



UvA-DARE (Digital Academic Repository)

Rapport onderzoek: De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Wissink, I.B.; Moonen, X.M.H.; Zand Scholten, A.; Stams, G.J.J.M.; Bindels, A.; Lekkerkerker, L.; van der Wal, M.

Publication date

2015

Document Version

Submitted manuscript

[Link to publication](#)

Citation for published version (APA):

Wissink, I. B., Moonen, X. M. H., Zand Scholten, A., Stams, G. J. J. M., Bindels, A., Lekkerkerker, L., & van der Wal, M. (2015). *Rapport onderzoek: De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs*. Universiteit van Amsterdam.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Pedagogiek en Onderwijskunde

Forensische Orthopedagogiek

Nieuwe Achtergracht 127
1018 WS Amsterdam
Postbus 15776
1001 NG Amsterdam

T 020 525 1412

Rapport onderzoek

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Datum:

4 februari 2015

I.B. Wissink (UvA)

X.M.H. Moonen (UvA)

A. Zand Scholten (UvA)

G.J.J.M. Stams (UvA)

A. Bindels (UvA)

L. Lekkerkerker (NJI)

M. van der Wal (GGD)

Met dank aan:

M.R. Bongaards (UvA)

Samenvatting

Achtergrond. Het is van belang om een mogelijke Licht Verstandelijke Beperking (LVB) vroegtijdig te signaleren bij kinderen, zodat tijdig de juist hulp geboden kan worden en ernstige gedrags- en emotionele problemen voorkomen kunnen worden. De school kan een belangrijke rol spelen in de vroegtijdige signalering, maar er is nog geen theoretisch onderbouwd en gevalideerd instrumentarium beschikbaar om de scholen hierbij te helpen. In de DSM-5 wordt bij de definiëring van een verstandelijke beperking meer nadruk gelegd op de term ‘adaptief functioneren’. Een verstandelijke beperking verwijst naar gebrekkige algemene mentale vermogens die invloed hebben op het adaptief functioneren binnen drie domeinen (conceptueel, sociaal en praktisch). Deze drie domeinen bepalen hoe goed een individu omgaat met dagelijkse taken. Onderhavig onderzoek is gericht op de validering van de nieuw ontwikkelde Schaal Adaptief Functioneren (SAF; totaalschaal 15 items; drie subschalen: conceptueel, sociaal en praktisch), een screeningsinstrument dat ingezet kan worden om een risico op LVB mee te signaleren. Indien de score op de SAF laag is, kan vervolgdagnostiek ingezet worden om uit te wijzen of er bij het kind daadwerkelijk sprake is van een LVB.

Methode. De dataverzameling heeft plaatsgevonden in twee fases: de ontwikkelfase en de toepassingfase. Tijdens de ontwikkelfase (fase 1) hebben leerkrachten van basisscholen in Amsterdam en Almere via een online vragenlijst de SAF, SDQ, en Vineland Adaptive Behavior Scales (VABS) ingevuld voor leerlingen in hun klas. Tijdens deze fase zijn SAF-gegevens ingevuld over 1047 leerlingen (gemiddelde leeftijd = 8.9 jaar). De SAF is tevens na twee maanden nogmaals ingevuld over 756 leerlingen. Tijdens de toepassingsfase (fase 2) hebben leerkrachten via het Hart & Ziel leerlingvolgsysteem SDQ en cito gegevens ingevuld. De SAF werd door de leerkrachten ingevuld wanneer de leerling laag scoorde op de ‘pro-sociaal gedrag’ subschaal van de SDQ en op tenminste één van de cito-onderdelen en de leerling geen relatief hoge score (A of ii) had op de cito-onderdelen. Tijdens deze fase zijn SDQ en cito-gegevens ingevuld over 685 leerlingen en SAF-gegevens over 317 leerlingen. Tot slot zijn van deze groep 45 leerlingen verder onderzocht met de zogenaamde ‘Gouden standaard’, oftewel, LVB-experts hebben bij deze kinderen beoordeeld of er sprake was van een LVB op basis van de resultaten op een uitgebreide intelligentietest (VALT) en na een gesprek met de ouders en/of leerkracht van de leerling.

Resultaten. De resultaten van huidig onderzoek bevestigen de betrouwbaarheid (interne consistentie, test-hertest betrouwbaarheid, intra-beoordelaars betrouwbaarheid), validiteit (inhoudsvaliditeit, meet invariantie voor kinderen met een autochtoon Nederlandse en allochtoon etnische achtergrond, constructvaliditeit), en er zijn aanwijzingen voor het discriminerend vermogen van de SAF, als screeningsinstrument waarmee gesignaleerd kan worden of er sprake is van een risico op LVB. Daarnaast zijn de sekse- en etnische verschillen in de SAF verkend. Vervolgonderzoek is in ontwikkeling en zal zich richten op het verzamelen van gegevens over meer leerlingen, zodat huidige gegevens gerepliceerd kunnen worden.

Introductie

Een groot aantal kinderen en jongeren in Nederland ondervindt ernstige problemen doordat zij een Licht Verstandelijke Beperking (LVB) hebben. Kinderen en jongeren met een LVB begrijpen anderen vaak verkeerd en schatten hun reacties niet goed in. Als gevolg hiervan kunnen ze in de problemen komen, hierdoor weer gefrustreerd raken en in nieuwe problemen raken. Het is dan ook niet verwonderlijk dat LVB-jongeren vaker dan gemiddeld met justitie in aanraking komen. Naar schatting behoort ongeveer een kwart van de jongeren in de justitiële inrichtingen tot deze groep (Wegwijzer Jeugd en Veiligheid, 2014). LVB-jongeren met een justitieel verleden hebben extra veel moeite om aansluiting te vinden in de maatschappij. Adequate begeleiding is daarom noodzakelijk.

Bij een groot deel van de kinderen en jongeren met een LVB gaat de beperking samen met leerproblemen, gedragsproblemen, emotionele problemen en gezinsproblemen. Ook is bekend dat kinderen en jongeren met een LVB meer risico lopen om seksueel misbruikt te worden of seksueel grensoverschrijdend gedrag te vertonen (Wissink, Moonen, Van Vugt, Stams, & Vergeer, 2012; Wissink, Van Vugt, Moonen, Stams, & Hendriks, 2014). Veel problemen van kinderen en jongeren met een licht verstandelijke beperking worden veroorzaakt door het feit dat hun beperking te laat onderkend wordt. Een licht verstandelijke beperking wordt veelal pas vastgesteld als de hulpverlening op een dood spoor zit, soms pas na het basisonderwijs (Duits & Bartels, 2011, p. 122). Goede en vroegtijdige signalering zou dan ook veel problemen (en inefficiënte zorg) kunnen voorkomen (GGD Amsterdam, 2014).

