

DE RELATIE TUSSEN LEES- EN LUISTERVAARDIGHEID;

EEN META-ANALYSE

Bert Meuffels & Huub van den Bergh

1. Relevantie

Er is betrekkelijk veel empirisch onderzoek verricht naar de relatie tussen lees- en luistervaardigheid. Op grond van de sterkte van de samenhang tussen deze twee vaardigheden proberen onderzoekers een antwoord te geven op de zowel in theoretisch als praktisch opzicht relevante vraag of er bij lezen en luisteren dezelfde dan wel verschillende intellectuele disposities betrokken zijn.

In theoretisch opzicht is deze vraag van belang voor onder meer de inrichting van taxonomische systemen van taalvaardigheden en - in uiterste instantie - voor de cognitieve structuur van de menselijke taalverwerking. In praktisch opzicht is deze vraag relevant voor onder andere de opbouw en inhoud van het taalvaardigheidsonderwijs en voor de inrichting van toetsbatterijen in, bij voorbeeld, grootschalig peilingsonderzoek als de PPOON (de Periodieke Peiling van het Onderwijsniveau; zie Kuhlmeier 1996). Mocht uit het verrichte onderzoek immers blijken dat lees- en luistertoetsen in feite min of meer equivalente indicatoren zijn van hetzelfde onderliggende construct (i.c. verbaal begrip), dan kan bij de meting van taalvaardigheden als lezen en luisteren volstaan worden met één, zo efficiënt mogelijke, operationalisatie van het bedoelde construct en hoeft men niet meer, zoals thans de praktijk is, elk onderscheiden aspect van beide vaardigheden apart via een scala van operationalisaties uitputtend te bestrijken. Ongetwijfeld een enorme besparing in termen van tijd, ontwikkelingskosten, afname- en scoringskosten, enzovoort.

Daarnaast kunnen op grond van de hoogte van de samenhang tussen lees- en luistertoetsen conclusies getrokken worden over mentale processen die aan beide vaardigheden ten grondslag liggen (Carroll 1993). Als de correlatie tussen lees- en luistervaardigheid bij voorbeeld 1.0 zou zijn, dan zijn beide vaardigheden niet onderscheidbaar, en doen lees- en luistervaardigheid vermoedelijk een beroep op dezelfde verwerkingsmechanismen. Dit impliceert dat onderzoek naar cognitieve processen voortbouwt c.q. kan voortbouwen op de resultaten van onderzoek naar de samenhang van vaardigheden (Carroll 1993: 194).

In de praktijk gaan onderzoekers ter beantwoording van de vraag of er bij lezen en luisteren dezelfde dan wel verschillende vaardigheden gemobiliseerd worden, in procedureel opzicht algemeen gesteld als volgt te werk: aan een steekproef proefpersonen wordt zowel een lees- als een luistertoets afgenomen, waarna de correlatie tussen de scores op beide toetsen berekend wordt. Indien deze correlatie, al dan niet gecorrigeerd voor meetfouten, substantieel afwijkt van 1.0, dan wordt in de regel de conclusie getrokken dat er bij lezen en luisteren - tot op zekere hoogte en gedeeltelijk althans - verschillende intellectuele vaardigheden in het geding zijn en dat dus beide vaardigheden in conceptueel opzicht een min of meer onafhankelijke status mag worden toegekend. Wijkt de berekende correlatie daarentegen niet significant af van 1.0, dan luidt de conclusie doorgaans dat lees- en luistervaardigheid empirisch niet onderscheidbaar zijn en dat het - alleen al

vanuit het oogpunt van spaarzaamheid - de voorkeur verdient beide variabelen met behulp van één construct (verbaal begrip) te representeren.

De analysetechnieken die gehanteerd zijn (en worden) ter beantwoording van de hier aan de orde zijnde vraag, blijven niet beperkt tot bivariate regressie-analyse, zoals hierboven is gesuggereerd: principale componentenanalyse, exploratieve factoranalyse, hypothese-toetsende factoranalyse, analyse van multi-trait-multi-method matrixen, analyse van covariantie-structuren, t-toetsen, variantie-analyses - kortom, een scala aan uiteenlopende dataat technieken is uitgeprobeerd om greep te krijgen op de dimensionaliteit van de betreffende taalvaardigheden.

De resultaten van het empirisch onderzoek naar de relatie tussen lezen en luisteren zijn, al het uitgevoerde onderzoek ten spijt, verre van consistent. In een aantal studies wordt een uitermate hoge correlatie aangetroffen en wordt op grond daarvan de conclusie getrokken dat lees- en luistervaardigheid in feite één ondeelbaar construct vormen (zie bij voorbeeld Sticht & James 1984), in andere studies daarentegen blijkt de correlatie laag (tot wel -.38) en wordt precies de tegenovergestelde conclusie afgeleid.

Voordat men nieuw onderzoek entameert naar de relatie tussen lees- en luistervaardigheid, zal toch op zijn minst enig inzicht verkregen moeten zijn in de vraag, waarom de resultaten van al het uitgevoerde correlatieve onderzoek nou zo sterk divergeren en nu eens voor de ene, dan weer voor de andere opvatting over de dimensionaliteit van lees- en luistervaardigheid lijken te pleiten. De gedachte dat de opbrengst van één enkel onderzoek - met als resultaat één of enkele correlaties, toegevoegd aan de honderden reeds verzamelde - de balans ten gunste van één van beide opvattingen over de dimensionaliteit van lees- en luistervaardigheid definitief zou doen doorslaan en de kwestie daarmee zou doen beslechten, is - gegeven de hier geschetste chaotische situatie - bepaald onrealistisch te noemen.

