,00	2,00	Total
OBIS-leesbegrip	1,00	33	41	74
	2,00	37	233	270
Total		70	274	344

OBIS-leesbegrip + OBIS-klankonderscheiding

R=Kappa= .42		OBIS-klankonderscheiding 1,00	2,00	Total
OBIS-leesbegrip	1,00	45	29	74
	2,00	43	227	270
Total		88	256	344

OBIS-leesbegrip + OBIS-letterkennis

R=Kappa= .44		OBIS-letterkennis 1,00	2,00	Total
OBIS-leesbegrip	1,00	42	32	74
	2,00	34	236	270
Total		76	268	344

OBIS-leesbegrip + OBIS-rekenbegrip

R=Kappa= .23		OBIS-rekenbegrip 1,00	2,00	Total
OBIS-leesbegrip	1,00	31	43	74
	2,00	49	221	270
Total		80	264	344

OBIS-leesbegrip + OBIS-geheugen

R=Kappa= .39		OBIS-geheugen 1,00	2,00	Total
OBIS-leesbegrip	1,00	42	32	74
	2,00	43	227	270
Total		85	259	344

OBIS-leesbegrip + OBIS-sommen A

R=Kappa= .41		OBIS-sommen A 1,00	2,00	Total
OBIS-leesbegrip	1,00	46	28	74
	2,00	47	223	270
Total		93	251	344

OBIS-leesbegrip + OBIS- getallenkennis

R= .37 Kappa=.36		OBIS-getallenkennis 1,00	2,00	Total
OBIS-leesbegrip	1,00	43	31	74
	2,00	50	220	270
Total		93	251	344

OBIS-leesbegrip + OBIS-vormenkennis

R=Kappa= .37		OBIS-vormenkennis 1,00	2,00	Total
OBIS-leesbegrip	1,00	37	37	74
	2,00	35	235	270
Total		72	272	344

OBIS-leesbegrip + OBIS-sommen B

R=Kappa=.47		OBIS-sommen B 1,00	2,00	Total
OBIS-leesbegrip	1,00	43	31	74
	2,00	31	239	270
Total		74	270	344

OBIS-klankarticulatie + OBIS-klankonderscheiding

R=Kappa=.43		OBIS-klankonderscheiding 1,00	2,00	Total
OBIS-klankarticulatie	1,00	44	26	70
	2,00	44	230	274
Total		88	256	344

OBIS-klankarticulatie + OBIS-letterkennis

R= .38 Kappa=.37		OBIS-letterkennis 1,00	2,00	Total
OBIS-klankarticulatie	1,00	37	33	70
	2,00	39	235	274
Total		76	268	344

OBIS-klankarticulatie + OBIS-rekenbegrip

R=Kappa= .18		OBIS-rekenbegrip 1,00	2,00	Total
OBIS-klankarticulatie	1,00	27	43	70
	2,00	53	221	274
Total		80	264	344

OBIS-klankarticulatie + OBIS-geheugen

R=Kappa=.31		OBIS-geheugen 1,00	2,00	Total
OBIS-klankarticulatie	1,00	36	34	70
	2,00	49	225	274
Total		85	259	344

OBIS-klankarticulatie + OBIS-sommen A

R =.41 Kappa=.40		OBIS-sommen A 1,00	2,00	Total
OBIS-klankarticulatie	1,00	44	26	70
	2,00	49	225	274
Total		93	251	344

OBIS-klankarticulatie + OBIS-getallenkennis

R=Kappa= .26		OBIS-getallenkennis 1,00	2,00	Total
OBIS-klankarticulatie	1,00	35	35	70
	2,00	58	216	274
Total		93	251	344

OBIS-klankarticulatie + OBIS-vormenkennis

R=Kappa= .27		OBIS-vormenkennis 1,00	2,00	Total
OBIS-klankarticulatie	1,00	30	40	70
	2,00	42	232	274
Total		72	272	344

OBIS-klankarticulatie + OBIS-sommen B

R=Kappa= .42		OBIS-sommen B 1,00	2,00	Total
OBIS-klankarticulatie	1,00	39	31	70
	2,00	35	239	274
Total		74	270	344

OBIS-klankonderscheiding + OBIS-letterkennis

R= .51 Kappa=.50		OBIS-letterkennis 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	51	37	88
	2,00	25	231	256
Total		76	268	344

OBIS-klankonderscheiding + OBIS-rekenbegrip

R=Kappa= .21		OBIS-rekenbegrip 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	34	54	88
	2,00	46	210	256
Total		80	264	344

OBIS-klankonderscheiding + OBIS-geheugen

R=Kappa= .41		OBIS-geheugen 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	48	40	88
	2,00	37	219	256
Total		85	259	344

OBIS-klankonderscheiding + OBIS-sommen A

R=Kappa= .51		OBIS-sommen A 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	58	30	88
	2,00	35	221	256
Total		93	251	344

OBIS-klankonderscheiding + OBIS-getallenkennis

R=Kappa= .36		OBIS-getallenkennis 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	48	40	88
	2,00	45	211	256
Total		93	251	344

OBIS-klankonderscheiding + OBIS-vormenkennis

R= .44 Kappa=.43		OBIS-vormenkennis 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	45	43	88
	2,00	27	229	256
Total		72	272	344

OBIS-klankonderscheiding + OBIS-sommen B

R=Kappa= .55		OBIS-sommen B 1,00	2,00	Total
OBIS-klankonderscheiding	1,00	53	35	88
	2,00	21	235	256
Total		74	270	344

OBIS-letterkennis + OBIS-rekenbegrip

R=Kappa= .22		OBIS-rekenbegrip 1,00	2,00	Total
OBIS-letterkennis	1,00	31	45	76
	2,00	49	219	268
Total		80	264	344

OBIS-letterkennis + OBIS-geheugen

R=Kappa=.38		OBIS-geheugen 1,00	OBIS-geheugen 2,00	Total
OBIS-letterkennis	1,00	42	34	76
	2,00	43	225	268
Total		85	259	344

OBIS-letterkennis + OBIS-sommen A

R=Kappa= .51		OBIS-sommen A 1,00	OBIS-sommen A 2,00	Total
OBIS-letterkennis	1,00	53	23	76
	2,00	40	228	268
Total		93	251	344

OBIS-letterkennis + OBIS-getallenkennis

R=Kappa= .54		OBIS-getallenkennis 1,00	OBIS-getallenkennis 2,00	Total
OBIS-letterkennis	1,00	55	21	76
	2,00	38	230	268
Total		93	251	344

OBIS-letterkennis + OBIS-vormenkennis

R=Kappa= .36		OBIS-vormenkennis 1,00	OBIS-vormenkennis 2,00	Total
OBIS-letterkennis	1,00	37	39	76
	2,00	35	233	268
Total		72	272	344

OBIS-letterkennis + OBIS-sommen B

R=Kappa= .52		OBIS-sommen B 1,00	2,00	Total
OBIS-letterkennis	1,00	47	29	76
	2,00	27	241	268
Total		74	270	344

OBIS-rekenbegrip + OBIS-geheugen

R=Kappa= .16		OBIS-geheugen 1,00	2,00	Total
OBIS-rekenbegrip	1,00	30	50	80
	2,00	55	209	264
Total		85	259	344

OBIS-rekenbegrip + OBIS-sommen A

R=Kappa= .25		OBIS-sommen A 1,00	2,00	Total
OBIS-rekenbegrip	1,00	38	42	80
	2,00	55	209	264
Total		93	251	344

OBIS-rekenbegrip + OBIS-getallenkennis

R=Kappa= .30		OBIS-getallenkennis 1,00	2,00	Total
OBIS-rekenbegrip	1,00	41	39	80
	2,00	52	212	264
Total		93	251	344

OBIS-rekenbegrip + OBIS-vormenkennis

R=Kappa= .19		OBIS-vormenkennis 1,00	2,00	Total
OBIS-rekenbegrip	1,00	28	52	80
	2,00	44	220	264
Total		72	272	344

OBIS-rekenbegrip + OBIS-sommen B

R=Kappa= .21		OBIS-sommen B 1,00	2,00	Total
OBIS-rekenbegrip	1,00	30	50	80
	2,00	44	220	264
Total		74	270	344

OBIS-geheugen + OBIS-sommen A

R=Kappa= .49		OBIS-sommen A 1,00	2,00	Total
OBIS-geheugen	1,00	55	30	85
	2,00	38	221	259
Total		93	251	344

OBIS-geheugen + OBIS-getallenkennis

R= .43 Kappa=.42		OBIS-getallenkennis 1,00	2,00	Total
OBIS-geheugen	1,00	51	34	85
	2,00	42	217	259
Total		93	251	344

OBIS-geheugen + OBIS-vormenkennis

R=Kappa= .35		OBIS-vormenkennis 1,00	2,00	Total
OBIS-geheugen	1,00	39	46	85
	2,00	33	226	259
Total		72	272	344

OBIS-geheugen + OBIS-sommen B

R=Kappa= .49		OBIS-sommen B 1,00	2,00	Total
OBIS-geheugen	1,00	48	37	85
	2,00	26	233	259
Total		74	270	344

OBIS-sommen A + OBIS-getallenkennis

R=Kappa= .57		OBIS-getallenkennis 1,00	2,00	Total
OBIS-sommen A	1,00	64	29	93
	2,00	29	222	251
Total		93	251	344

OBIS-sommen A + OBIS-vormenkennis

R=Kappa= .44		OBIS-vormenkennis 1,00	2,00	Total
OBIS-sommen A	1,00	47	46	93
	2,00	25	226	251
Total		72	272	344

OBIS-sommen A + OBIS-sommen B

R= .81 Kappa=.80		OBIS-sommen B 1,00	2,00	Total
OBIS-sommen A	1,00	71	22	93
	2,00	3	248	251
Total		74	270	344

OBIS-getallenkennis + OBIS-vormenkennis

R=Kappa=.41		OBIS-vormenkennis 1,00	2,00	Total
OBIS-getallenkennis	1,00	45	48	93
	2,00	27	224	251
Total		72	272	344

OBIS-getallenkennis + OBIS-sommen B

R=Kappa= .54		OBIS-sommen B 1,00	2,00	Total
OBIS-getallenkennis	1,00	54	39	93
	2,00	20	231	251
Total		74	270	344

OBIS-vormenkennis + OBIS sommen B

R=Kappa=.41		OBIS-sommen B 1,00	2,00	Total
OBIS-vormenkennis	1,00	39	33	72
	2,00	35	237	272
Total		74	270	344

Test hertest betrouwbaarheid per (sub)toets in het oormerken van het onderste kwartiel van de leerlingen ofwel de 25% slechtst presterende leerlingen.