Het College van B&W van de gemeente Amsterdam investeert in een betere vroegtijdige signalering van problematiek om tijdig de juiste hulp te kunnen bieden en hiermee verergering van de problematiek te voorkomen. Daarnaast wilde het College van B&W van de gemeente Amsterdam wat meer zicht krijgen op de mate waarin psychosociale problemen en een licht verstandelijke beperking (LVB) zich voordoen onder de Amsterdamse jeugd. Het College was hierbij ook geïnteresseerd in de samenhang met de etnische achtergrond van deze kinderen. De indruk bestond dat met name de problematiek bij kinderen uit niet-westerse migrantengezinnen onvoldoende erkend werd (Onderzoeksrapport GGD Amsterdam-UvA, 2008).

Het College vroeg dan ook om onderzoek naar een methode waarmee kinderen op jonge leeftijd kunnen worden gescreend op LVB-kenmerken om zo vroegtijdige onderkenning van LVB mogelijk te maken en om uiteindelijk uitspraken te kunnen doen over de mate van voorkomen van LVB bij de Amsterdamse jeugd. Bij de ontwikkeling van een dergelijk instrument is voortgebouwd op het BSA-instrument (Beoordeling Sociaal Aanpassingsvermogen; Schouten, Lekkerkerker, & Konijn, 2009), een instrument dat tussen 2006 en 2008 is ontwikkeld voor toepassing binnen de jeugdzorg door het Nederlands Jeugdinstituut en de Vereniging Orthopedagogische Behandelcentra LVB (destijds in opdracht van het ministerie van Volksgezondheid, Welzijn en Sport). Met de BSA kon het sociaal aanpassingsvermogen van jongeren binnen de klinische populatie beoordeeld worden, met als doel een betere onderkenning van problemen die te maken hebben met een LVB binnen de klinische populatie. Er was echter nog geen (kort) screeningsinstrument beschikbaar voor toepassing binnen de niet-klinische populatie en om die reden is de Schaal Adaptief Functioneren (SAF) ontwikkeld (voor een niet-klinische, schoolgaande populatie).

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Op basis van geschatte landelijke percentages met betrekking tot de verdeling van intelligentie zouden er in Amsterdam ongeveer 31.000 kinderen zijn die qua gemeten intelligentie functioneren op het niveau van een licht verstandelijke beperking (Onderzoeksrapport GGD Amsterdam-UvA, 2008). Hoeveel kinderen daadwerkelijk een LVB hebben is echter onbekend. Zowel de landelijke als Amsterdamse cijfers zijn dus schattingen, aangezien goede meetinstrumenten ontbreken. LVB wordt in Nederland gedefinieerd op basis van een gemeten IQ tussen de 50-55 en 70 of een IQ tussen de 71-85, in combinatie met een gebrekkig sociaal aanpassingsvermogen en bijkomende medische en/of psychiatrische problemen en het vermoeden van een langdurige behoefte aan ondersteuning (Moonen & Verstegen, 2006). Er zijn ook verschillende relevante omgevingsfactoren die met LVB in verband worden gebracht. Zo komt LVB vaker voor in lagere sociale milieus. In de wetenschappelijke literatuur wordt ook vaak over LVB-problematiek gesproken vanuit het cumulatief risicomodel: hoe meer opeenstapeling van negatieve factoren (zoals cognitieve en sociale zwakte in combinatie met armoede, stress, middelengebruik, opvoedingsfalen), des te groter de kans op maatschappelijke problemen, zoals storend gedrag en criminaliteit.

Schattingen laten zien dat er in Nederland ongeveer 55.000 (maar mogelijke tienduizenden meer) mensen zijn met een LVB (IQ 50-69) en waarschijnlijk enkele honderdduizenden zwakbegaafde mensen (IQ 70-85) met bijkomende problemen (Ras, Woittiez, Van Kempen, & Sadiraj, 2010). Van deze mensen komen jaarlijks zo'n 11.000 kinderen en jongeren in ernstige problemen op basis van meervoudig complexe problematiek, waarbij gedacht moet worden aan onder andere ernstige gedragsproblemen, verslaving, schooluitval en misbruik (Ras e.a., 2010). Daarnaast is er nog een groep kinderen en jongeren met LVB die enkelvoudige problematiek kent. Deze groep komt ook vaker in aanraking met jeugdzorg vergeleken met gemiddeld begaafde kinderen en jongeren.

Gezien het gebrek aan exacte cijfers omtrent de mate van het voorkomen van LVB bij kinderen in Nederland, maar ook gezien de ontwikkelingsrisico's bij deze groep kinderen, is het van groot belang een goede methode te ontwikkelen waarmee op jonge leeftijd (bij kinderen op basisscholen) een risico op LVB kan worden vastgesteld, zodat vervolgdagnostiek kan plaatsvinden. Hiermee zou niet alleen tijdig de juiste zorg geboden kunnen worden, maar zou ook meer zicht kunnen worden verkregen in de prevalentie van LVB. Het onderzoeksproject waarover hier gerapporteerd wordt, richt zich op de ontwikkeling en validatie van een dergelijke LVB-screeningsmethode voor toepassing bij kinderen in het basisonderwijs.

In het kader van het onderzoeksproject zijn in twee fases gegevens verzameld. De eerste fase, de ‘ontwikkelfase’ is gebruikt voor het valideren, onderzoeken en testen van mogelijke items en uiteindelijke bepaling van de items voor opname in de Schaal Adaptief Functioneren (SAF). Voor de dataverzameling in de tweede fase, de ‘toepassingsfase’ is een getrapte screeningsmethode gebruikt die bestond uit de volgende 3 stappen (waaronder de ontwikkelde SAF): in de eerste plaats zijn kinderen die zowel lage CITO-scores als lage door de leerkracht gerapporteerde SDQ (‘Strengths and Difficulties Questionnaire’) scores hadden geselecteerd (stap 1), zodat leerkrachten over die leerlingen het nieuw ontwikkelde instrument, de SAF (Schaal Adaptief Functioneren), konden invullen (stap 2). Uiteindelijk is bij een selectieve groep leerlingen over wie de SAF was ingevuld daarnaast nog een uitgebreide ‘intelligente’ testbatterij afgenomen, de Verkorte Adaptief Leervermogen Test (VALT; hier bestempeld als ‘Gouden Standaard’) (stap 3), om de SAF verder te kunnen valideren aan de hand van dit criterium. In huidig rapport worden de resultaten beschreven van de analyses ter validatie van het nieuw ontwikkelde instrument, de SAF, dat gebruikt wordt in de tweede stap van de screeningsmethode. De SDQ en CITO testen (eerste stap) zijn beproefde methoden. Voor meer informatie over de VALT verwijzen wij naar de methodesectie in dit onderzoeksrapport en naar de publicatie van het Landelijk Kenniscentrum LVB van Moonen en Wissink (2015). In huidig onderzoeksrapport zal tevens een eerste verkenning worden uitgevoerd van mogelijke verschillen tussen kinderen met een Nederlands etnische achtergrond en kinderen met een niet-Nederlands etnische achtergrond in het adaptief functioneren, gemeten met de SAF (als indicatie van mogelijk bestaan van een LVB). Het gehele onderzoek is uitgevoerd met goedkeuring van de Ethische Commissie van de afdeling Pedagogiek en Onderwijskunde van de Universiteit van Amsterdam.