2. Problemen

Onderzoek naar de relatie tussen lees- en luistervaardigheid en de daaraan te verbinden conclusies ten aanzien van de (on)onderscheidbaarheid van deze twee taalvaardigheden wordt door een aantal factoren ernstig gecompliceerd; aan drie daarvan zullen we een korte bespreking wijden.

In de eerste plaats is uit onderzoek bij herhaling gebleken dat de ene leestoets (of om het even: luistertoets) de andere niet is. In verschillende studies worden in toetsen die nominaal dezelfde vaardigheid bestrijken, in feite geheel verschillende onderwerpen aan de orde gesteld en bevaagd. Maar niet alleen de onderwerpen kunnen per studie verschillen, ook de vraagvorm (bijvoorbeeld open vragen versus multiple choice), de betrouwbaarheid waarmee de vaardigheid in kwestie gemeten wordt, enzovoort. Al deze factoren kunnen een aanzienlijke invloed uitoefenen op datgene wat precies gemeten wordt (zie o.a. Kuhlmeier 1996; Bennett & Ward 1993).

Het vervelende gevolg van dit type divergenties is dat op grond van de scores op één of enkele lees- of luistervaardigheidstoetsen niet of nauwelijks generaliseerd kan worden naar de leesvaardigheid of de luistervaardigheid - een ondubbelzinnig antwoord op de vraag naar de dimensionaliteit van lees- en luistervaardigheid is dan ook uitgesloten.

Voor een verantwoordde generalisatie zou men idealiter een aselecte steekproef moeten kunnen trekken uit een gespecificeerde, welomschreven populatie van mogelijke operationalisaties van lees- en luistervaardigheid - helaas ontbreekt het tot nu toe aan een dergelijke, theoretisch gefundeerde domeinbeschrijving. Het gevolg hiervan is dat bij de samenstelling van een toetsbatterij een zekere vorm van willekeur, en dus van ongewisse externe validiteit, niet uitgesloten is. Zo maakt het, bij voorbeeld, heel wat uit of een proefpersoon wiens leesvaardigheid aan de hand van een leesvaardigheidstoets geschat wordt, een formulier moet invullen naar aanleiding van fictieve correspondentie, of dat hij een dienstregeling moet raadplegen, of een tekst van enige omvang moet lezen waarbij hij tekstbegripvragen in de vorm van vierkeuze-alternatieven moet beantwoorden, enzovoort. Wordt er systematisch en specifiek onderzoek naar deze zogeheten toetsspecifieke effecten verricht, dan blijken deze zich telkens, zowel bij de meting van de lees- als de luistervaardigheid - en trouwens ook bij de meting van de schrijfvaardigheid -, voor te doen, en wel in vrij sterke mate (Fiske 1987; Kuhlmeier 1996).

In aansluiting op het voorafgaande wordt een tweede complicerende factor gevormd door de omstandigheid dat, wil men althans de relatie tussen lees- en luistervaardigheid adequaat kunnen onderzoeken, aan proefpersonen in feite niet twee geheel verschillende toetsen zouden moeten worden voorgelegd (i.c. de één voor de meting van de leesvaardigheid, de ander voor luistervaardigheid), maar juist één en dezelfde toets; deze toets zou idealiter 'uitwisselbaar' moeten zijn, d.w.z. de verschillen tussen de lees- en luistertoets zouden louter en alleen de wijze van aanbidding (visueel dan wel auditief) moeten betreffen. Juist om alternatieve verklaringen met betrekking tot toetsspecifieke effecten uit te kunnen sluiten, zouden in beide toetsen het aan de orde gestelde onderwerp, de vraagvorm, enzovoort, hetzelfde moeten zijn. Wanneer proefpersonen echter twee maal dezelfde tekst voorgeschoteld krijgen (één keer schriftelijk en één keer auditief), en over diezelfde tekst twee maal precies dezelfde vragen moeten beantwoorden, dan gaan geheugeneffecten e.d. vanzelfsprekend een rol spelen, hetgeen ongetwijfeld leidt tot een overschatting van de samenhang tussen lees- en luistervaardigheid.

Een derde complicerende factor wordt gevormd door de aard van het beslissingscriterium, i.c. de maatstaf die door onderzoekers gehanteerd wordt om de conclusie te funderen dat lezen en schrijven beide een beroep doen op dezelfde dan wel op verschillende intellectuele vaardigheden. Veelal komt die in concreto neer op de numerieke hoogte van de correlatie tussen de scores op de lees- en luistertoets. De hoogte van een correlatiecoëfficiënt is echter - onder meer - afhankelijk van de betrouwbaarheid van de beide toetsen. Indien de betrouwbaarheid van de lees- en luistertoets in beide gevallen bij voorbeeld .70 bedraagt, dan kan de correlatie tussen die twee toetsen nooit hoger uitvallen dan .70. Het theoretische maximum van 1.0 kan bij niet-perfect betrouwbare metingen niet gehaald worden. Om in dit verband onverantwoorde conclusies te voorkomen, wordt de correlatiecoëfficiënt daarom niet zelden gecorrigeerd voor attenuatie (i.c. voor de onbetrouwbaarheid van beide toetsen). Vervolgens wordt getoetst of deze voor attenuatie gecorrigeerde correlatiecoëfficiënt al dan niet significant afwijkt van 1.0.

De vraag of met deze correctie voor attenuatie wel een goede schatting wordt verkregen van de 'ware' correlatie, is echter afhankelijk van de wijze waarop de betrouwbaarheid van beide toetsen geschat wordt. In de regel wordt voor die schatting coëfficiënt α (Cronbach's α) gebruikt; deze interne consistentie maat is echter, zoals uit de psychometrische vakliteratuur genoegezaam bekend is,

een onderschatting van de 'werkelijke' betrouwbaarheid (Lord & Novick 1962: 91; 135). Het gevolg daarvan is dat de voor attenuatie gecorrigeerde correlatiecoëfficiënt een overschatting is van de werkelijke samenhang tussen de beide toetsen (zie noot 1). Hoe sterk die overschatting is, valt in een concreet geval niet te zeggen.