TvK

R= Kappa=.81		test 1,00	test 2,00	Total
hertest	1,00	13	2	15
	2,00	3	74	77
Total		16	76	92

Peabody

R= Kappa=.63		test 1,00	test 2,00	Total
hertest	1,00	11	4	15
	2,00	6	75	81
Total		17	79	96

TAK-klankonderscheiding

R= .61 Kappa=..58		test 1,00	test 2,00	Total
hertest	1,00	25	2	27
	2,00	17	50	67
Total		42	52	94

TAK-klankarticulatie

R= Kappa=.65		test 1,00	test 2,00	Total
hertest	1,00	25	5	30
	2,00	10	54	64
Total		35	59	94

TAK-passieve woordenschat

R=.62 Kappa=.59		test 1,00	test 2,00	Total
hertest	1,00	22	3	25
	2,00	14	55	69
Total		36	58	94

OBIS-schrijven

R= Kappa=.75		test 1,00	test 2,00	Total
hertest	1,00	14	5	19
	2,00	2	61	63
Total		16	66	82

OBIS-passieve woordenschat

R=.81 Kappa=.80		test 1,00	test 2,00	Total
hertest	1,00	13	1	14
	2,00	4	64	68
Total		17	65	82

OBIS-leesbegrip

R=.36 Kappa=.35		test 1,00	test 2,00	Total
hertest	1,00	5	4	9
	2,00	9	64	73
Total		14	68	82

OBIS-klankarticulatie

R= Kappa=.87		test 1,00	test 2,00	Total
hertest	1,00	12	2	14
	2,00	1	67	68
Total		13	69	82

OBIS-klankonderscheiding

R=.77 Kappa=.76		test 1,00	test 2,00	Total
hertest	1,00	12	1	13
	2,00	5	64	69
Total		17	65	82

OBIS-letterkennis

R= Kappa=.83		test 1,00	test 2,00	Total
hertest	1,00	16	2	18
	2,00	3	61	64
Total		19	63	82

OBIS-rekenbegrip

R=.59 Kappa=.55		test 1,00	test 2,00	Total
hertest	1,00	9	1	10
	2,00	10	62	72
Total		19	63	82

OBIS-geheugen

R= Kappa=.69		test 1,00	test 2,00	Total
hertest	1,00	12	5	17
	2,00	3	62	65
Total		15	67	82

OBIS-sommen A

R=.75 Kappa=.72		test 1,00	test 2,00	Total
hertest	1,00	14	0	14
	2,00	8	60	68
Total		22	60	82

OBIS-getallenkennis

R= Kappa=.85		test 1,00	test 2,00	Total
hertest	1,00	20	1	21
	2,00	4	57	61
Total		24	58	82

OBIS-vormen

R=.57 Kappa=.56		test 1,00	test 2,00	Total
hertest	1,00	10	4	14
	2,00	7	61	68
Total		17	65	82

OBIS-somen B

R=.75 Kappa=.73		test 1,00	test 2,00	Total
hertest	1,00	12	0	12
	2,00	7	63	70
Total		19	63	82

Kruistabellen leerkrachtoordelen receptieve en productieve spreekvaardigheid versus de op het onderste kwartiel gesplitste scores op de (sub)toetsen.

DOCENTOORDEEL VERSTAAN/BEGRIJPEN NEDERLANDS en …

OBIS-schrijven

		OBIS-schrijven		
		1,00	2,00	Total
Docentoordeel mate begrip van gesproken Nederlands	3	2	1	3
	4	8	1	9
	5	9	12	21
	6	13	30	43
	7	19	77	96
	8	11	75	86
	9	3	36	39
	10	1	7	8
Total		66	239	305

Peabody

		dsompea dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	2	1	0	1
	3	1	1	2
	4	7	4	11
	5	16	7	23
	6	11	13	24
	7	34	93	127
	8	7	79	86
	9	2	39	41
	10	0	10	10
Total		79	246	325

TAK-klankonderscheiding

		dsomtko dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	2	1	0	1
	3	4	2	6
	4	5	1	6
	5	17	7	24
	6	16	30	46
	7	22	78	100
	8	13	66	79
	9	5	37	42
	10	3	20	23
Total		86	241	327

TAK-klankarticulatie

		dsomtka dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	2	1	0	1
	3	6	0	6
	4	6	0	6
	5	14	10	24
	6	12	34	46
	7	20	80	100
	8	16	63	79
	9	1	41	42
	10	3	20	23
Total		79	248	327

TAK-passieve woordkennis

		dsomtpw dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	2	1	0	1
	3	4	2	6
	4	4	2	6
	5	16	8	24
	6	15	31	46
	7	24	76	100
	8	10	69	79
	9	5	37	42
	10	1	22	23
Total		80	247	327

OBIS-passieve woordkennis

		dsomopw dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	3	0	3
	4	8	1	9
	5	14	7	21
	6	18	25	43
	7	21	75	96
	8	9	77	86
	9	0	39	39
	10	1	7	8
Total		74	231	305

OBIS-leesbegrip

		dsomolb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	5	4	9
	5	10	11	21
	6	11	32	43
	7	19	77	96
	8	12	74	86
	9	1	38	39
	10	1	7	8
Total		61	244	305

OBIS-klankarticulatie

		dsomonaz dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	3	0	3
	4	8	1	9
	5	11	10	21
	6	13	30	43
	7	18	78	96
	8	7	79	86
	9	1	38	39
	10	1	7	8
Total		62	243	305

OBIS-klankonderscheiding

		dsomoko dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	3	0	3
	4	8	1	9
	5	13	8	21
	6	15	28	43
	7	23	73	96
	8	11	75	86
	9	3	36	39
	10	2	6	8
Total		78	227	305

OBIS-letterkennis

		dsomolk dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	6	3	9
	5	11	10	21
	6	10	33	43
	7	19	77	96
	8	7	79	86
	9	2	37	39
	10	0	8	8
Total		57	248	305

OBIS-rekenjargon

		dsomorb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	6	3	9
	5	15	6	21
	6	12	31	43
	7	18	78	96
	8	11	75	86
	9	5	34	39
	10	1	7	8
Total		70	235	305

OBIS-geheugen

		dsomogh dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	6	3	9
	5	10	11	21
	6	15	28	43
	7	21	75	96
	8	14	72	86
	9	5	34	39
	10	3	5	8
Total		76	229	305

OBIS-sommen A

		dsomosa dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	3	0	3
	4	9	0	9
	5	13	8	21
	6	17	26	43
	7	23	73	96
	8	8	78	86
	9	3	36	39
	10	2	6	8
Total		78	227	305

OBIS-getallenkennis

		dsomoget dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	7	2	9
	5	10	11	21
	6	19	24	43
	7	20	76	96
	8	9	77	86
	9	1	38	39
	10	2	6	8
Total		70	235	305

OBIS-vormenkennis

		dsomovm dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	1	2	3
	4	9	0	9
	5	10	11	21
	6	17	26	43
	7	20	76	96
	8	4	82	86
	9	5	34	39
	10	1	7	8
Total		67	238	305

OBIS-sommen B

		dsomosb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vbegrijp Taalvaardigheid sniveau Nederlands verstaan/begrijpen	3	2	1	3
	4	9	0	9
	5	10	11	21
	6	14	29	43
	7	19	77	96
	8	4	82	86
	9	2	37	39
	10	1	7	8
Total		61	244	305

DOCENTOORDEEL SPREKEN NEDERLANDS en …

OBIS-schrijven

		dq5 dichotome op 25% nietgecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	2	2	4
	4	8	5	13
	5	12	18	30
	6	15	50	65
	7	16	86	102
	8	8	38	46
	9	2	33	35
	10	0	7	7
Total		66	239	305

TvK

		dsomtvk15 dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	1	3	0	3
	2	6	1	7
	3	2	2	4
	4	7	4	11
	5	11	11	22
	6	21	43	64
	7	14	82	96
	8	2	59	61
	9	3	35	38
	10	2	10	12
Total		71	247	318

Peabody

		dsompea dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	1	1	0	1
	2	3	1	4
	3	3	1	4
	4	11	2	13
	5	10	10	20
	6	23	26	49
	7	22	105	127
	8	6	56	62
	9	0	37	37
	10	0	8	8
Total		79	246	325

TAK-klankonderscheiding

		dsomtko dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	1	1	0	1
	2	7	2	9
	3	2	2	4
	4	13	3	16
	5	12	17	29
	6	18	36	54
	7	19	78	97
	8	11	56	67
	9	2	31	33
	10	1	16	17
Total		86	241	327

TAK-klankarticulatie

		dsomtka dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	1	1	0	1
	2	9	0	9
	3	3	1	4
	4	8	8	16
	5	12	17	29
	6	19	35	54
	7	14	83	97
	8	11	56	67
	9	0	33	33
	10	2	15	17
Total		79	248	327

TAK-passieve woordkennis

		dsomtpw dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	1	1	0	1
	2	9	0	9
	3	2	2	4
	4	12	4	16
	5	13	16	29
	6	16	38	54
	7	16	81	97
	8	10	57	67
	9	1	32	33
	10	0	17	17
Total		80	247	327

OBIS-passieve woordkennis

		dsomopw dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	4	0	4
	4	12	1	13
	5	18	12	30
	6	17	48	65
	7	16	86	102
	8	4	42	46
	9	0	35	35
	10	0	7	7
Total		74	231	305

OBIS-leesbegrip

		dsomolb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	3	1	4
	4	8	5	13
	5	11	19	30
	6	15	50	65
	7	15	87	102
	8	6	40	46
	9	1	34	35
	10	0	7	7
Total		61	244	305

OBIS-klankarticulatie

		dsomonaz dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	3	1	4
	4	11	2	13
	5	11	19	30
	6	17	48	65
	7	11	91	102
	8	5	41	46
	9	1	34	35
	10	0	7	7
Total		62	243	305

OBIS-klankonderscheiding

		dsomoko dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	4	0	4
	4	10	3	13
	5	15	15	30
	6	18	47	65
	7	17	85	102
	8	7	39	46
	9	3	32	35
	10	1	6	7
Total		78	227	305

OBIS-letterkennis

		dsomolk dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	3	1	4
	4	8	5	13
	5	13	17	30
	6	12	53	65
	7	14	88	102
	8	3	43	46
	9	2	33	35
	10	0	7	7
Total		57	248	305

OBIS-rekenjargon

		dsomorb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	4	0	4
	4	8	5	13
	5	14	16	30
	6	14	51	65
	7	16	86	102
	8	7	39	46
	9	4	31	35
	10	1	6	7
Total		70	235	305

OBIS-geheugen

		dsomogh dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	3	1	4
	4	9	4	13
	5	12	18	30
	6	18	47	65
	7	14	88	102
	8	12	34	46
	9	4	31	35
	10	2	5	7
Total		76	229	305

OBIS-sommen A

		dsomosa dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	4	0	4
	4	11	2	13
	5	13	17	30
	6	18	47	65
	7	19	83	102
	8	7	39	46
	9	2	33	35
	10	1	6	7
Total		78	227	305

OBIS-getallenkennis

		dsomoget dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	3	1	4
	4	10	3	13
	5	13	17	30
	6	20	45	65
	7	15	87	102
	8	6	40	46
	9	0	35	35
	10	1	6	7
Total		70	235	305