Achtergrond SAF

De DSM-5 verschilt in bepaalde opzichten van de DSM-IV als het gaat om de definitie van een verstandelijke beperking. Zo is er een verandering in terminologie: waar eerst werd gesproken over een ‘mental retardation’ gebruikt men nu de term ‘intellectual disability’ (ID, oftewel verstandelijke beperking). Men spreekt van een verstandelijke beperking bij een IQ van ongeveer 2 standaarddeviaties (of meer) onder het populatiegemiddelde. Dit komt neer op een IQ van 70 of lager. Deze IQ test scores worden overigens niet meer genoemd bij de diagnostische criteria, maar wel in de bijbehorende tekst van de DSM-5. Een andere verandering is dat er in de DSM-5 meer nadruk wordt gelegd op de term ‘adaptief functioneren’, gedefinieerd als intellectueel functioneren in het dagelijks leven, of het vermogen om onafhankelijk te functioneren en met een gevoel van sociale verantwoordelijkheid. Voorbeelden van adaptief functioneren die worden genoemd zijn begrip van tijd en van de waarde van geld, perceptie en interpretatie van sociale ‘cues’ en persoonlijke zorg (eten, aankleden e.d.). De ontwikkelde SAF beoogt dit concept ‘adaptief functioneren’ te meten als kenmerk van een verstandelijke beperking. Op die manier kan het instrument ingezet worden als screeningsinstrument voor LVB (een lage score op de SAF, oftewel een gebrekkig adaptief functioneren, verwijst dan naar een risico op LVB).

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Een verstandelijke beperking verwijst naar gebrekkige algemene mentale vermogens die invloed hebben op het adaptief functioneren binnen drie domeinen (APA, 2013). Deze domeinen bepalen hoe goed een individu omgaat met dagelijkse taken. Het *conceptuele* domein omvat taalvaardigheden, lezen, schrijven, rekenen, redeneren, kennis en geheugen. Het *sociale* domein verwijst naar capaciteiten als empathie, sociaal oordeel, interpersoonlijke communicatievaardigheden, het vermogen om vriendschappen te maken en vrienden te behouden. Het *praktische* domein, tenslotte, draait om zelfmanagement in gebieden zoals persoonlijk zorg, werk (verantwoordelijkheden), geld management, recreatie, en het organiseren van taken voor school en werk (APA, 2013).

De SAF is ontwikkeld door een team van experts op het gebied van kinderen met een LVB, alsmede instrumentontwikkeling en onderzoek op dit terrein (zie methode). Hierbij is, zoals gezegd, voortgebouwd op de BSA (Beoordeling Sociaal Aanpassingsvermogen, Schouten e.a., 2009) die was ontwikkeld voor gebruik binnen de jeugdzorg (via hulpverlenerrapportages). In het onderhavige onderzoek wordt beoogd de betrouwbaarheid en validiteit van de SAF (voor gebruik op basisscholen via leerkrachtrapportages) te onderzoeken. Hiertoe zal gekeken worden naar de inhoudsvaliditeit, construct (convergente) validiteit, interne consistentie, test- hertest betrouwbaarheid en het discriminerend vermogen van de SAF. Speciale aandacht zal uitgaan naar de resultaten voor kinderen met een Nederlands etnische achtergrond en kinderen met een niet-Nederlands etnische achtergrond gezien de wens van de gemeente Amsterdam om uiteindelijk iets te kunnen zeggen over mogelijke verschillen tussen deze groepen in het adaptief functioneren (als indicatie van het mogelijk bestaan van een LVB).

Methodede

Procedure

Zoals gezegd heeft de dataverzameling voor huidig onderzoek plaatsgevonden in twee fases: de eerste fase (ontwikkelfase) was gericht op de dataverzameling ten behoeve van de vaststelling van de test-hertest betrouwbaarheid, en de convergente validiteit van de SAF, fase 2 was meer gericht op het toepassen van de SAF, als onderdeel van het Hart & Ziel leerlingvolgsysteem (toepassingsfase) (zie ook onderstaand ‘Overzicht dataverzameling’).

Ontwikkelfase (fase I)

Tijdens de ontwikkelfase hebben leerkrachten via een online vragenlijst in Qualtrics de SAF (werknaam ‘ISA’), SDQ, en Vineland Adaptive Behavior Scales (VABS) ingevuld voor leerlingen in hun klas (zie onder). Tijdens deze fase zijn SAF-gegevens ingevuld over 1047 leerlingen door 113 leerkrachten.

Toepassingsfase (fase II)

Tijdens de toepassingsfase hebben leerkrachten via het Hart & Ziel leerlingvolgsysteem de SDQ en cito gegevens ingevuld. Hart & Ziel is een methode om de sociaal-emotionele ontwikkeling van kinderen te volgen. Met Hart & Ziel worden sociaal-emotionele problemen op tijd gesignaleerd en kunnen kinderen passend begeleid en geholpen worden. De SAF werd door de leerkrachten ingevuld wanneer de leerling een somscore op de ‘pro-sociaal gedrag’ subschaal van de SDQ had van lager dan 10, en op tenminste één van de cito-onderdelen een relatief lage score had (D of E of iv of v) en de leerling geen relatief hoge score (A of ii) had op de cito-onderdelen. Tijdens deze fase zijn SDQ en cito-gegevens ingevuld over 685 leerlingen door 63 leerkrachten en de SAF-gegevens over 317 leerlingen. Uiteindelijk zijn 45 leerlingen uitgebreid getest met behulp van de VALT en VTIQ (WISC-verkort) (meting 5).

Overzicht dataverzameling.

	Meting 1	Meting 2	Meting 3	Meting 4	Meting 5
	Via online vragenlijst		Via Hart en Ziel systeem		Via scholen
I	SAF	SAF			
Ontwikkelfase	(N = 1047)	(N = 756)			
	SDQ				
	(N = 1037)				
	VABS				
	(N = 1039)				
	IQ (LVB)				
	(N = 170)				
II			SDQ	SAF	VTIQ
Toepassingsfase			(N = 685)	(N = 317)	(N = 45)
			CITO		VALT
			(N = 685)		(N = 45)

Schaal Adaptief Functioneren (SAF) (constructvaliditeit, betrouwbaarheid)

De Schaal Adaptief Functioneren (SAF) is tijdens de ontwikkelfase ingevuld over 1047 leerlingen (113 leerkrachten) (393 autochtone leerlingen; 650 allochtone leerlingen¹). De gemiddelde leeftijd van de leerlingen was 8.87 jaar ($SD = .811$) en het ging om 553 jongens en 490 meisjes (53% jongens; 47% meisjes; de sekseverdeling was gelijk over de etnische groepen, $\chi^2(1) = .000$, $p = .998$). Deze meting is na 2 maanden herhaald (meting 2) ten behoeve van de analyse van de test-hertest betrouwbaarheid. Tijdens deze hertest is de SAF ingevuld over 756 leerlingen (uitval 27.8%; zie tevens resultaten). De leerkrachten ontvingen IRIS-cheques voor hun medewerking en ouders van de kinderen konden deelname weigeren nadat zij een informatiebrief hadden ontvangen. Slechts een klein percentage van de ouders heeft deelname geweigerd (< 2%).