Met het oog op het hier aan de orde zijnde probleem: de relatie tussen lees- en luistervaardigheid, lijkt een minder strikt statistisch, meer inhoudelijk gericht beslissingscriterium waarin recht gedaan wordt aan het feitelijk gegeven dat de ene leestoets (of luistertoets) nu eenmaal de andere niet is, meer adequaat. De correlatie tussen verschillende leestoetsen zal, uitgaande van die gedachte, onveranderlijk afwijken van 1.0 (ongeacht de vraag of wel of niet gecorrigeerd is voor onbetrouwbaarheid); hetzelfde geldt voor de luistertoetsen. Als beslissingscriterium voor de (on)onderscheidbaarheid van lees- en luistervaardigheid fungeert vervolgens de verhouding van de correlatie tussen lees- en luistertoetsen ten opzichte van de correlatie tussen verschillende leestoetsen onderling en de correlatie tussen verschillende luistertoetsen onderling, en wel in dier voege dat indien de correlatie tussen lees- en luistertoetsen even hoog is als de wortel uit het product van de correlatie tussen verschillende leestoetsen onderling en die tussen verschillende luistertoetsen onderling, dat dan gesproken kan worden van één construct².

Gezien het bovenstaande commentaar zal duidelijk zijn dat één enkel onderzoek praktisch gezien nimmer uitsluit kan geven over de vraag of er bij lees- en luistervaardigheid een beroep moet worden gedaan op één en dezelfde vaardigheid dan wel of bij beide vaardigheden verschillende intellectuele disposities betrokken zijn. In lijn met de bovenstaande opmerkingen over complicerende factoren kan wellicht enig inzicht in deze netelige problematiek betreffende de conceptuele relatie tussen beide vaardigheden verkregen worden door niet één of enkele, doch een zo groot mogelijk aantal correlaties tussen lees- en luistertoetsen die in de literatuur gerapporteerd worden, aan de hand van een meta-analyse te beschrijven en - zo mogelijk - te verklaren.

3. Gegevens en analyse

Aan de hand van geautomatiseerde zoektochten door verschillende literatuurbanken (waaronder ERIC, Psychological Abstracts, Psychlit) zijn voor de periode 1966 - 1996 studies gezocht waarin de relatie tussen leesvaardigheids- en luistervaardigheidstoetsen onderling, of de relatie tussen beide typen toetsen centraal staat. Deze zoektochten resulteerden in 95 verschillende studies, waaruit in totaal 976 correlatiecoëfficiënten gedestilleerd konden worden (zie tabel 1).

In lang niet alle studies werden er ook correlatiecoëfficiënten gerapporteerd: ook factorladingen, F-ratio's, t-toetsen, effectgroottes en dergelijke bleken, afhankelijk van de specifieke aard van de analyse, te fungeren als rapportage-eenheid. Aangezien in de onderhavige meta-analyse de (hoogte van de) correlatiecoëfficiënt als basale afhankelijkheidsvariabele zal fungeren, werden al deze 'afwijkende' beschrijvende statistieken getransformeerd tot simpele correlatiecoëfficiënten³.

Per studie werden vervolgens de volgende vijf variabelen geannoteerd:

- leeftijd van de proefpersonen;

- percentage jongens in de steekproef;
- correlatie wel versus niet gecorrigeerd voor attenuatie;
- 1^e versus 2^e taal (d.w.z. wordt de lees- of luistertoets afgenomen in de moedertaal of in een vreemde, op latere leeftijd onderwezen taal?);
- open versus gesloten vragen.

Deze vijf variabelen - andere dan deze werden er in de regel niet gerapporteerd, zodat onze keuze noodgedwongen tot deze vijf beperkt moest blijven - zullen in de onderhavige meta-analyse als potentiële predictoren van de variabiliteit in correlatie fungeren. Zie voor een beschrijving van het verzamelde materiaal en de daaruit afgeleide predictoren tabel 1.

Tabel 1. *Aantal studies (N_s), aantal correlaties (N_r) en (gemiddeld) aantal proefpersonen (N_{pp}) per afhankelijke variabele; beschrijvende grootheden voor de vijf predictoren (Lft: gemiddelde leeftijd proefpersonen; % Att: percentage studies waarin correctie voor attenuatie plaatsvond; 1^eTaal: aantal studies waarin de toetsen in de eerste taal afgenomen zijn; % OV: aantal studies met alleen open vragen).*

Relatie	N _s	N _r	N _{pp}	% Att	1 ^e Taal	Lft	% OV
Lezen - Lezen	28	317	116	18	97	12.7	38
Lezen - Luisteren	44	330	78	9	88	13.1	12
Luisteren - Luisteren	23	329	92	16	85	12.7	45

Dat de theoretische status van de vijf predictoren verschilt, is evident. De variabele 'correctie voor attenuatie' bij voorbeeld vervult in het onderhavige onderzoek veeleer de functie van externe controle-variabele bij de uit te voeren statistische analyse (i.c. een multivariate regressie-analyse): het vertrouwen in de analyse neemt immers toe, indien de geschatte correlaties bij correctie hoger uitvallen dan bij niet-correctie.

De predictor 'leeftijd' daarentegen is in theoretisch opzicht veel interessanter, en tot deze predictor zullen we de beschouwingen, gezien de beperkte ruimte, dan ook grotendeels beperken. Is de relatie tussen lees- en luistervaardigheid invariant over leeftijd, of verandert deze afhankelijk van de leeftijd zoals Sticht (1984) in een tentatieve meta-studie suggereert?