OBIS-vormenkennis

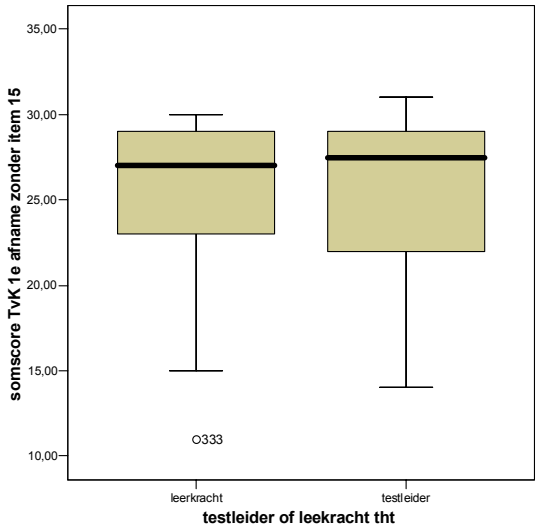
		dsomovm dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	2	1	3
	3	3	1	4
	4	12	1	13
	5	9	21	30
	6	20	45	65
	7	13	89	102
	8	4	42	46
	9	4	31	35
	10	0	7	7
Total		67	238	305

OBIS-sommen B

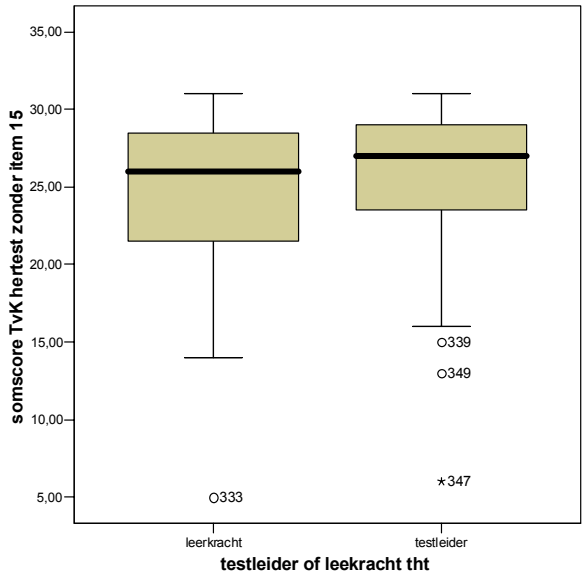
		dsomosb dichotome op 25% niet gecorrigeerd voor lft		
		1,00	2,00	Total
vspreek Taalvaardigheid niveau Nederlands spreken	2	3	0	3
	3	3	1	4
	4	11	2	13
	5	11	19	30
	6	14	51	65
	7	12	90	102
	8	5	41	46
	9	2	33	35
	10	0	7	7
Total		61	244	305

APPENDIX BOXPLOTS TESTLEIDERS VERSUS LEERKRACHTEN

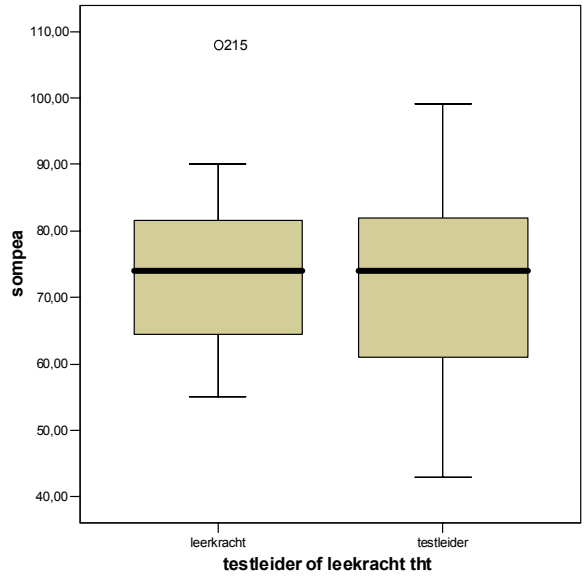
Figuur A45: Boxplots TvK voor leerkrachten en testleiders, 1e afname



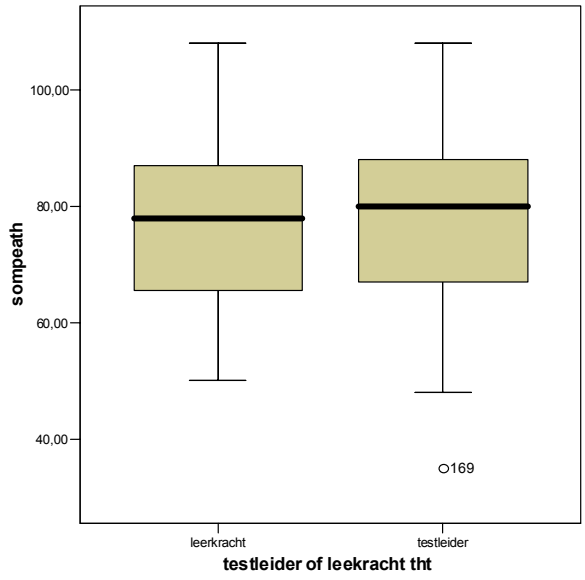
Figuur A46: Boxplots TvK voor leerkrachten en testleiders, hertest



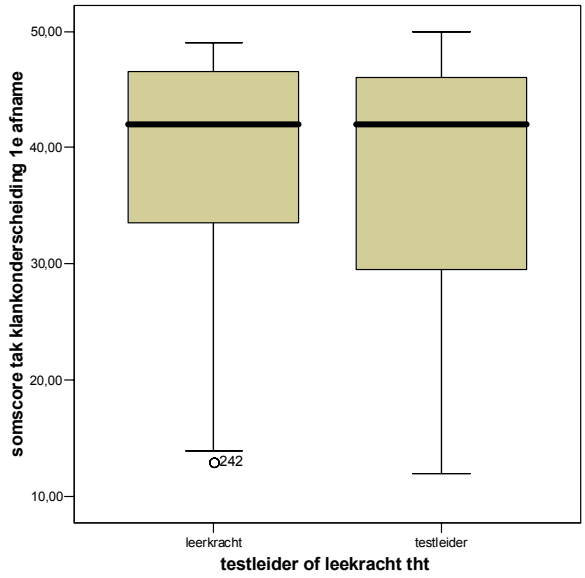
Figuur A47: Boxplots Peabody voor leerkrachten en testleiders, 1e afname



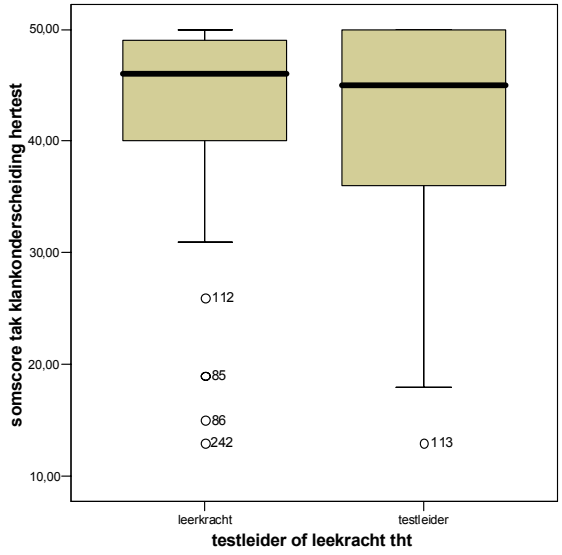
Figuur A48: Boxplots Peabody voor leerkrachten en testleiders, hertest



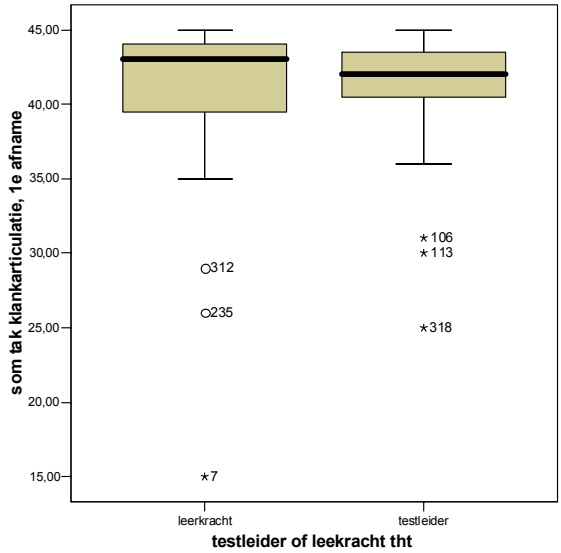
Figuur A49: Boxplots TAK-klankarticulatie voor leerkrachten en testleiders, 1e afname



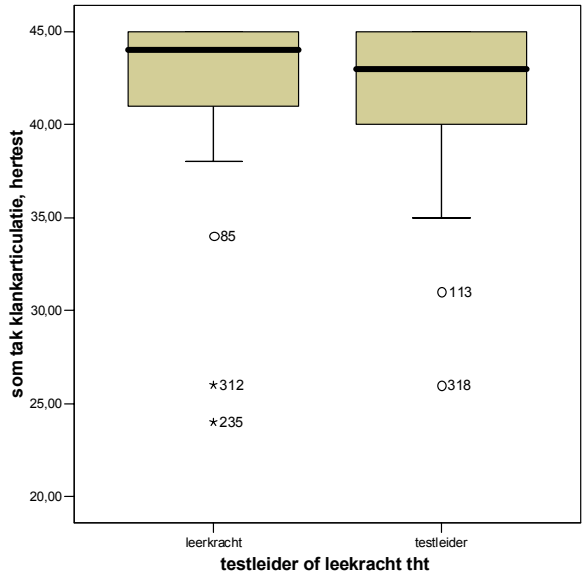
Figuur A50: Boxplots TAK-klankarticulatie voor leerkrachten en testleiders, hertest



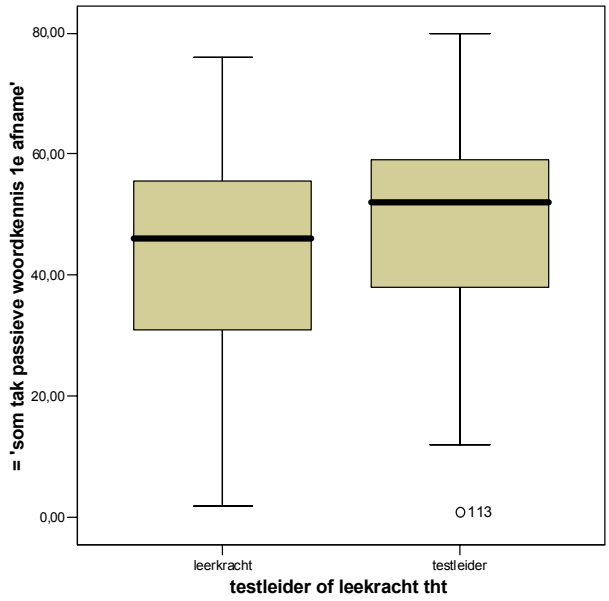
Figuur A51: Boxplots TAK-klankonderscheiding voor leerkrachten en testleiders, 1e afname