In de ontwikkelfase van de SAF zijn in eerste instantie 41 stellingen geformuleerd door een team van experts. Deze 41 items waren voor een deel afkomstig uit de BSA, waarvoor destijds de expertise van professionals in het werkveld van de zorg aan jeugdigen met een LVB, een internationale literatuurverkenning, en een aantal bestaande nationale en internationale instrumenten (SRZ-P, Vineland-Z, ABAS-II; Adaptive Behavior Assessment System-Second Edition, ABI; Adaptive Behavior Inventory, ABES, SIB-R, ICAP; Inventory for Client and Agency Planning, ABS, NABC) als basis dienden (Schouten e.a., 2009). Voor de ontwikkeling van de SAF is zowel de internationale als nationale literatuur wederom bestudeerd (e.g., McKenzie & Megson, 2012) en zijn nieuw ontwikkelde instrumenten bekeken (ABES-R2, ABS-S:2, SEV; Sociaal-Emotionele Vragenlijst, VISEON; Volginstrument sociaal-emotionele ontwikkeling en het volgsysteem ZIEN!). Tenslotte zijn zowel experts als leerkrachten geraadpleegd.

Uiteindelijk is het aantal items van 41 teruggebracht naar 15 SAF-items die het totale adaptief functioneren beogen te meten en deze items zijn gebruikt in de toepassingsfase (met 5 items per subschaal). Items zijn verwijderd naar aanleiding van opmerkingen van de leerkrachten, wanneer er sprake was van een hoog percentage missende waarden ($\geq 10\%$; indicatief voor lastig te beantwoorden vraag) of in geval van lage correlaties of factorladingen in de resultaten van een exploratieve factoranalyse tijdens de ontwikkelfase (< .35; indicatief voor inhoudelijk zwakke vraag).

Een voorbeeld van een stelling van de uiteindelijke conceptuele subschaal is 'Heeft moeite om een verhaallijn te volgen' (gespiegeld item), van de sociale subschaal 'Lijkt zich in een ander te kunnen verplaatsen' en van de praktische subschaal 'Controleert taak (kijkt eigen werk na)'. Leerkrachten konden op een drie-puntsschaal aangeven in hoeverre de stelling van toepassing was op de desbetreffende leerling (0 = niet waar, 1 = een beetje waar, 2 = zeker waar). Een hogere score op de SAF betekent een hoger adaptief functioneren (en lager risico op LVB).

¹ In navolging van de definitie van het Centraal Bureau voor de Statistiek is een leerling als 'allochtoon' beschouwd als tenminste 1 van de ouders buiten Nederland was geboren.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Constructvaliditeit (convergente validiteit)

Voor de analyse van de construct (convergente) validiteit van de SAF zijn de SDQ en VABS van de ontwikkelfase en cito-gegevens van de toepassingsfase gebruikt (zie hieronder).

SDQ

Met de 'Strengths and Difficulties Questionnaire' (SDQ; Goodman, 1997) zijn de gedragsproblemen gemeten van 1037 leerlingen (ontwikkelfase). De SDQ is een korte vragenlijst, die bedoeld is voor kinderen van drie tot zestien jaar, en bestaat uit 25 items die 5 subschalen meten: emotionele problemen, gedragsproblemen, hyperactiviteit/aandachtsproblemen, problemen met leeftijdgenoten, pro-sociaal gedrag. De totale SDQ bleek voldoende betrouwbaar (Cronbach's $\alpha = .71$; emotionele problemen $\alpha = .77$; gedragsproblemen $\alpha = .67$; hyperactiviteit/aandachtsproblemen $\alpha = .75$; problemen met leeftijdgenoten $\alpha = .51$; pro-sociaal gedrag $\alpha = .84$).

Vineland Adaptive Behavior Scales (VABS)

De subschaal 'Interpersoonlijke relaties' van het socialisatiedomein van de Vineland Adaptive Behavior Scales (VABS; 23 items; Sparrow, Cicchetti, & Balla, 2006) is ingevuld voor 1039 leerlingen (ontwikkelfase). De VABS is een uitgebreide vragenlijst om het adaptieve gedrag van kinderen te meten. De 23 items van de subschaal 'Interpersoonlijke relaties' bleken intern consistent: Cronbach's $\alpha = .94$.

Cito

Voor de cito-gegevens is gebruik gemaakt van gegevens die zijn verzameld via het Hart & Ziel leerlingvolgsysteem (toepassingsfase). Via het Hart & Ziel systeem zijn over een groep van 685 leerlingen (vergelijkbare) cito-gegevens ingevuld door leerkrachten van 21 basisscholen.² Het ging om de subtests rekenen/wiskunde en begrijpend lezen. Er is voor deze twee subtests gekozen aangezien deze op alle scholen werden afgenomen en gezien de correlaties met intelligentietesten (Driessen, 2005). Over een subgroep van 317 leerlingen waren naast de cito-scores (via de getrapte screeningsmethode) de SAF-gegevens ingevuld. Deze gegevens zijn gebruikt voor de analyse van de convergente validiteit van de SAF (op basis van de cito-gegevens). De verschillende typen cito-gegevens zijn zo gehercodeerd dat een hogere score op de variabele 'cito' een hoger cognitief vermogen aangaf.³

Discriminerend vermogen

Voor de analyse van het discriminerend vermogen van de SAF zijn in de eerste plaats officiële IQ test-scores die bekend waren bij de basisschoolleerkrachten ($N = 170$ leerlingen met een IQ-score op LVB niveau; gegevens ontwikkelfase) gebruikt. Daarnaast zijn op basis van de gegevens van de leerkrachten uit de toepassingsfase zowel leerlingen die een risico op LVB vertoonden als leerlingen die een dergelijk risico niet vertoonden geselecteerd voor verder onderzoek. Bij deze groep van 45 leerlingen is extra informatie verzameld, waaronder de testgegevens op de WISC-verkort en de VALT, hier bestempeld als de 'Gouden Standaard' (zie hieronder).

IQ-test

² Bij 1 school heeft de IB-er ook gegevens ingevuld.