In de uit te voeren statistische analyses zijn drie afhankelijke variabelen relevant: de correlaties tussen leestoetsen onderling, de correlaties tussen luistertoetsen onderling en de correlaties tussen lezen en luisteren. Om verschillende redenen (redenen betreffende onder meer stochastische afhankelijkheid van de drie afhankelijke variabelen) verdient het de voorkeur deze drie afhankelijke variabelen niet separaat, maar simultaan te analyseren. Daartoe is gebruik gemaakt van een multivariaat model. Bovendien lijkt het alleszins aannemelijk dat de geobserveerde correlaties binnen één studie, door diverse factoren, meer overeenkomst vertonen dan de correlaties tussen studies (i.e. de observaties zijn immers, in statistische zin, niet onafhankelijk). Vandaar dat uitgegaan wordt van een multivariaat multivariate regressie-analyse.

In dit multivariate model werd per afhankelijke variabele het gemiddelde geschat alsmede twee variantiecomponenten: de eerste variantiecomponent heeft betrekking op de variantie tussen de correlatiecoëfficiënten binnen studies, de tweede component betreft de variantie tussen studies. Het

spreekt voor zich dat in het kader van de onderhavige meta-analyse de aandacht speciaal uitgaat naar deze tweede variantiecomponent: kan de variantie in deze component systematisch worden toegeschreven aan de variantie in de vijf onderscheiden predictor, of moeten de verschillen in correlatie tussen studies onderling 'onverklaard' blijven?

4. Resultaten

De op basis van het toegepaste multivariate model geschatte gemiddelde correlaties en de mate waarin deze van studie tot studie verschillen, staan samengevat in tabel 2.

Tabel 2. *Geschatte gemiddelde correlatie tussen leestoetsen onderling, tussen luistertoetsen onderling, en tussen lees- en luistertoetsen; de standaarddeviatie tussen studies en binnen studies).*

Correlatie tussen:	Standaarddeviatie	
	gemiddelde	tussen studies
Leestoetsen	.60	.18
Lees- en luistertoetsen	.54	.18
Luistertoetsen	.57	.16

Uit tabel 2 blijkt dat de (geschatte) gemiddelde correlatie tussen verschillende leestoetsen .60 bedraagt, en die voor luisteren .57. De gemiddelde correlatie tussen de verschillende lees- en luistertoetsen bedraagt .54. Toepassing van het in paragraaf 2 ontwikkelde beslissingscriterium dat wil leesvaardigheid in conceptueel opzicht onderscheiden kunnen worden van luistervaar-digheid - de correlatie tussen lezen en luisteren lager moet zijn dan (de wortel uit het product van) de correlatie tussen verschillende leestoetsen onderling en die tussen verschillende luistertoetsen onderling, leidt tot de conclusie dat lees- en luistervaardigheid in empirische zin wel degelijk onderscheidbaar zijn, zij het dat de overlap bij beide vaardigheden aanzienlijk is (zie noot 2).

Binnen studies zijn de verschillen in correlatiecoëfficiënten aanzienlijk. Een verschil van bijvoorbeeld .34 (twee standaarddeviaties) tussen verschillende leestoetsen in één studie is niet ongewoon. Hetzelfde geldt voor de verschillen binnen studies in de hoogte van de correlatiecoëfficiënten tussen luistertoetsen onderling, of voor de hoogte van correlatiecoëfficiënten tussen lees- en luistertoetsen.

Ook tussen de verschillende studies echter blijkt er een aanzienlijke variabiliteit te bestaan in de hoogte van de correlaties die worden aangetroffen tussen leestoetsen, luistertoetsen en lees- en luistertoetsen: de spreiding bedraagt liefst .18 voor zowel de correlatie tussen verschillende leestoetsen als die tussen lees- en luistertoetsen; de spreiding van de correlaties voor verschillende luistertoetsen doet daar maar weinig voor onder (.16). Naar alle waarschijnlijkheid mag men uit deze overeenkomstige verschillen in correlaties binnen studies en tussen studies afleiden dat verschillen

ten gevolge van verschillende steekproeven (van verschillende leeftijd), verschillende talen, enzovoort (grotweg) net zo relevant zijn als verschillen ten gevolge van specifieke operationalisaties.

Op grond van deze spreidingen kunnen waarschijnlijkheidsintervallen voor de betreffende correlatiecoëfficiënten geschat worden. Volgens deze schattingen blijkt in 90% van de studies⁴ de correlatiecoëfficiënt tussen lezen en luisteren te variëren tussen .24 tot .84. Voor de correlaties tussen leestoetsen zijn de grenzen van dit 90% - waarschijnlijkheidsinterval .30 en .90, en voor luisteren respectievelijk .31 en .83. Voor de vraag of lees- en luistervaardigheid een aparte status kan worden toegerekend, maakt het kennelijk heel wat uit op welke specifieke studie men zijn conclusies wenst te baseren.

Kunnen de verschillen in hoogte van de correlatiecoëfficiënten ook worden toegeschreven aan de verschillen tussen de studies, althans wanneer die verschillen betrekking hebben op de leeftijd van de proefpersonen, de taal waarin getoetst wordt, het percentage jongens in de steekproef, enzovoort, kortom betrekking hebben op de vijf onderscheiden predictor? Toepassing van een (multivariaat) multiple regressiemodel (zie o.a. Bryke & Raudenbush 1992: 155 e.v.), waarin de effecten van de vijf onderzochte predictor op de correlaties geschat worden, leert dat de betreffende verschillen voor een aanzienlijk deel kunnen worden toegeschreven aan de vijf variabelen (i.c. predictor) waarin de studies verschillen (zie tabel 3).