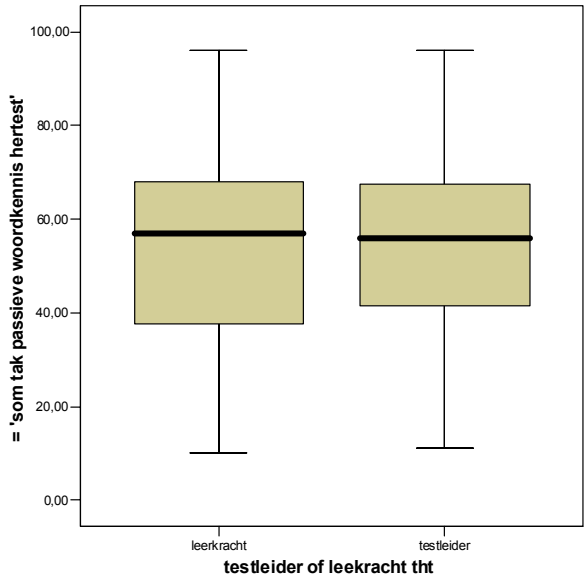
Figuur A52: Boxplots TAK-klankonderscheiding voor leerkrachten en testleiders, hertest



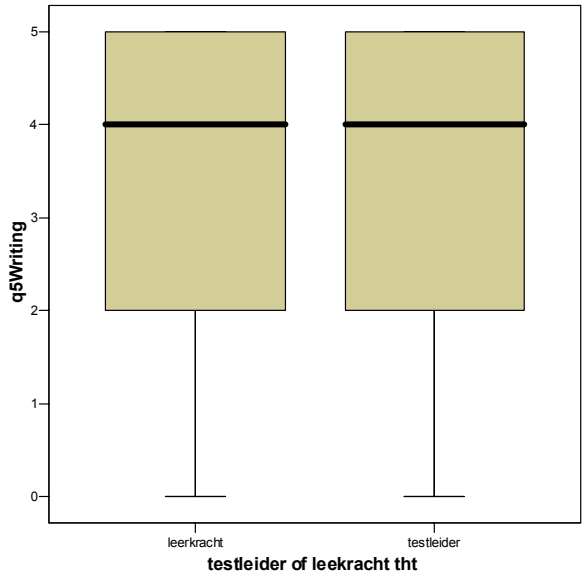
Figuur A53: Boxplots TAK-passieve woordkennis voor leerkrachten en testleiders, 1e afname



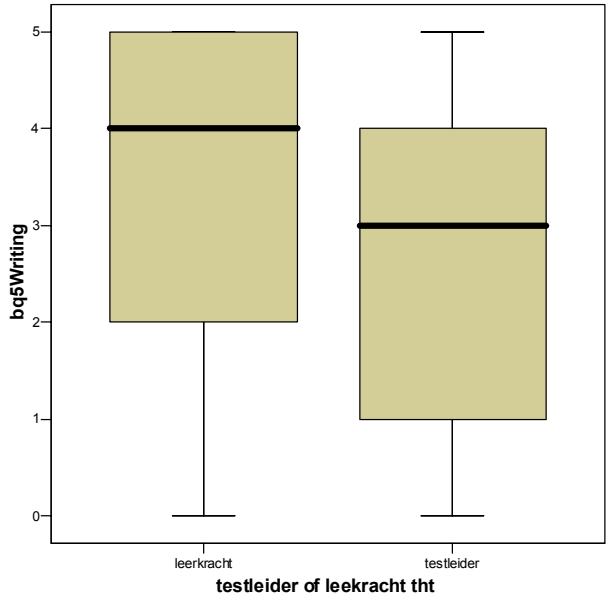
Figuur A54: Boxplots TAK-passieve woordkennis voor leerkrachten en testleiders, hertest



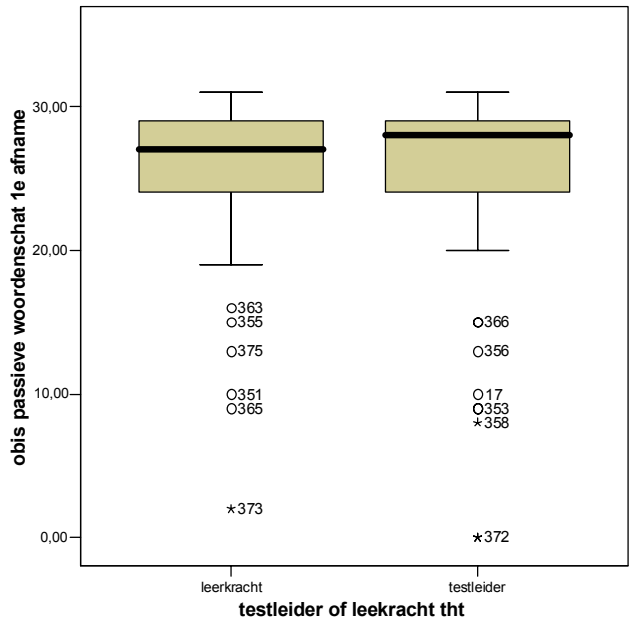
Figuur A55: Boxplots OBIS-schrijven voor leerkrachten en testleiders, 1e afname



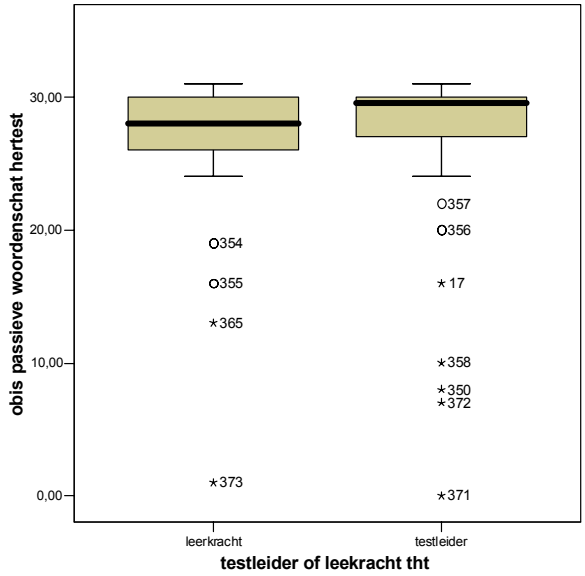
Figuur A56: Boxplots OBIS-schrijven voor leerkrachten en testleiders, hertest



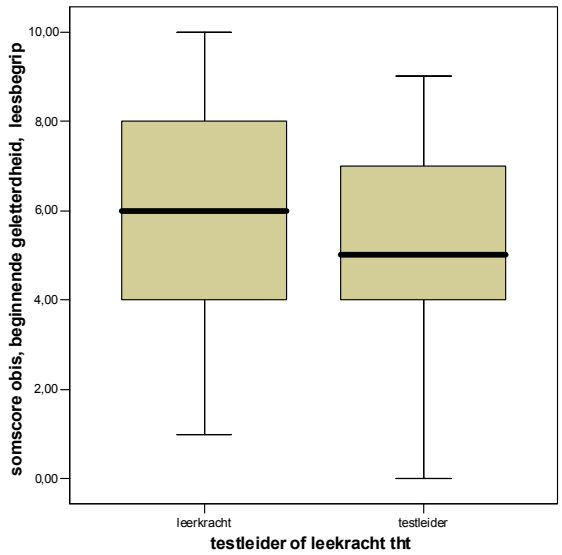
Figuur A57: Boxplots OBIS-passieve woordkennis voor leerkrachten en testleiders, 1e afname



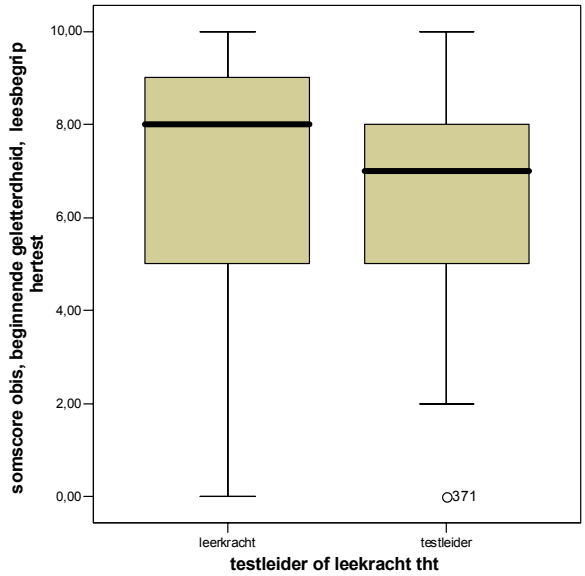
Figuur A58: Boxplots OBIS-passieve woordkennis voor leerkrachten en testleiders, hertest



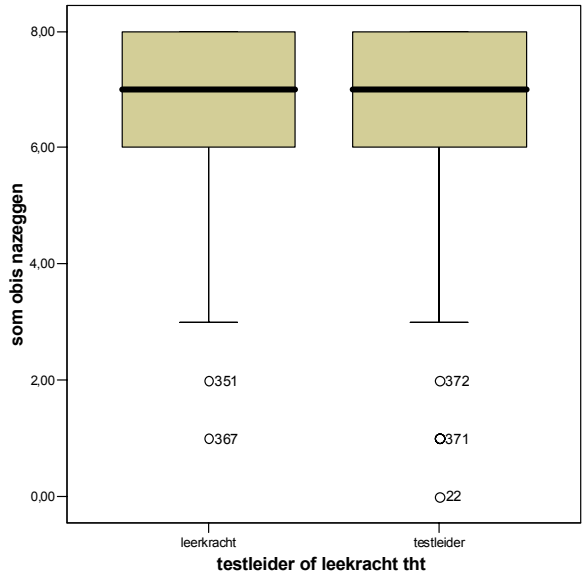
Figuur A59: Boxplots OBIS-leesbegrip voor leerkrachten en testleiders, 1e afname



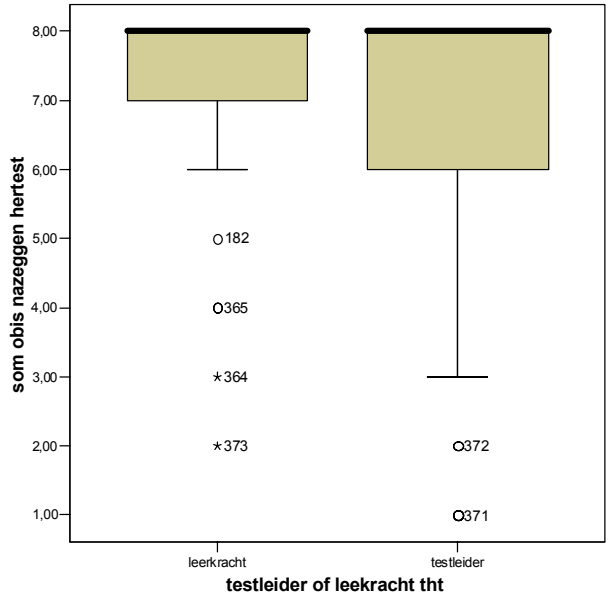
Figuur A60: Boxplots OBIS-leesbegrip voor leerkrachten en testleiders, hertest