³ De verschillende typen cito-gegevens (met numerieke waarden i t/m v of letters A t/m E) zijn allen gehercodeerd naar een schaal van 1 – 5, waarbij een hogere score een beter cognitief vermogen aangaf.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Bij 170 leerlingen (van de 1047 in de oorspronkelijke dataset van de ontwikkelfase) was de leerkracht op de hoogte van een gestandaardiseerde IQ-testuitslag van lager dan 85, oftewel, van een LVB op basis van een IQ-test. Deze LVB-leerlingen zijn vergeleken met de overige leerlingen.

WISC-verkort (VTIQ)

Daarnaast is bij 45 leerlingen uit het databestand van de toepassingsfase de WISC-III-NL-verkort (ook wel: VTIQ) afgenomen. De WISC-III-NL is een bewerking van de Engelstalige WISC-III (1992), de tweede herziening van de 'Wechsler Intelligence Scale for Children' (1949). De verkorte versie bestond uit de onderdelen 'Onvolledige tekeningen', 'Informatie', 'Substitutie', 'Blokpatronen' en 'Woordkennis' (Dumont & Faro, 1993). De scores op de WISC-verkort zijn eveneens gebruikt als criterium (zie analyseplan).

VALT ('Gouden Standaard')

Tenslotte is bij 45 leerlingen uit het databestand van de toepassingsfase een uitgebreide testbatterij afgenomen, die door experts is ontwikkeld om een indruk te krijgen van zowel de zwakke als sterke punten van iedere leerling met betrekking tot het brede spectrum van het begrip 'intelligentie' (Moonen & Ponsioen, 2013; Moonen & Wissink, 2015). De testbatterij bestond uit de VALT (Verkorte Adaptief Leervermogen Test). De VALT omvatte de subtests 'Woordkennis', 'Informatie', 'Substitutie' en 'Cijfers voorwaarts zeggen' uit de WISC III 6-17 jaar, de subtest 'Woordbetekenis' uit de RAKIT-2; 4-11 jaar (Resing, Bleichrodt, Drenth, & Zaal, 2012), en de subtest 'Woordreeksen nazeggen' en 'Ontbrekende figuren' uit de CAS 5-18 (Ponsioen, 2010). De benodigde tijd voor de afname van de WISC-verkort en de VALT was ongeveer 1 ½ uur. De leerlingen ontvingen een Bart Smit waardebond voor hun deelname (en ouders konden uiteraard deelname van hun kind weigeren). Op basis van de resultaten van deze uitgebreide testbatterij en na een gesprek met de ouder(s) en leerkracht is door twee LVB-experts (dr. X. Moonen en dr. A. Ponsioen) een oordeel gevormd of er bij de leerlingen sprake was van een LVB. Deze procedure wordt gezien als de 'Gouden Standaard'.

Analyseplan

De *interne consistentie* is een aspect van de betrouwbaarheid en heeft betrekking op de mate waarin onderdelen van een meetinstrument met elkaar samenhangen. Men kan de mate waarin delen van een meetinstrument met elkaar samenhangen berekenen met de coëfficiënt van interne consistentie. Een test die intern consistent blijkt te zijn, zal in het algemeen ook betrouwbaar zijn (Van Peet, Van den Wittenboer, & Hox, 2001). De interne consistentie van de SAF is beoordeeld aan de hand van de Cronbach's alfa waarden, zowel in de totale groep als in de 2 etnische groepen afzonderlijk. Algemene richtlijnen schrijven voor dat Cronbach's alfa's van tussen de .70 en .90 als acceptabel tot goed worden beschouwd. In de onderzoekspraktijk wordt ook wel een alfa van .60 als ondergrens gehanteerd (Bijleveld, 2009).

Een andere vorm van betrouwbaarheid is de stabiliteit, of *test-hertest betrouwbaarheid*. Informatie hierover wordt verkregen door te kijken naar de correlatie r tussen de scores op twee paralleltests. De test-hertest betrouwbaarheid van de SAF is gebaseerd op de correlaties tussen de eerste en de tweede afname van de SAF (meting 1 en 2; ontwikkelfase). Daarnaast is de intra-beoordelaarsbetrouwbaarheid per item berekend (Cohen's Kappa).

Behalve dat een instrument betrouwbaar moet zijn moet het ook een hoge validiteit bezitten. Validiteit kan hierbij worden omschreven als de mate waarin een meetinstrument aan zijn doel

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

beantwoordt, oftewel, of het datgene meet wat het verondersteld wordt te meten (Van Peet e.a., 2001). Er zijn verschillende manieren om de validiteit vast te stellen of om argumenten ter verdediging van de validiteit aan te voeren.

Inhoudsvaliditeit heeft te maken met de mate waarin een meetinstrument het bedoelde terrein bestrijkt. Het begrip ‘inhoudsvaliditeit’ impliceert dat de onderzoeker in staat is om precies te omschrijven wat het terrein is waarop het instrument betrekking heeft. Om de inhoudsvaliditeit van de SAF te onderzoeken is via een confirmatieve factoranalyse (CFA) getoetst of de veronderstelde factorstructuur met de 3 SAF-subschalen een goede fit vertoonde met de gegevens.

Vervolgens is de inhoudsvaliditeit verder onderzocht door met een multigroep CFA te toetsen of er sprake was van *meetinvariantie* over de 2 etnische groepen. Meetinvariantie kan bekeken worden op verschillende niveaus. De volgende niveaus zijn van belang voor een vergelijking van groepsgemiddelden: configurele invariantie, metrische invariantie en ‘scalar’ invariantie. Configurele invariantie verwijst naar een gelijke basis factorstructuur. Metrische invariantie houdt in dat de factorladingen vergelijkbaar zijn over verschillende groepen en van ‘scalar’ invariantie, tenslotte, is sprake wanneer de intercepten ook vergelijkbaar zijn voor de verschillende groepen (Vandenberg & Lance, 2000). Er is getoetst of er sprake is van deze 3 niveaus van meetinvariantie over de twee etnische groepen, zodat een vergelijking van gemiddelden op de SAF in de etnische groepen mogelijk is.

Een ander aspect van de validiteit is de *begripsvaliditeit*, oftewel, is het meetinstrument in kwestie een adequate representant van het begrip zoals het is bedoeld (Van Peet e.a., 2001) en heeft te maken met de mate waarin de resultaten op een bepaald meetinstrument voor een begrip overeenkomen met de resultaten op een ander meetinstrument. Men kan uitspraken doen over de *begripsvaliditeit* door het patroon van correlaties met diverse meetinstrumenten te bekijken. Als indicatie van de *begripsvaliditeit* is de *convergente validiteit* van de SAF onderzocht aan de hand van de Pearson correlaties met de SDQ-subschalen en VABS (ontwikkelfase) en cito-scores (toepassingsfase).