De gegevens in tabel 3 moeten als volgt gelezen worden: het intercept geeft telkens de correlatie voor de drie afhankelijke variabelen voor leerlingen - en daar gaat het hier om - met een leeftijd van dertien jaar. Voor deze groep van dertienjarigen bedraagt de correlatie tussen leestoetsen dus .48, de correlatie tussen de luistertoetsen .41, en die tussen lees- en luistertoetsen .43. Het intercept is hier met opzet 'gefixeerd' voor leerlingen met een leeftijd van dertien jaar: deze leeftijd vormt immers bij benadering de gemiddelde leeftijd van de proefpersonen in het door ons onderzochte materiaal. Voor interpretatieve doeleinden geldt het hier gefixeerde intercept dus als base-line, waartegen de resultaten van andere leeftijdsgroepen afgezet en gecontrasteerd moeten worden. Hoe staat het nu met de drie betreffende correlaties bij andere proefpersonen dan dertienjarigen?

De correlatie tussen verschillende leestoetsen neemt .03 af met elk jaar dat de in de steekproef opgenomen proefpersonen ouder zijn dan dertien, en 0.3 toe met elk jaar dat de proefpersonen jonger zijn. Op tienjarige leeftijd bij voorbeeld bedraagt de correlatie tussen verschillende leestoetsen $(.48 - [(10-13) * -.03] =) .57$, op zestienjarige leeftijd daarentegen bedraagt deze 'slechts' .39.

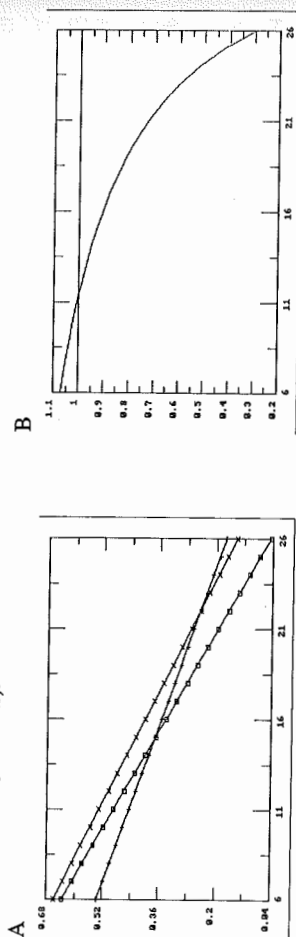
Hetzelfde type interpretatie is van toepassing op de correlatie tussen verschillende luistertoetsen onderling, en op die tussen lees- en luistertoetsen. In alle drie de gevallen is er dus een duidelijk, aan toename in leeftijd gekoppelde, dalende trend in de correlaties waarneembaar. In figuur 1A is die (geschatte) daling van de drie correlaties, conditioneel aan de toename in leeftijd, grafisch weergegeven.

Tabel 3 Effecten van de vijf predictoren op de hoogte van de correlatiecoëfficiënten tussen leestoetsen onderling, luistertoetsen onderling en lees- en luistertoetsen.

	Correlatie tussen:		
	Lezen - lezen	Lezen - luisteren	Luisteren - luisteren
Intercept	.48	.43	.41
Leeftijd ¹	-.03	-.03	-.02
Proportie jongens ¹	-.02	-.04	-.04
Correctie attenuatie	.47	.40	.53
2 ^e taal	.13*	.15	.15*
Gesloten vragen	-.02	.02	.09

1: gecentreerd rond 13 jaar; 2: gecentreerd rond 50%; * niet significant (p > .05).

Figuur 1 De veranderingen in de hoogte van de correlaties tussen lees- en luistertoetsen met leeftijd (A; x: correlatie tussen leestoetsen; +: correlatie tussen luistertoetsen; □: correlatie tussen lees- en luistertoetsen) en de correlatie tussen lezen en luisteren in relatie tot de correlatie tussen leestoetsen onderling en luistertoetsen onderling (B; zie noot 2).



Uit figuur 1A blijkt dat de correlatie tussen leestoetsen varieert van .67 bij zesjarigen tot (±) .15 bij 26-jarigen. Ook bij de luistertoetsen neemt de correlatie af met leeftijd, zij het dat in dit geval de daling wat minder geproforceerd is dan bij lezen het geval is: de betreffende correlatie neemt af van (±) .53 bij zesjarigen tot (±) .25 voor 26-jarigen. Ook de samenhang tussen de lees- en luistertoetsen neemt af met leeftijd, van ongeveer .65 bij zesjarigen tot iets boven de 0 (!) bij 26-jarigen⁵.

De vraag naar de (van leeftijd kennelijk afhankelijk) correlatie tussen enerzijds leestoetsen en anderzijds luistertoetsen in verhouding tot de correlatie tussen leestoetsen onderling respectievelijk luistertoetsen onderling, kortom de vraag in welke mate de gegevens beantwoorden aan het in paragraaf 2 geformuleerde beslissingscriterium ten aanzien van de dimensionaliteit van lees- en luistervaardigheid, wordt beantwoord in Figuur 1B (zie hiervoor: noot 2): de relatie tussen

lezen en luisteren blijkt duidelijk te divergeren naarmate de leeftijd van de proefpersonen toeneemt, althans vanaf een bepaald punt: tot ongeveer elf jaar is er sprake van een bijna perfect verband tussen lezen en luisteren, d.w.z. de (gemiddelde) correlatie tussen lees- en luistertoetsen wijkt tot omstreeks elf jaar nauwelijks af van de gemiddelde correlatie tussen leestoetsen onderling en tussen luistertoetsen onderling. Hierna gaan beide vaardigheden echter duidelijk hun eigen weg: vanaf dat moment is er sprake van een gescheiden ontwikkeling van lees- en luistervaardigheid. Hoe zouden deze bevindingen inhoudelijk verklaard kunnen worden?