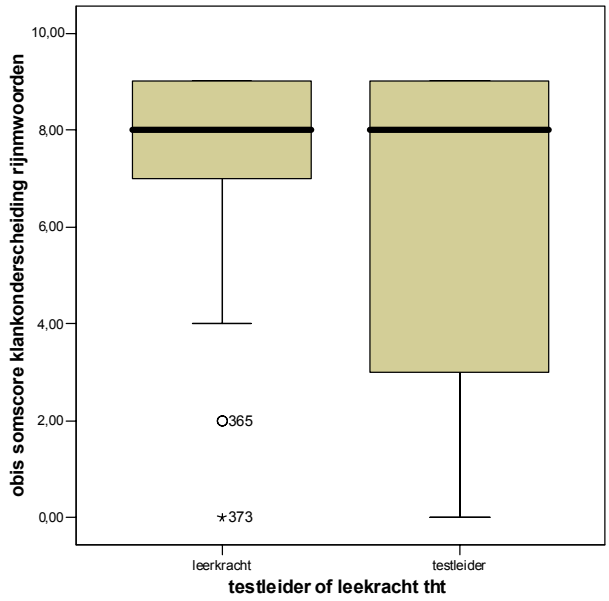
Figuur A61: Boxplots OBIS-klankarticulatie voor leerkrachten en testleiders, 1e afname



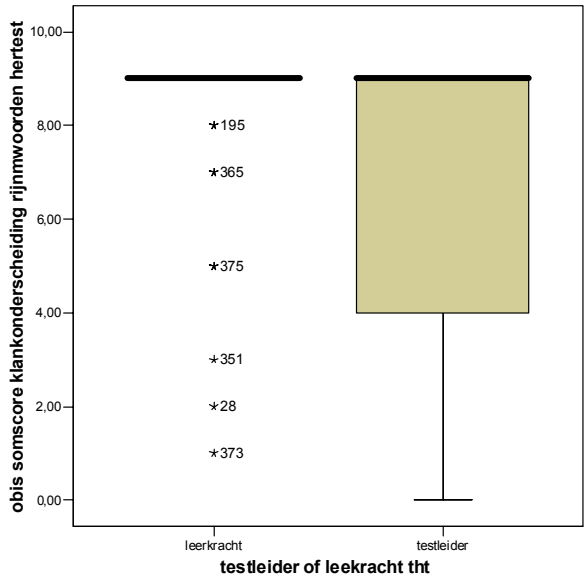
Figuur A62: Boxplots OBIS-klankarticulatie voor leerkrachten en testleiders, hertest



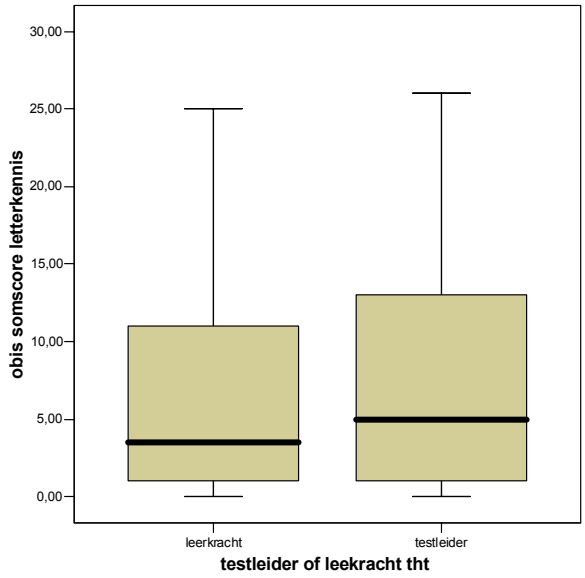
Figuur A63: Boxplots OBIS-klankonderscheiding voor leerkrachten en testleiders, 1e afname



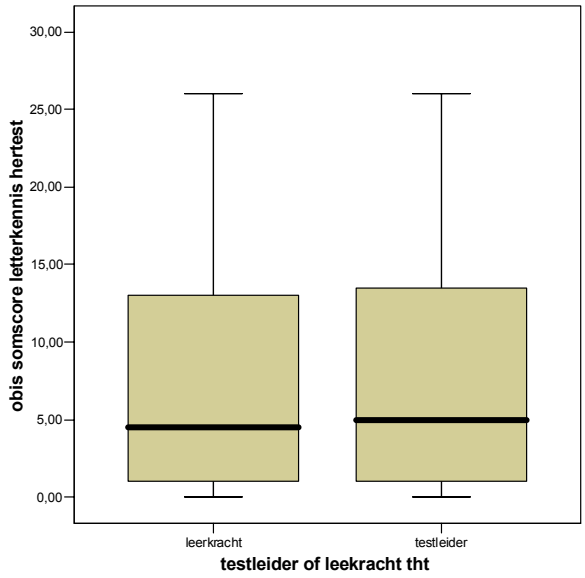
Figuur A64: Boxplots OBIS-klankonderscheiding voor leerkrachten en testleiders, hertest



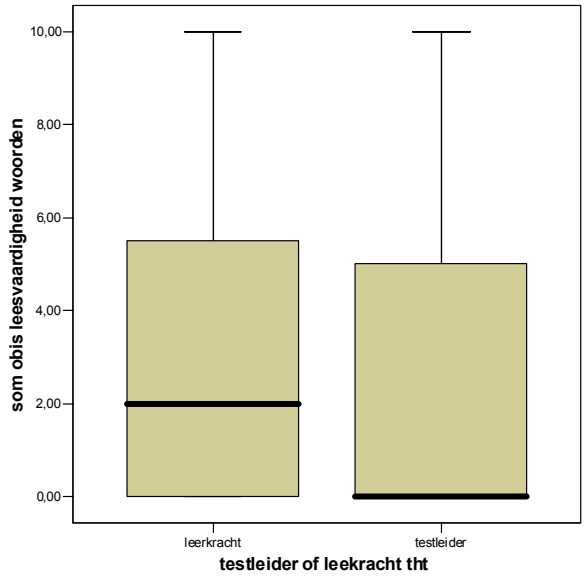
Figuur A65: Boxplots OBIS-letterkennis voor leerkrachten en testleiders, 1e afname



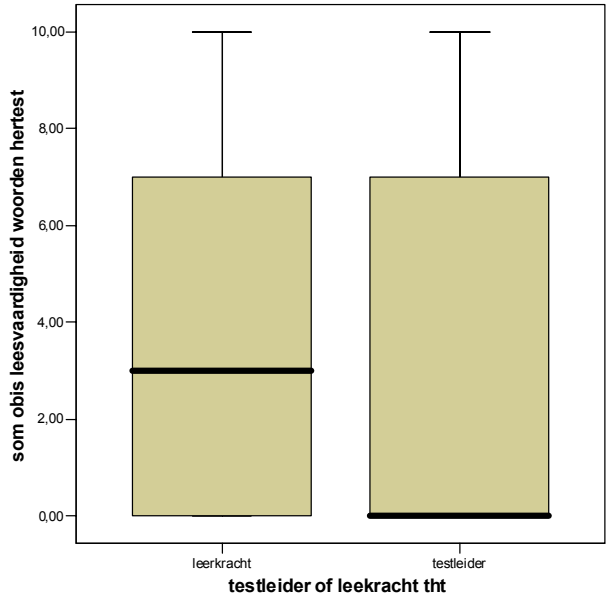
Figuur A66: Boxplots OBIS-letterkennis voor leerkrachten en testleiders, hertest



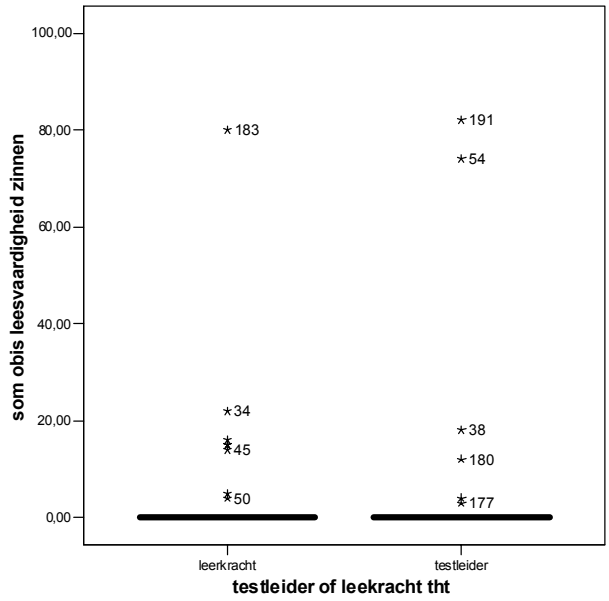
Figuur A67: Boxplots OBIS-leesvaardigheid woorden voor leerkrachten en testleiders, 1e afname



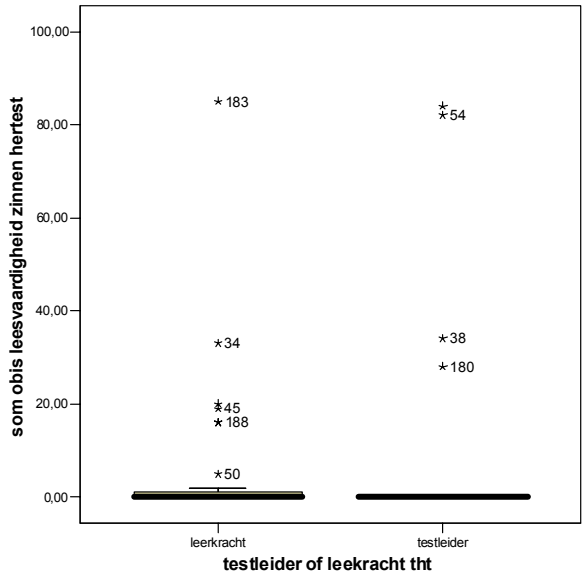
Figuur A68: Boxplots OBIS- leesvaardigheid woorden voor leerkrachten en testleiders, hertest



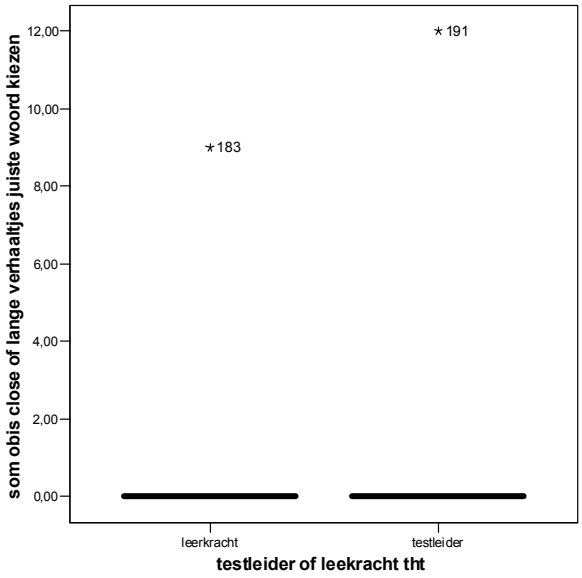
Figuur A69: Boxplots OBIS-leesvaardigheid zinnen voor leerkrachten en testleiders, 1e afname



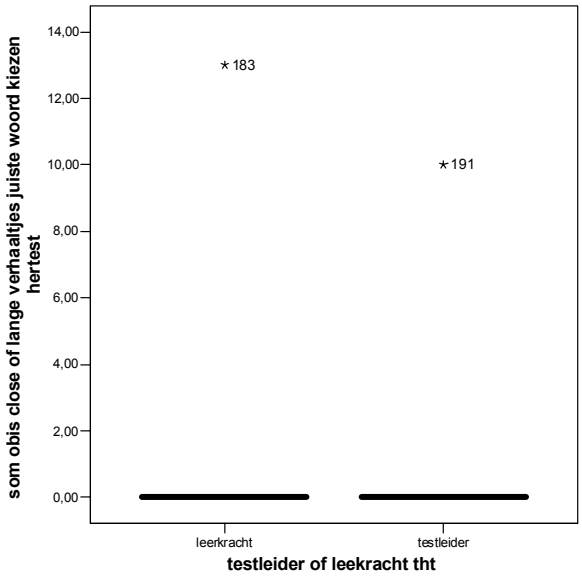
Figuur A70: Boxplots OBIS-leesvaardigheid zinnen voor leerkrachten en testleiders, hertest