Criteriumvaliditeit, tenslotte, heeft te maken met de vraag in hoeverre een meetinstrument in staat is een bepaald gedrag te voorspellen. Dat bepaalde gedrag wordt het criterium genoemd. De correlatie tussen het meetinstrument en het criterium wordt dan de validiteitscoëfficiënt genoemd. Hoe sterker de samenhang, des te meer valide de test is voor het criterium. Om de criterium validiteit te onderzoeken is een ROC-curve analyse uitgevoerd met als eerste ‘criterium’ de IQ-test scores bekend bij de scholen (ontwikkelfase), als tweede ‘criterium’ het oordeel op basis van de WISC-verkort scores (VTIQ; 45 leerlingen toepassingsfase), en als laatste de oordelen op basis van de ‘Gouden Standaard’ (VALT) als criterium, (45 leerlingen toepassingsfase). Op deze manier is een indruk verkregen van het *discriminerende vermogen* van de SAF.

Ter afsluiting zijn de SAF gegevens uit het databestand van de ontwikkelfase gebruikt voor een eerste verkenning van mogelijke verschillen tussen leerlingen met een Nederlands etnische achtergrond en leerlingen met een niet-Nederlands etnische achtergrond in het adaptief functioneren (als indicatie van een LVB).

Resultaten

Interne consistentie

Voor de interne consistentie zijn Cronbach's alfa's berekend voor de subschalen van de SAF en de totale SAF schaal. De alfa's zijn weergegeven in Tabel 1 voor zowel de totale groep leerlingen als voor de autochtone en allochtone groep afzonderlijk (gezien het uiteindelijke doel van de Gemeente Amsterdam). Met een alfatest is getoetst of de betrouwbaarheden in de twee etnische groepen significant van elkaar verschilden.

Tabel 1. Interne consistentie.

	Totale Groep	Autochtone leerlingen	Allochtone leerlingen	<i>p</i> -waarde alfatest
Conceptueel	.889	.885	.891	.591
Sociaal	.851	.858	.846	.431
Praktisch	.833	.838	.830	.641
Totale SAF	.735	.747	.731	.554

De resultaten toonden aan dat alle SAF-subschalen in zowel de totale groep als in beide etnische groepen een goede interne consistentie vertoonden en dat de totale SAF een acceptabele interne consistentie vertoonde (.735 totale groep). Daarnaast toonden de resultaten van de alfatests dat deze betrouwbaarheden niet significant verschilden wanneer leerkrachten rapporteerden over leerlingen met een autochtoon Nederlands etnische achtergrond danwel over leerlingen met een niet-Nederlandse etnische achtergrond.

Test-hertest betrouwbaarheid

De test-hertest betrouwbaarheid is op twee manieren onderzocht. In de eerste plaats door Pearson's correlatie coëfficiënten te berekenen tussen de SAF-scores op meting 1 en de SAF-scores op meting 2 (over dezelfde leerlingen). Niet onbelangrijk is om hierbij te vermelden dat wij eerst getoetst hebben of er verschillen bestonden tussen de leerlingen over wie we wel meting 2 gegevens hadden en de leerlingen over wie we geen gegevens hadden ontvangen tijdens de herhaalmeting. De resultaten lieten zien dat er geen significante verschillen bestonden, dit gold zowel voor de conceptuele subschaal, $t(1043) = 1.036$, $p = .300$, de sociale subschaal, $t(559) = -.141$, $p = .888$, de praktische subschaal, $t(1043) = .664$, $p = .507$, als voor de totale SAF-schaal, $t(1043) = .646$, $p = .519$. Er waren dus geen significante verschillen tussen de leerlingen die zijn 'uitgevallen' op het tweede meetmoment en de leerlingen over wie we wel herhaalgegevens hadden ($n = 756$) en waar onderstaande resultaten op zijn gebaseerd. Met andere woorden, dit is geen selecte groep en daardoor zijn de resultaten beter generaliseerbaar. Nadat we dit hadden bekeken, zijn de Pearson's correlatiecoëfficiënten berekend. Alle Pearson's correlatiecoëfficiënten waren significant ($p < .001$) en wezen op een sterk verband tussen de SAF-scores op meting 1 en meting 2 (gekoppeld op leerlingniveau). De test-hertest betrouwbaarheid was het sterkst voor de totale SAF ($r = .865$, $p < .001$), gevolgd door de conceptuele subschaal ($r = .811$, $p < .001$), de sociale subschaal ($r = .786$, $p < .001$) en de praktische subschaal ($r = .780$, $p < .001$).

In de tweede plaats is voor elk SAF-item afzonderlijk Cohen's Kappa (κ) uitgerekend voor overeenstemming tussen de inschatting op meting 1 en de inschatting op meting 2. Met uitzondering van 1 item (op de praktische subschaal met 'redelijke' overeenstemming;

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Cohen's Kappa (κ) = .385, $p < .001$) toonden alle items een gemiddelde mate van overeenstemming (Cohen's Kappa (κ) > .40), tien van de vijftien items vertoonden zelfs een Cohen's Kappa (κ) van > .50 (Altman, 1999; Landis & Koch, 1977). De gemiddelde Cohen's Kappa (κ) was .531. Dit is de proportie overeenstemming bovenop de overeenstemming gebaseerd op kans. Daar komt nog bij dat alle Kappa's (κ) een p-waarde van $p < .000$ hadden, wat duidt op kappa coëfficiënten (κ) die statistisch significant afwijken van nul en dus duiden op overeenstemming. Tezamen met de test-herstest resultaten op basis van Pearson's correlatiecoëfficiënten bevestigen deze gegevens de test-herstest betrouwbaarheid van de SAF.

Inhoudsvaliditeit (meet invariantie)

Vervolgens is met een confirmatieve factoranalyse (CFA) getoetst of de veronderstelde factorstructuur met de drie SAF-subschalen een goede fit vertoonde met de gegevens. Verschillende fitmaten worden gebruikt om een indruk te krijgen van de juistheid van een model. Voor modellen met 75 tot 200 'cases' wordt de Chi^2 -waarde gezien als een redelijke fitmaat (Kenny, 2014). Bij modellen met meer respondenten is de Chi^2 -waarde echter al snel significant. Om die reden worden vaak andere fitmaten gerapporteerd, zoals de CFI en de RMSEA. De CFI varieert tussen de 0 en 1, waarbij waarden dicht bij 1 een betere fit aantonen. Voor de CFI wordt doorgaans een waarde van tussen de .90 en .95 gezien als middelmatige fit (> .95 als goede fit; Hu & Bentler, 1999).⁴ Een RMSEA waarde van .05 wordt gezien als indicatief voor een goede fit (.01 = excellente fit en .08 = middelmatige fit; MacCallum, Browne, & Sugawara, 1996). In het ideale geval ligt de lage grens van het 90% betrouwbaarheidsinterval van de RMSEA dichtbij 0 (niet slechter dan .05) en de hoge grens lager dan .08 (Kenny, 2014). Afgaande op deze gegevens vertoonde de basisstructuur van het model een redelijke fit met de data ($\text{Chi}^2(174) = 920.771$, $p < .001$; CFI = .886; RMSEA = .064), maar met toevoeging van 3 correlaties tussen errortermen van items met gelijke woorden of vergelijkbare inhoud een goede fit met de data ($\text{Chi}^2(168) = 594.578$, $p < .001$; CFI = .935; RMSEA = .049) (zie Figuur 1). Deze resultaten duiden op configurele meet invariantie (niveau 1).