Eén mogelijke verklaring is van cognitieve aard: tot elf, twaalf jaar bevinden de betreffende twee taalvaardigheden zich nog in een ontwikkelingsstadium; lees- en luistervaardigheid hebben het volwassen niveau van beheersing nog niet bereikt en zijn derhalve nog niet uitgekristalliseerd. Verschillen in mogelijke beheersing van die lees- of luistervaardigheid binnen een persoon - en dus afname van de betreffende correlatie - treden nog nauwelijks aan het licht. Vanaf elf, twaalf jaar echter, na het verlaten van de basisschool, hebben de meeste leerlingen de betreffende twee taalvaardigheden redelijk onder de knie en kan een verdere uitbouw en differentiatie plaatsvinden. Merk op dat deze cognitieve verklaring precies tegengesteld is aan die van Sticht & James (1984), die in een meta-studie vinden dat de correlatie (geheel anders en precies tegengesteld aan de resultaten van ons onderzoek) juist toeneemt met de leeftijd, maar geheel in lijn is met de zogenaamde differentiatiehypothese (Garett 1946; Fergusson 1954, 1956; Hoeks, Molenaar & Mellenbergh 1990).

Cognitieve verklaringen als bovenstaande, ongeacht de precieze aard en uitwerking daarvan, zijn echter naar ons oordeel in dit stadium van het onderzoek naar de relatie tussen lees- en luistervaardigheid prematuur. Zulke soort verklaringen die een zware wissel trekken op ontwikkelingspsychologische inzichten, zijn pas zinvol en het overwegen waard als eerst alternatieve, meer 'plavloerse' verklaringen van de gevonden trends in de data zijn uitgesloten - en uitgesloten is dit type meer 'triviale' verklaringen in het onderhavige geval beslist niet. Instrument-bias (en de daarmee onvermijdelijk gepaard gaande toetsspecificiteit) kan in dit geval immers een heel aannemelijke verklaring vormen voor bij voorbeeld het feit, dat zowel de correlatie tussen de leestoetsen onderling als die tussen de luistertoetsen onderling consistent afneemt met de leeftijd: hoe ouder proefpersonen worden, hoe moeilijker het voor toetsconstructeurs doorgaans is om toetsen te ontwikkelen die betrouwbaar differentiëren tussen de onderzochte proefpersonen. Een van de meest efficiënte manieren om die differentiatie te bewerkstelligen en daarmee meetfouten te reduceren, is de aard van het te bevragen onderwerp complex te maken, bij voorbeeld door onderwerpen aan te bieden en te bevragen waarmee vrijwel niemand van de onderzochte proefpersonen al enige kennis of ervaring heeft opgedaan. Het onvermijdelijk gevolg van dit soort instrument-bias is een reductie van de correlatie tussen de lees- en luistertoetsen onderling, en - generaliserend - tevens een reductie van de correlatie tussen enerzijds de leesvaardigheidstoetsen en anderzijds de luistervaardigheidstoetsen.

Wat de invloed van de overige predictoren betreft, merken we hier in dit kort bestek enkel op dat ook de samenstelling van de steekproef, de correctie voor attenuatie en, zij het wat minder eenduidig, de taal waarin getoetst wordt, elk een substantiële invloed uitoefenen op de hoogte van de drie correlatiecoëfficiënten. We volstaan met de constatering dat met name het - overigens

voorspelde - effect van de correctie voor attenuatie op de hoogte van de correlaties een duidelijke steun vormt voor de stabiliteit van de hier gerapporteerde bevindingen.

5. Discussie

De aard van de hier uitgevoerde meta-studie kan wellicht vraagtekens en kritiek oproepen: anders dan in meta-studies gebruikelijk is, worden in het onderhavige onderzoek de resultaten van alle empirisch onderzoek naar de relatie tussen lees- en luistervaardigheid op één hoop geveegd, en dat zonder enige vorm van kwalitatieve discriminatie of differentiatie tussen de studies onderling. Het ene onderzoek is het andere toch niet?

Meta-studies hebben in de regel het karakter van een 'horse race' studie: om in de tegenstrijdigheid aan onderzoeksresultaten en in de chaotische wirwar aan conclusies op een bepaald onderzoeksveld enige helderheid te scheppen, tracht men in de veelheid van het uitgevoerde empirisch onderzoek ter zake het kaf van het koren te scheiden door successtevelijk steeds stringenter beoordelingscriteria voor de kwaliteit van onderzoek aan te leggen, totdat tenslotte één of enkele substantiële onderzoeken overblijven die de toets der kritiek wél kunnen doorstaan; op de resultaten van deze 'happy few' worden dan voornamelijk de conclusies gebaseerd (zie voor een voor het vak Taalbeheersing relevant voorbeeld van dit type meta-studie: Tordoir en Westdorp (1979); in deze horse-race studie worden de effecten van het traditionele grammatica-onderwijs op aspecten van de taalbeheersing (schrijf- en leesvaardigheid nader geanalyseerd). Horse-race studies worden overigens niet alleen uitgevoerd om methodologisch verantwoordere inhoudelijke inzichten op een bepaald terrein te verwerven - hoe zit het nu 'werkelijk'?, - ze fungeren vaak ook als opmaat voor nieuw te entameren onderzoek waarin getracht wordt de geconstateerde feiten in het eerder uitgevoerde onderzoek te vermijden.

Een meta-studie van het type 'horse race' is voor het onderhavige onderzoeksveld praktisch gezien niet goed uitvoerbaar, en theoretisch gezien evenmin wenselijk. Praktisch gezien stuit men op het probleem dat de traditioneel in meta-studies aangelegde beoordelingscriteria, die in de regel methodologisch en/of statistisch van aard zijn en door de bank genomen toegesmeden zijn op onderzoeksontwerpen van experimentele aard, op het onderhavige niet-experimentele onderzoeksterrein onvoldoende discrimineren. Na achtereenvolgende toepassing van criteria als 'onderscheidingsvermogen van de statistische toets', 'adequatheid van statistische analyse', 'effectgrootte', 'scope van de studie', enzovoort, blijven er veel te veel studies over om een helder, consistent zicht te krijgen op de hier centraal gestelde relatie tussen lees- en luistervaardigheid.