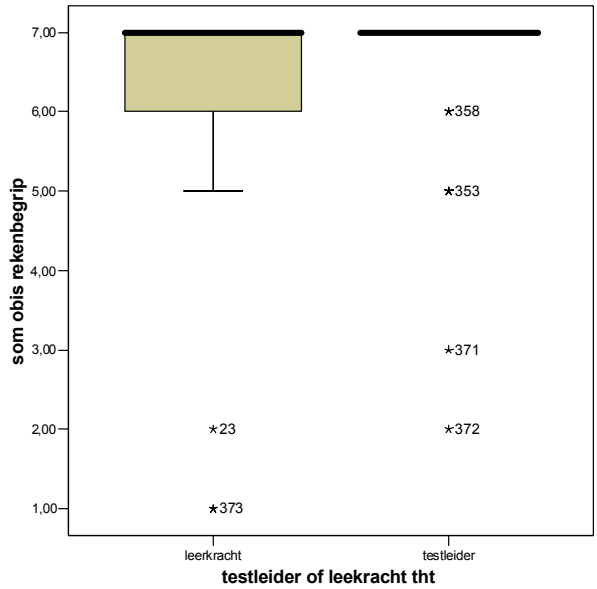
Figuur A71: Boxplots OBIS-leesvaardigheid teksten voor leerkrachten en testleiders, 1e afname



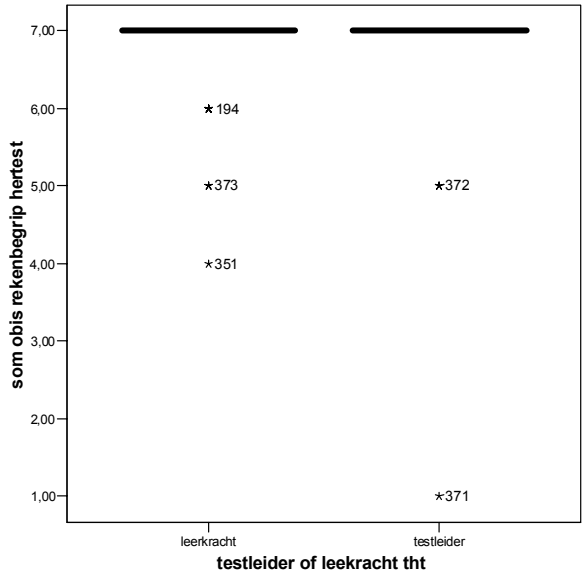
Figuur A72: Boxplots OBIS-leesvaardigheid teksten voor leerkrachten en testleiders, hertest



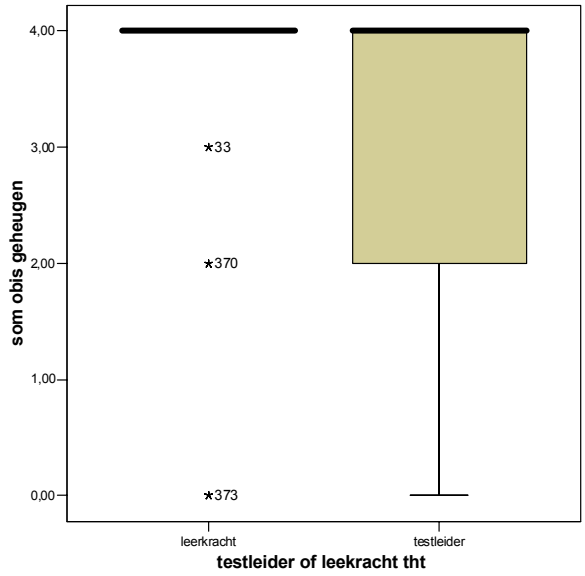
Figuur A73: Boxplots OBIS-rekenjargon voor leerkrachten en testleiders, 1e afname



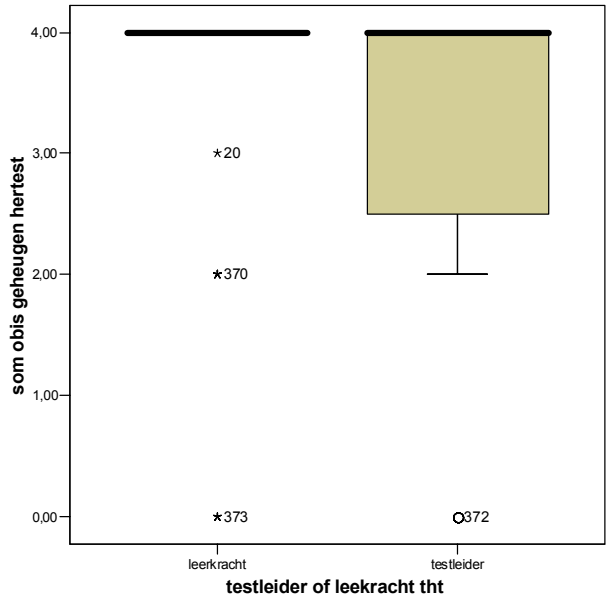
Figuur A74: Boxplots OBIS-rekenjargon voor leerkrachten en testleiders, hertest



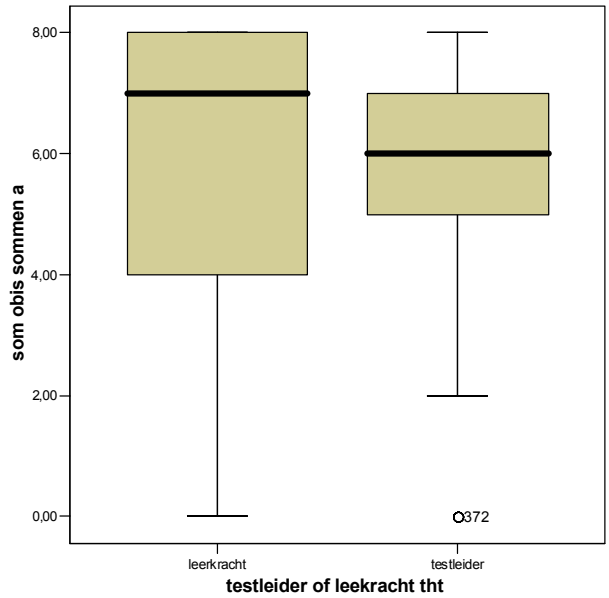
Figuur A75: Boxplots OBIS-geheugen voor leerkrachten en testleiders, 1e afname



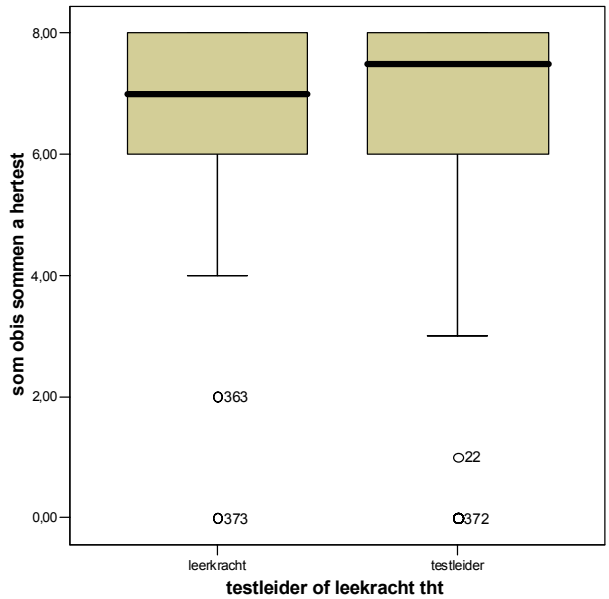
Figuur A76: Boxplots OBIS-geheugen voor leerkrachten en testleiders, hertest



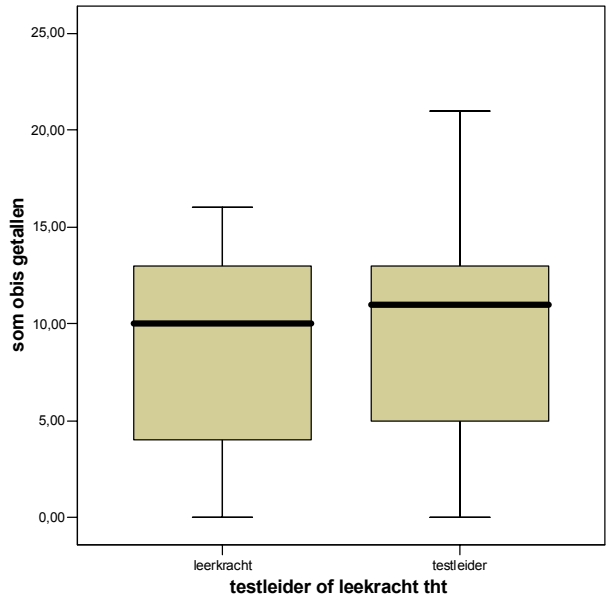
Figuur A77: Boxplots OBIS-sommen A voor leerkrachten en testleiders, 1e afname



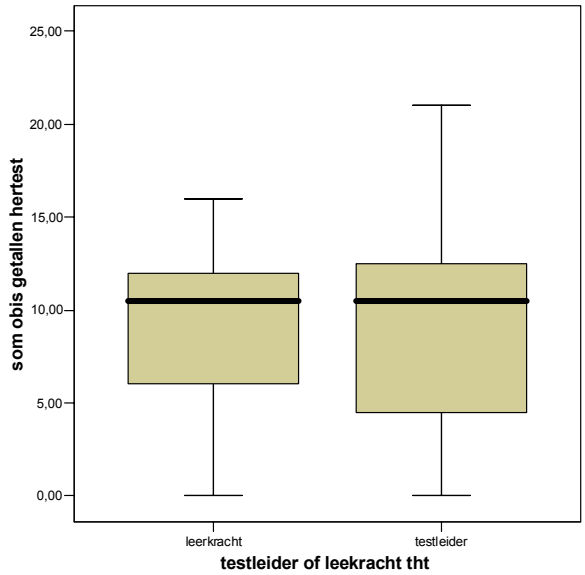
Figuur A78: Boxplots OBIS-sommen A voor leerkrachten en testleiders, hertest