Vervolgens is getoetst of de factorladingen gelijk gesteld konden worden over de etnische groepen zonder significante verslechtering in model fit, als indicatie van metrische invariantie (niveau 2). Cheung en Rensvold (1999) stellen dat het beter is om naar een verschil in andere fitmaten dan de Chi^2 te kijken wanneer (geneste) modellen met elkaar vergeleken worden. Veranderingen in de CFI van .01 of minder worden dan gezien als niet significant (en als bewijs voor meetinvariantie of meet equivalentie). In Tabel 2 kunnen de fitmaten van het model met gelijkgestelde factorladingen worden afgelezen. Zoals te zien was het verschil in CFI .003, dit is minder dan .01 en dus kan dit gezien worden als indicatie voor metrische invariantie over de etnische groepen.

Voor het laatste niveau (niveau 3; scalar invariantie) zijn de intercepten gelijk gesteld over de etnische groepen en wederom gekeken naar de fitmaten en het verschil in CFI. De resultaten toonden een goede fit en een verschil in CFI dat niet groter was dan .01. Hiermee

⁴ Hierbij moet echter rekening worden gehouden met het feit dat als de RMSEA van het 'nullmodel' lager is dan .158, een te lage waarde wordt verkregen voor de CFI (homepage D.A. Kenny, 2014). De RMSEA is dan informatiever.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

wordt de hypothese dat er sprake is van scalar equivalentie niet verworpen (oftewel, de hypothese van scalar invariantie wordt aangenomen).

Tabel 2. Drie niveaus meet invariantie.

	Chi ² (df)	CFI	RMSEA	ΔCFI (genest)
1) Configurele invariantie	594.578(168)***	.935	.049	
2) Metrische invariantie	627.586 (183)***	.932	.048	-.003
3) Scalar invariantie	707.968(198)***	.922	.050	-.01

Opm. *** = $p < .001$.

Begripsvaliditeit (convergente validiteit)

Voor de convergente validiteit zijn de correlaties met verschillende andere instrumenten (waarmee men een samenhang zou verwachten) onderzocht. Zo is de samenhang tussen de SAF en de subschalen van de SDQ, tussen de SAF en de VABS, en tussen de SAF en de cito-onderdelen (rekenen/wiskunde en begrijpend lezen) berekend. Aangezien de cito-scores worden gezien als variabelen met een ordinaal meetniveau zijn voor deze samenhangen Spearman's rangcorrelatie coëfficiënten berekend; in de overige gevallen zijn Pearson's correlatie coëfficiënten berekend. In Tabel 3 worden alle correlaties getoond.

Tabel 3. Convergente validiteit.

	SAF concept	SAF soc	SAF prakt	SAF totaal
SDQ Emot. probl.	-.333***	-.321***	-.363***	-.414***
SDQ Gedragsprobl.	-.204***	-.412***	-.482***	-.442***
SDQ Hyper/aand. probl.	-.226***	-.338***	-.532***	-.440***
SDQ Peer probl.	-.042	-.129***	.037	-.056 ^{ms}
SDQ Pro-soc. Gedrag	.359***	.760***	.487***	.652***
VABS	.489***	.733***	.439***	.677***
CITO RW	.653***	.187**	.342**	.472***
CITO BL	.541***	.246***	.375***	.466***

Opm. CITO RW = Citoscore rekenen wiskunde; CITO BL = Citoscore begrijpend lezen. *** = $p < .001$, ** = $p < .01$, ^{ms} = $p < .10$.

Alle correlaties waren in de verwachte richting. Echter, voor de convergente validiteit is een aantal correlaties met name interessant en deze willen we hier extra belichten. Er zijn geen duidelijke grenzen bekend om te kunnen spreken van convergente validiteit. Echter, het is zo dat de correlaties met meetinstrumenten waar men convergentie mee zou verwachten zo hoog mogelijk moeten zijn (en hoger dan correlaties met instrumenten waarmee men dit niet zou verwachten) (Web center for Social Research Methods, 2014).

Volgens de theorie achter de SAF (zie pagina 5/6) zou de conceptuele subschaal moeten samenhangen met de cito-scores voor rekenen-wiskunde en begrijpend lezen. Deze subschaal vertoonde inderdaad de sterkste en vrij hoge correlaties met de cito-scores van de leerlingen (resp. -.653*** en -.541***). Voor de sociale subschaal zou je de sterkste correlaties verwachten met de subschaal pro-sociaal gedrag van de SDQ en met de VABS (Vineland screener, subschaal interpersoonlijke relaties), en deze waren inderdaad het sterkst en hoog (resp. .760*** en .733***). Opvallend is dat de correlatie met de subschaal peer problemen van de SDQ vrij laag was, maar het is bekend dat deze subschaal vaak onbetrouwbaar is (zie bijvoorbeeld Koskelainen, Sourander, & Vauras, 2001), en dit kan de lage correlatie verklaren. Voor de praktische subschaal zou je, gezien de schoolse context, de sterkste samenhang verwachten met de de subschaal hyperactiviteit/ aandachtsproblemen, en deze correlatie bleek inderdaad de hoogste te zijn (-.532***). Tot slot kan nog gekeken

worden naar de correlaties tussen de totale SAF-score en de overige gemeten concepten. Uit deze correlaties blijkt dat in de totale SAF-score het sociale aspect het sterkst vertegenwoordigd lijkt, gezien de bevinding dat de totale SAF-score het sterkst positief samenhangt met de meetinstrumenten die sociale aspecten beogen te meten (SDQ pro sociaal gedrag en VABS). Desalniettemin vertoonde de totale SAF-score, volgens verwachting, ook vrij sterke significant positieve correlaties met beide CITO-scores, en significant negatieve correlaties met alle probleemgedrag metingen. Over het geheel genomen bevestigen de gevonden correlaties de convergente validiteit van de SAF.

Criteriumvaliditeit (discriminerend vermogen)

Het is van belang dat de SAF goed discrimineert tussen kinderen met een risico op LVB en kinderen zonder een dergelijk risico. Hiervoor zijn ROC analyses uitgevoerd die een indicatie van de AUC-(Area Under the Curve) waarde geven. Een AUC-waarde van rond de .70 of hoger duidt op een veel betere classificatie dan een classificatie gebaseerd op kans (Dolan & Doyle, 2000; Shapiro, 1999).