Een andere, belangrijker reden om doelbewust af te zien van het elimineren van een aantal studies op grond van externe criteria betreft de omstandigheid, dat in een dergelijk opgezette kwalitatieve meta-studie in het geheel geen recht gedaan wordt aan het in paragraaf 2 aan de orde gestelde probleem van toetspecificiteit en de daarmee gepaard gaande onzekerheid in externe validiteit - sterker, dat probleem wordt geheel ontkend en wordt juist door het terzijde schuiven van onderzoek dat niet aan de gestelde maatstaven zou voldoen, in ernstige mate geggraveerd. Juist met het oog op die altijd aanwezige toetsspecificiteit is het van fundamenteel belang om, zoals in de

onderhavige studie is gebeurd, zoveel mogelijk onderzoek in de beschouwingen te betrekken, en juist niet zo min mogelijk, zoals het doel van horse race studies is.

Of men nu in kwalitatieve zin tussen studies onderscheidt of niet, het belang van meta-studies als de onderhavige is primair gelegen in het gegeven dat met behulp daarvan in beginsel een meer systematische kijk op relaties in de werkelijkheid mogelijk wordt. Dat kwalitatieve verschillen tussen studies (zoals operationalisaties) van even grote invloed zijn op de uiteindelijke correlaties, als meer kwantitatieve verschillen is opmerkelijk. Duidelijk komt immers uit de hier uitgevoerde meta-analyse naar voren dat de relatie tussen lezen en luisteren geen chaotisch, grillig en onvoorspelbaar verloop vertoont, integendeel: die relatie blijkt onder meer conditioneel afhankelijk te zijn van de leeftijd.

Noten

1. Deze correctie van een correlatiecoëfficiënt voor onbetrouwbaarheid, of attenuatie, wordt als volgt uitgevoerd:

$$r'_{xy} = \frac{r_{xy}}{\sqrt{p_x * p_y}}$$

waarbij r_{xy} de correlatie tussen de scores op toets x en toets y is, met p_x en p_y de betrouwbaarheid van respectievelijk toets x en y wordt aangegeven, en r'_{xy} tenslotte, de gecorrigeerde correlatiecoëfficiënt is.

2. Stel, r_{lezu} is de correlatie tussen lees- en luistertoetsen, r_{ez} is de correlatie tussen leestoetsen, en r_{lu} is de correlatie tussen luistertoetsen. Een meer formele weergave van het opgestelde criterium is dan:

$$r'_{lezu} = \frac{r_{lezu}}{\sqrt{r_{lez} * r_{lui}}}$$

Indien r'_{lezu} gelijk is aan 1.0, dan is de correlatie tussen lezen en luisteren net zo hoog als de correlatie tussen verschillende leestoetsen of verschillende luistertoetsen. Een onderscheid tussen beide constructen is dan empirisch gezien betekenisloos.

3. Enkele voorbeelden van conversie van toetsingsgrootheden naar correlatiecoëfficiënten (t: toetsingsgrootte; F: F-ratio; χ^2 : toetsingsgrootte χ^2 ; d: Cohen's effectgrootte; df: vrijheidsgraden; n: steekproefgrootte):

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

$$r = \sqrt{\frac{F}{F + df \text{ (ruis)}}}$$

$$r = \sqrt{\frac{\chi^2}{n}}$$

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

4. De grenzen van een waarschijnlijkheidsinterval kunnen eenvoudig bepaald worden als:

$$r'_{xy} = r_{xy} \pm z_{90} * SD_{r_{xy}} = r_{xy} \pm 1.65 * SD_{r_{xy}}$$

waarbij het minteken gebruikt wordt voor de bepaling van de linker, en het plusteken voor de bepaling van de rechter betrouwbaarheidsgrens.
Zonder uitzondering bleken nonlineaire relaties tussen de hoogte van de correlatiecoëfficiënten en leeftijd niet aantoonbaar.

5.