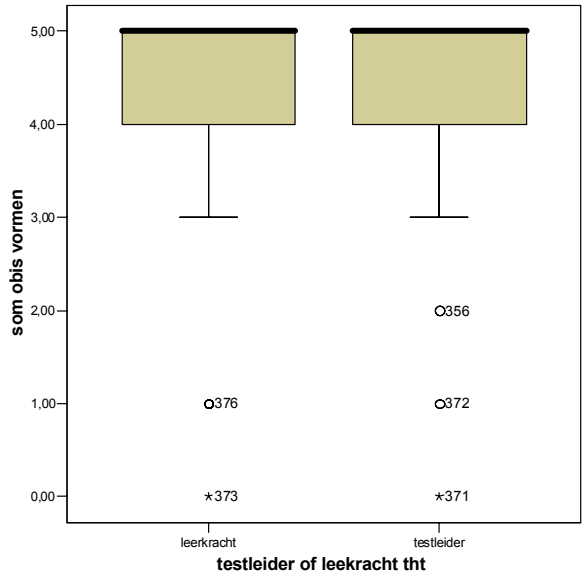
Figuur A79: Boxplots OBIS-getallenkennis voor leerkrachten en testleiders, 1e afname



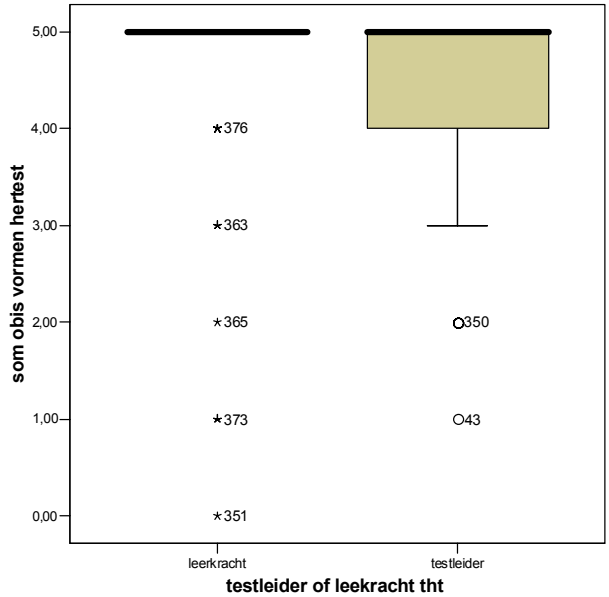
Figuur A80: Boxplots OBIS-getallenkennis voor leerkrachten en testleiders, hertest



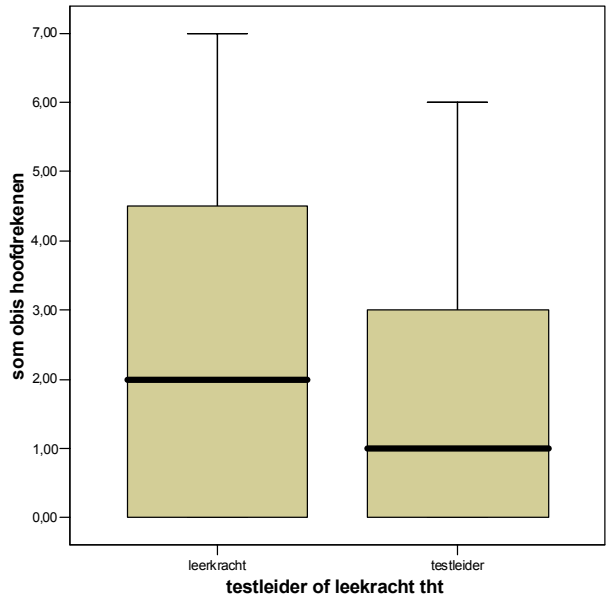
Figuur A81: Boxplots OBIS-vormen voor leerkrachten en testleiders, 1e afname



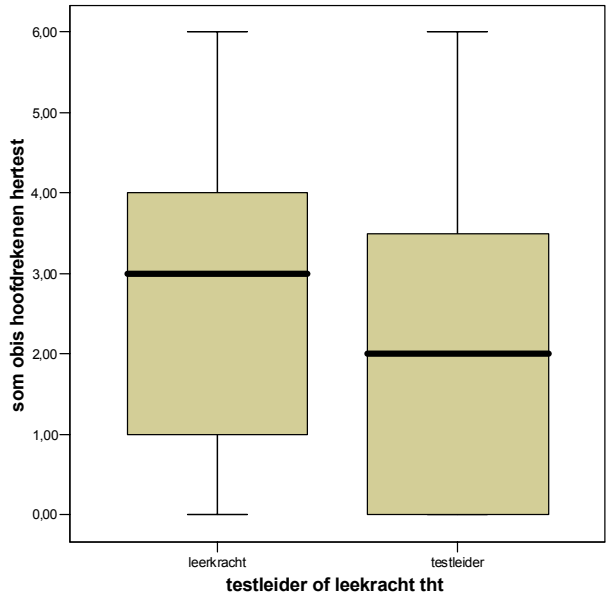
Figuur A82: Boxplots OBIS-vormen voor leerkrachten en testleiders, hertest



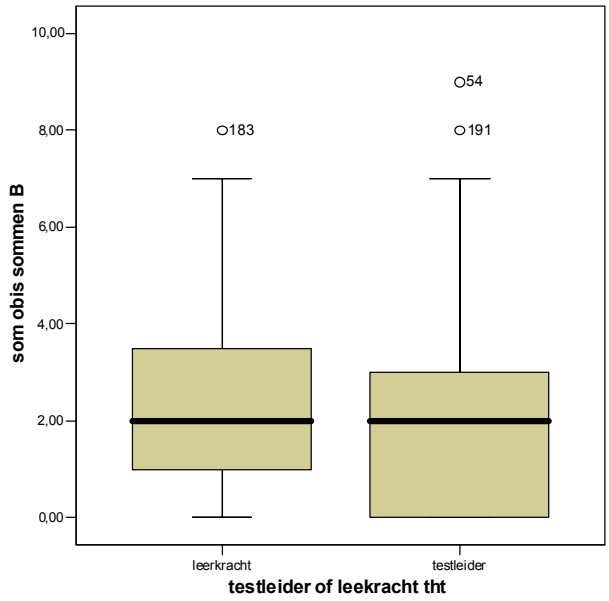
Figuur A83: Boxplots OBIS-hoofdrekenen voor leerkrachten en testleiders, 1e afname



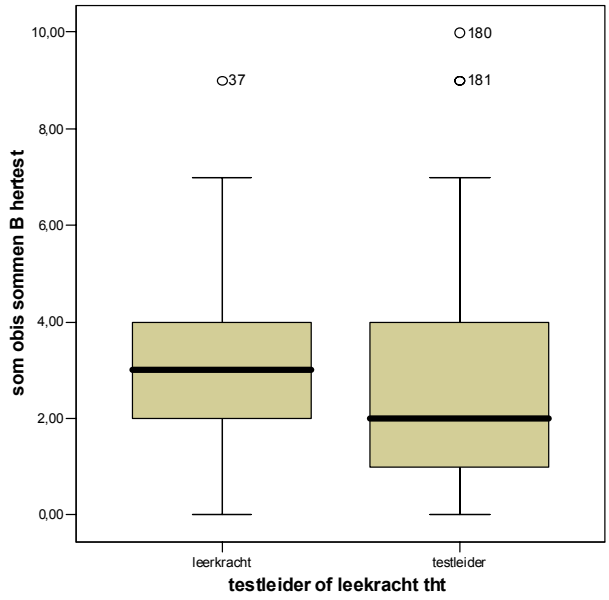
Figuur A84: Boxplots OBIS-hoofdrekenen voor leerkrachten en testleiders, hertest



Figuur A85: Boxplots OBIS-Sommen B voor leerkrachten en testleiders, 1e afname



Figuur A86: Boxplots OBIS-Sommen B voor leerkrachten en testleiders, hertest



STABILITEIT VAN TOETSEN IN HET OORMERKEN VAN DE 25% SLECHTST PRESTERENDE LEERLINGEN

Test-hertest kappa's tussen dichotome variabelen die aangeven of leerling hoort bij de 25% slechtst presterende leerlingen dan wel bij de 75% hoogst presterende leerlingen.

TVK

		hertest		
		1,00	2,00	Total
test	1,00	13	3	16
	2,00	2	74	76
Total		15	77	92

Kappa = .81

Peabody

		hertest		
		1,00	2,00	Total
test	1,00	11	6	17
	2,00	4	75	79
Total		15	81	96

Kappa = .63

TAK-klankonderscheiding

		hertest		
		1,00	2,00	Total
test	1,00	25	17	42
	2,00	2	50	52
Total		27	67	94

Kappa = .58

TAK-klankarticulatie

		hertest		
		1,00	2,00	Total
test	1,00	25	10	35
	2,00	5	54	59
Total		30	64	94

Kappa= .65

TAK-passieve woordkennis

		hertest		
		1,00	2,00	Total
test	1,00	22	14	36
	2,00	3	55	58
Total		25	69	94

Kappa = .59

OBIS-passieve woordkennis

		hertest		
		1,00	2,00	Total
test	1,00	13	4	17
	2,00	1	64	65
Total		14	68	82

Kappa = .80

OBIS-leesbegrip

		hertest		
		1,00	2,00	Total
test	1,00	5	9	14
	2,00	4	64	68
Total		9	73	82

Kappa = .35

OBIS-klankarticulatie

		hertest		
		1,00	2,00	Total
test	1,00	12	1	13
	2,00	2	67	69
Total		14	68	82

Kappa = .87

OBIS-klankonderscheiding

		hertest		
		1,00	2,00	Total
test	1,00	12	5	17
	2,00	1	64	65
Total		13	69	82

Kappa = .76

OBIS-kennis van letters

		hertest		
		1,00	2,00	Total
test	1,00	16	3	19
	2,00	2	61	63
Total		18	64	82

Kappa = .83

OBIS-aanvankelijk schrijven

		hertest		
		1,00	2,00	Total
test	1,00	14	2	16
	2,00	5	61	66
Total		19	63	82

Kappa = .75

OBIS-Rekenjargon

		hertest		
		1,00	2,00	Total
test	1,00	9	10	19
	2,00	1	62	63
Total		10	72	82

Kappa = .55

OBIS-geheugen

		hertest		
		1,00	2,00	Total
test	1,00	12	3	15
	2,00	5	62	67
Total		17	65	82

Kappa = .69

OBIS-sommen A

		hertest		
		1,00	2,00	Total
test	1,00	14	8	22
	2,00	0	60	60
Total		14	68	82

Kappa = .72

OBIS-kennis van getallen

		hertest		
		1,00	2,00	Total
test	1,00	20	4	24
	2,00	1	57	58
Total		21	61	82

Kappa = .85

OBIS-kennis van vormen

		hertest		
		1,00	2,00	Total
test	1,00	10	7	17
	2,00	4	61	65
Total		14	68	82

Kappa = .56

OBIS-sommen B

		hertest		
		1,00	2,00	Total
test	1,00	12	7	19
	2,00	0	63	63
Total		12	70	82

Kappa = .73

Kruistabellen met op de ene as de dichotome variabele die aangeeft of een leerling behoort bij de 25% slechtst presterende leerlingen op de betreffende subtoets (1) danwel bij de 75% hoogst presterende leerlingen (2) en op de andere as of de leerling een .90-leerling is (1) of niet (0). 'Bij 'Verwacht Aantal' staat de celfrequentie die bij afwezigheid van samenhang tussen beide dichotome variabelen verwacht wordt (de celfrequentie berekend uit de randfrequenties)