Literatuur

- Abbott, R.D. & V.W. Berninger (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology*, 85, 478-508.
- Assessment of Performance Unit (1985). *English language in Scottish schools: A monitoring project*. Edinburgh.
- Bachman, L.F. (1982). The trait structure of cloze test scores. *TESOL Quarterly*, 16, 61-70.
- Bachman, L.F. & A.S. Palmer (1981). The construct validation of the FSI oral interview. *Language Learning*, 31, 67-86.
- Bachman, L.F. & A.S. Palmer (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16, 449-464.
- Balke, G. (1991). *Multilevel factor analysis of proficiency in English as a foreign language*. Paper presented at the annual meeting of the American Educational Research Association on April 3-7 in Chicago, Ill. (ED 334 216).
- Bennett, R.E. & Ward, W.C. (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment*. Hillsdale, NJ: Lawrence Erlbaum Ass.
- Blok, H. (1987). *Taal voor alle dagen: Feiten en meningen over het taalgebruik van LBO- en MAVO-leerlingen in alledaagse situaties* (Diss.). Den Haag: SVO.
- Bostrom, R.N. & E.S. Waldhart (1988). Memory models and the measurement of listening. *Communication Education*, 37, 1-13.
- Boyle, J.P. (1987). Listening comprehension a separate factor in language ability? *IRAL*, 25, 238-256.
- Bryke, A.S. & Kaudenbush, S.W. (1992). *Hierarchical linear models: Applications and data analysis methods* (Advanced quantitative techniques in the social sciences 1). Newbury Park: Sage Publications.
- Brumbaek, R.A. (1989). Comparison of reading and listening-reading techniques for administration of SAT reading comprehension subtest: justification for the bypass approach. *Perceptual and Motor Skills*, 68, 1015-1018.
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42, 313-357.
- Carroll, J.B. (1941). A factor analysis of verbal abilities. *Psychometrika*, 6, 279-307.
- Carroll, J.B. (1962). Factors of verbal achievement. *Proceedings of the 1961 invitational conference on testing problems*. Princeton, NJ: Educational Testing Service.
- Carroll, J.B. (1983). Psychometric theory and language testing. In: J.W. Oller (ed.), *Issues in Language Testing Research*. Rowley, Mass.: Newbury House, 80-107.
- Carroll, J.B. (1993). *Human cognitive abilities a survey of factor-analytic studies*. Cambridge: University press.
- Carlisle J.F. & Felbinger L. (1991). Profiles of listening and reading comprehension. *Journal of Educational Research*, 84, 345-354.
- Clark, M.L. (1979). *Hierarchical structure of comprehension skills; vol.2: Factorial and 'smallest space analysis' of primary reading and listening test correlations: an empirical study of grade 7 children's performance*. Australian Council for Educational Research.
- Crockett, D.J. (1974). Component analysis of within correlations of language-skill tests in normal children. *The Journal of Special Education*, 8, 361-375.
- Curtis M.E. (1980). Development of components of reading skill. *Journal of Educational Psychology*, 72, 656-669.
- Duker, S. (1965). Listening and reading. *The Elementary School Journal*, 65, 321-329.
- Duker, S. (1964). *Listening biography*. New York: Scarecrow Press.
- Fergusson, G.A. (1954). On learning and human ability. *Canadian Journal of Psychology*, 8, 95-112.
- Fergusson, G.A. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, 10, 121-131.
- Fiske, D.W. (1987). Construct invalidity comes from method effects. *Educational and Psychological Measurement*, 47, 285-307.
- Garrett, H.E. (1946). A developmental theory of intelligence. *American Psychologist*, 1, 372-378
- Gelderen, A. van (1987). *Taalmaten. Deel 3: Relaties tussen spreek-, schrijf-, lees- en luisterprestaties*. Amsterdam: SCO.
- Hauser, M. & A. Hughes (1987). A factor analytic study of four listening tests. *Journal of the Listening Association*, 129-147.
- Hinofitis, F.B. (1980). Cloze as an alternative method of ESL placement and proficiency testing. In: J.W. Oller & K. Perkins (eds.), *Research into language testing*. Rowley, Mass.: Newbury House, 121-128
- Hoeks, J., Molenaar, P.C.M. & Mellenbergh, G.J. (1990). Fitting linear models to semantic tests constructed to Guilford's facet design. *Multivariate Behavioural Research* (in press).
- Horowitz, M.W. & A. Berkowitz (1967). Listening and reading, speaking and writing: an experimental investigation of differential acquisition and reproduction of memory. *Perceptual and Motor Skills*, 24, 207-215.
- Hosley, D. & K. Meredith (1979). Inter- and intra- test correlates of the TOEFL. *TESOL Quarterly*, 13, 209-217.
- Jong, J.H.A.L. de (1987). Defining tests for listening comprehension. In: A. Valdman (ed.) *Proceedings of the symposium on the evaluation of foreign language proficiency*. Bloomington, Indiana Univers. Found.
- Jong, J.H.A.L. de (1988). Rating scales and listening comprehension. *Australian Review of Applied Linguistics*, 11, 73-85.
- Kuhlemeier, H. (1996). *Taakvaardigheid, taalactiviteiten en taalattitudes: Een validatiestudie* (Diss.). Arnhem: Cito.
- Lord F.M. & Novick, R.M. (1962). *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley Publishing Company.
- Lumley, T. (1993). The notion of subskills in reading comprehension tests: an EAP example. *Language Testing*, 10, 211-234
- Oller, J.W. (1979). *Language tests at school*. London: Langman.
- Oller, J.W. & F.B. Hinofitis (1980). Two mutually exclusive hypotheses about second language ability: indivisible or partially divisible competence. In: J.W. Oller & K. Perkins (eds.), *Research in language testing*. Rowley, Mass.: Newbury House, 13-23.
- Ross, R. (1964). A look at listeners. *The Elementary School Journal*, 64, 369-372
- Rost, D.H. (1993). Assessing different components of reading comprehension: fact or fiction? *Language Testing*, 10, 79-91.
- Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning*, 43, 313-344.
- Scholtz, G. e.a. (1980). Is language ability divisible or unitary? A factor analysis of 22 english proficiency tests. In: J.W. Oller & K. Perkins (eds.), *Research in Language Testing*. Rowley, Mass.: Newbury House, 24-33.
- Schreuder, K. (1987). *Hörverstandnistests und Lesefertigkeit*. Paper gepresenteerd op het 18^e congres voor Angewandte Linguistik Und Computer.
- Sinatra, G.M. (1990). Convergence of listening and reading processing. *Reading Research Quarterly*, 25, 115-130.
- Sticht T.G. & James, J.H. (1984). Listening and reading. In: P.D. Pearson (Ed.), *Handbook of reading research*. New York: Longman, 293-317
- Stump, T.A. (1978). Cloze and dictation tasks as predictors of intelligence and achievement scores. In: J.W. Oller & K. Perkins (eds.), *Language in education - testing the tests*. Rowley, Mass.: Newbury House, 36-64.
- Tordoir, A. & H. Wesdorp (1979). *Het grammatica-onderwijs in Nederland. Een research-overzicht betreffende de effecten van grammatica-onderwijs en een verslag van een onderzoek naar de praktijk van dit onderwijs in Nederland*. Den Haag: Staatsuitgeverij.
- Zwrick, R. (1987). Assessing the dimensionality of NAEP reading data. *Journal of Educational Measurement*, 24, 293-308.