OBIS schrijven		.90 leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	40	38	78
	Verwacht Aantal	52,4	25,6	78,0
	% binnen toets	51,3%	48,7%	100,0%
	% binnen .90	19,1%	37,3%	25,1%
	% van Totaal	12,9%	12,2%	25,1%
Bovenste 75%	Aantal	169	64	233
	Verwacht Aantal	156,6	76,4	233,0
	% binnen toets	72,5%	27,5%	100,0%
	% binnen .90	80,9%	62,7%	74,9%
	% van Totaal	54,3%	20,6%	74,9%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .196

TvK		.90 leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	28	44	72
	Verwacht Aantal	51,3	20,7	72,0
	% binnen toets	38,9%	61,1%	100,0%
	% binnen .90	12,7%	49,4%	23,2%
	% van Totaal	9,0%	14,2%	23,2%
Bovenste 75%	Aantal	193	45	238
	Verwacht Aantal	169,7	68,3	238,0
	% binnen toets	81,1%	18,9%	100,0%
	% binnen .90	87,3%	50,6%	76,8%
	% van Totaal	62,3%	14,5%	76,8%
Totaal	Aantal	221	89	310
	Verwacht Aantal	221,0	89,0	310,0
	% binnen toets	71,3%	28,7%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	71,3%	28,7%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .394

Peabody		.90-leerling niet	.90-leerling wel	Totaal
Onderste 25%	Aantal	33	48	81
	Verwacht Aantal	58,4	22,6	81,0
	% binnen toets	40,7%	59,3%	100,0%
	% binnen .90	14,5%	54,5%	25,6%
	% van Totaal	10,4%	15,2%	25,6%
Bovenste 75%	Aantal	195	40	235
	Verwacht Aantal	169,6	65,4	235,0
	% binnen toets	83,0%	17,0%	100,0%
	% binnen .90	85,5%	45,5%	74,4%
	% van Totaal	61,7%	12,7%	74,4%
Totaal	Aantal	228	88	316
	Verwacht Aantal	228,0	88,0	316,0
	% binnen toets	72,2%	27,8%	100,0%
	% binnen .90	100,0 %	100,0%	100,0%
	% van Totaal	72,2%	27,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .411

Tak klankonderscheiding		.90-leerling niet	.90-leerling wel	Totaal
Onderste 25%	Aantal	45	42	87
	Verwacht Aantal	54,6	32,4	87,0
	% binnen toets	51,7%	48,3%	100,0%
	% binnen .90	22,4%	35,3%	27,2%
	% van Totaal	14,1%	13,1%	27,2%
Bovenste 75%	Aantal	156	77	233
	Verwacht Aantal	146,4	86,6	233,0
	% binnen toets	67,0%	33,0%	100,0%
	% binnen .90	77,6%	64,7%	72,8%
	% van Totaal	48,8%	24,1%	72,8%
Totaal	Aantal	201	119	320
	Verwacht Aantal	201,0	119,0	320,0
	% binnen toets	62,8%	37,2%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	62,8%	37,2%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .140

TAK klankarticulatie		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	41	36	77
	Verwacht Aantal	48,4	28,6	77,0
	% binnen toets	53,2%	46,8%	100,0%
	% binnen .90	20,4%	30,3%	24,1%
	% van Totaal	12,8%	11,3%	24,1%
Bovenste 75%	Aantal	160	83	243
	Verwacht Aantal	152,6	90,4	243,0
	% binnen toets	65,8%	34,2%	100,0%
	% binnen .90	79,6%	69,7%	75,9%
	% van Totaal	50,0%	25,9%	75,9%
Totaal	Aantal	201	119	320
	Verwacht Aantal	201,0	119,0	320,0
	% binnen toets	62,8%	37,2%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	62,8%	37,2%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .111

TAK-passieve woordkennis		.90-leerling		Totaal
		,00	1,00	
Onderste 25%	Aantal	26	57	83
	Verwacht Aantal	52,1	30,9	83,0
	% binnen toets	31,3%	68,7%	100,0%
	% binnen .90	12,9%	47,9%	25,9%
	% van Totaal	8,1%	17,8%	25,9%
Bovenste 75%	Aantal	175	62	237
	Verwacht Aantal	148,9	88,1	237,0
	% binnen toets	73,8%	26,2%	100,0%
	% binnen .90	87,1%	52,1%	74,1%
	% van Totaal	54,7%	19,4%	74,1%
Totaal	Aantal	201	119	320
	Verwacht Aantal	201,0	119,0	320,0
	% binnen toets	62,8%	37,2%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	62,8%	37,2%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .386

OBIS passieve woordkennis		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	24	54	78
	Verwacht Aantal	52,4	25,6	78,0
	% binnen toets	30,8%	69,2%	100,0%
	% binnen .90	11,5%	52,9%	25,1%
	% van Totaal	7,7%	17,4%	25,1%
Bovenste 75%	Aantal	185	48	233
	Verwacht Aantal	156,6	76,4	233,0
	% binnen toets	79,4%	20,6%	100,0%
	% binnen .90	88,5%	47,1%	74,9%
	% van Totaal	59,5%	15,4%	74,9%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .449

OBIS leesbegrip			.90-leerling		Totaal
			niet	wel	
dsomolb	Onderste 25%	Aantal	30	37	67
		Verwacht Aantal	45,0	22,0	67,0
		% binnen toets	44,8%	55,2%	100,0%
		% binnen .90	14,4%	36,3%	21,5%
		% van Totaal	9,6%	11,9%	21,5%
	Bovenste 75%	Aantal	179	65	244
		Verwacht Aantal	164,0	80,0	244,0
		% binnen toets	73,4%	26,6%	100,0%
		% binnen .90	85,6%	63,7%	78,5%
		% van Totaal	57,6%	20,9%	78,5%
Totaal		Aantal	209	102	311
		Verwacht Aantal	209,0	102,0	311,0
		% binnen toets	67,2%	32,8%	100,0%
		% binnen .90	100,0%	100,0%	100,0%
		% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .250

OBIS klankarticulatie		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	32	32	64
	Verwacht Aantal	43,0	21,0	64,0
	% binnen toets	50,0%	50,0%	100,0%
	% binnen .90	15,3%	31,4%	20,6%
	% van Totaal	10,3%	10,3%	20,6%
Bovenste 75%	Aantal	177	70	247
	Verwacht Aantal	166,0	81,0	247,0
	% binnen toets	71,7%	28,3%	100,0%
	% binnen .90	84,7%	68,6%	79,4%
	% van Totaal	56,9%	22,5%	79,4%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .187

OBIS klankonderscheiding		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	40	43	83
	Verwacht Aantal	55,8	27,2	83,0
	% binnen toets	48,2%	51,8%	100,0%
	% binnen .90	19,1%	42,2%	26,7%
	% van Totaal	12,9%	13,8%	26,7%
Bovenste 75%	Aantal	169	59	228
	Verwacht Aantal	153,2	74,8	228,0
	% binnen toets	74,1%	25,9%	100,0%
	% binnen .90	80,9%	57,8%	73,3%
	% van Totaal	54,3%	19,0%	73,3%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .244

OBIS letterkennis		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	34	33	67
	Verwacht Aantal	45,0	22,0	67,0
	% binnen toets	50,7%	49,3%	100,0%
	% binnen .90	16,3%	32,4%	21,5%
	% van Totaal	10,9%	10,6%	21,5%
Bovenste 75%	Aantal	175	69	244
	Verwacht Aantal	164,0	80,0	244,0
	% binnen toets	71,7%	28,3%	100,0%
	% binnen .90	83,7%	67,6%	78,5%
	% van Totaal	56,3%	22,2%	78,5%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .184

OBIS rekenbegrip		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	36	38	74
	Verwacht Aantal	49,7	24,3	74,0
	% binnen toets	48,6%	51,4%	100,0%
	% binnen .90	17,2%	37,3%	23,8%
	% van Totaal	11,6%	12,2%	23,8%
Bovenste 75%	Aantal	173	64	237
	Verwacht Aantal	159,3	77,7	237,0
	% binnen toets	73,0%	27,0%	100,0%
	% binnen .90	82,8%	62,7%	76,2%
	% van Totaal	55,6%	20,6%	76,2%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .221

OBIS geheugen		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	41	36	77
	Verwacht Aantal	51,7	25,3	77,0
	% binnen toets	53,2%	46,8%	100,0%
	% binnen .90	19,6%	35,3%	24,8%
	% van Totaal	13,2%	11,6%	24,8%
Bovenste 75%	Aantal	168	66	234
	Verwacht Aantal	157,3	76,7	234,0
	% binnen toets	71,8%	28,2%	100,0%
	% binnen .90	80,4%	64,7%	75,2%
	% van Totaal	54,0%	21,2%	75,2%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .171

OBIS sommen A		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	42	46	88
	Verwacht Aantal	59,1	28,9	88,0
	% binnen toets	47,7%	52,3%	100,0%
	% binnen .90	20,1%	45,1%	28,3%
	% van Totaal	13,5%	14,8%	28,3%
Bovenste 75%	Aantal	167	56	223
	Verwacht Aantal	149,9	73,1	223,0
	% binnen toets	74,9%	25,1%	100,0%
	% binnen .90	79,9%	54,9%	71,7%
	% van Totaal	53,7%	18,0%	71,7%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .261

OBIS kennis van getallen		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	48	36	84
	Verwacht Aantal	56,5	27,5	84,0
	% binnen toets	57,1%	42,9%	100,0%
	% binnen .90	23,0%	35,3%	27,0%
	% van Totaal	15,4%	11,6%	27,0%
Bovenste 75%	Aantal	161	66	227
	Verwacht Aantal	152,5	74,5	227,0
	% binnen toets	70,9%	29,1%	100,0%
	% binnen .90	77,0%	64,7%	73,0%
	% van Totaal	51,8%	21,2%	73,0%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .130

OBIS vormen		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	41	26	67
	Verwacht Aantal	45,0	22,0	67,0
	% binnen toets	61,2%	38,8%	100,0%
	% binnen .90	19,6%	25,5%	21,5%
	% van Totaal	13,2%	8,4%	21,5%
Bovenste 75%	Aantal	168	76	244
	Verwacht Aantal	164,0	80,0	244,0
	% binnen toets	68,9%	31,1%	100,0%
	% binnen .90	80,4%	74,5%	78,5%
	% van Totaal	54,0%	24,4%	78,5%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .067

OBIS sommen B		.90-leerling		Totaal
		niet	wel	
Onderste 25%	Aantal	28	41	69
	Verwacht Aantal	46,4	22,6	69,0
	% binnen toets	40,6%	59,4%	100,0%
	% binnen .90	13,4%	40,2%	22,2%
	% van Totaal	9,0%	13,2%	22,2%
Bovenste 75%	Aantal	181	61	242
	Verwacht Aantal	162,6	79,4	242,0
	% binnen toets	74,8%	25,2%	100,0%
	% binnen .90	86,6%	59,8%	77,8%
	% van Totaal	58,2%	19,6%	77,8%
Totaal	Aantal	209	102	311
	Verwacht Aantal	209,0	102,0	311,0
	% binnen toets	67,2%	32,8%	100,0%
	% binnen .90	100,0%	100,0%	100,0%
	% van Totaal	67,2%	32,8%	100,0%

Correlatie tussen al dan niet bij onderste 25% horen en al dan niet .90-leerling zijn = .303