Allereerst is er een vergelijking gemaakt in de volledige dataset (met informatie verkregen van de leerkrachten) tussen de kinderen die op basis van een bekende intelligentietest tot de LVB-groep behoorden ($IQ < 85$) en de kinderen die daar niet toe behoorden. Een ROC analyse toonde een significante AUC-waarde van .682, $p < .001$. De gemiddelden in beide groepen op de totale SAF bleken eveneens significant van elkaar te verschillen, $t(267) = 8.540$, $p < .001$. De kinderen bij wie een intelligentietest een LVB had geïndiceerd scoorden significant lager op de SAF ($M = 14.39$; $SD = 5.44$) dan de kinderen bij wie dit niet het geval was ($M = 18.40$; $SD = 6.35$).

Daarnaast zijn de gegevens van de 45 leerlingen die verder getest zijn gebruikt. Bij deze leerlingen is in de eerste plaats de verkorte WISC test (VTIQ) afgenomen. In de tweede plaats is de nieuw ontwikkelde 'Gouden Standaard' (VALT) afgenomen bij deze 45 kinderen en op basis van die resultaten en na een gesprek met de leerkracht en ouder(s) hebben twee experts op het gebied van LVB (dr. X.Moonen & A. Ponsioen) beoordeeld of er bij een kind sprake was van een risico op LVB of niet (VALT).⁵ Deze 3 typen gegevens zijn eveneens gebruikt als 'criterium' in de analyses om de AUC (Area Under the Curve) gegevens te verkrijgen.

De AUC-waarde met als criterium het oordeel op basis van de WISC-verkort ($IQ < 85 =$ risico op LVB) was .673, $p = .053$. De AUC-waarde met als criterium het oordeel op basis van de 'Gouden Standaard' (VALT) testuitslagen van X. Moonen was .766, $p = .009$ en van A. Ponsioen .729, $p = .029$. Deze gegevens geven aanwijzingen dat de SAF onderscheid kan maken tussen kinderen met een risico op LVB en kinderen zonder een risico op LVB, met name wanneer men de 'Gouden Standaard' (VALT) als criterium voor het bestaan van een LVB gebruikt.

Wanneer de WISC-verkort uitslag werd aangehouden voor het onderscheid wel LVB – niet LVB bleek er een marginaal significant verschil te bestaan tussen de LVB en niet-LVB kinderen in de totale SAF-score, $t(43) = 1.709$, $p = .095$. De gemiddelden toonden aan dat de LVB-kinderen ($M = 12.71$; $SD = 3.53$) iets lager scoorden op de SAF dan de niet-LVB

⁵ Blind voor de SAF-scores van de leerlingen.

kinderen ($M = 14.18$; $SD = 2.26$). Wanneer de expert oordelen op basis van de Gouden Standaard (VALT) werden aangehouden, bleken de verschillen in SAF-scores volledig significant, respectievelijk $t(43) = 2.551$, $p = .014$ en $t(43) = 2.709$, $p = .010$. De LVB-kinderen scoorden significant lager op de SAF (respectievelijk $M = 11.82$; $SD = 3.68$ en $M = 11.60$; $SD = 2.80$) dan de niet-LVB kinderen (respectievelijk $M = 14.21$; $SD = 2.32$ en $M = 14.20$; $SD = 2.64$). Deze gegevens geven eerste aanwijzingen voor het onderscheidend vermogen van de SAF (als screeningsinstrument), maar meer gegevens zijn nodig om huidige resultaten te repliceren.

Op basis van de resultaten van de ROC analyses is een voorlopig afkappunt vastgesteld. In Tabel 4 worden de sensitiviteit (terecht positieven) en de specificiteit (terecht negatieven) voor de verschillende afkappunten weergegeven. Op basis van deze gegevens blijkt dat bij een afkapwaarde van 13.50 de sensitiviteit (% terecht positieven) 91% is en de specificiteit (% terecht negatieven) 59% is. Gezien het vrij hoge percentage van de onterecht negatieven bij deze grenswaarde (41%) is ervoor gekozen om twee voorlopige afkappunten te formuleren: bij een totale SAF-score van lager dan 12 is er voldoende reden om vervolgonderzoek uit te voeren en bij een totale SAF-score van 14 of hoger is dit niet direct geïndiceerd (tenzij er andere aanwijzingen zijn). Het tussenliggende gebied (een totale SAF-score van 12 of 13) wordt aangemerkt als het twijfelgebied (zie voor de meest recente gegevens omtrent de afkappunten: <http://www.uva.nl/profiel/i.b.wissink>).

Tabel 4. Sensitiviteit en specificiteit (in %).

Afkappunt (\leq)	Sensitiviteit (terecht positieven)	Specificiteit (terecht negatieven)	Onterecht positieven	Onterecht negatieven
6.00	0	100	100	0
7.50	9.1	100	90.9	0
8.50	18.2	100	81.8	0
9.50	27.3	100	72.7	0
10.50	27.3	94.1	72.7	5.9
11.50	45.5	88.2	54.5	11.8
12.50	63.6	73.5	36.4	26.5
13.50	90.9	58.8	9.1	41.2
14.50	90.9	47.1	9.1	52.9
15.50	90.9	32.4	9.1	67.6
16.50	90.9	14.7	9.1	85.3
17.50	90.9	8.8	9.1	91.2
18.50	90.9	2.9	9.1	97.1
20.00	90.9	0	9.1	100
22.00	100	0	0	100

Opn.: Resultaten op basis van VALT ('Gouden standaard') met eindoordeel dr. X. Moonen (hoogste AUC-waarde) als criterium.

Etnische verschillen in adaptief functioneren

De etnische verschillen in het adaptieve functioneren zijn verkend middels variantie-analyses. In de eerste plaats is een multivariate variantie-analyse (MANOVA) uitgevoerd met de drie SAF-subschalen als afhankelijke variabelen en zowel geslacht (gezien mogelijke sekseverschillen) als etniciteit als factor. De multivariate resultaten toonden aan dat er zowel sprake was van een significant sekseverschil, $F(3, 1032) = 39.820$, $p < .001$, power = 1.000, als van een significant etnisch verschil, $F(3, 1032) = 3.336$, $p = .019$, power = .760, in de SAF-subschalen. De interactie tussen geslacht en etniciteit bleek niet significant, $F(3, 1032) = .910$, $p = .436$, power = .251.

Tabel 5. Sekseverschillen in adaptief functioneren (subschalen en totale schaal).

	Meisjes	Jongens
Conceptueel AF	6.25 (.14)	6.11 (.13)
Sociaal AF***	7.16 (.13)	5.53 (.12)
Praktisch AF***	6.33 (.12)	5.10 (.12)
Totaal SAF***	19.75 (.32)	16.74 (.30)

Opm. AF = Adaptief Functioneren. *** = $p < .001$.

Gezien de significante hoofdeffecten is vervolgens naar de univariate resultaten gekeken en deze lieten zien dat er sprake was van een sekseverschil in zowel sociaal adaptief functioneren, $F(1, 1034) = 86.978$, $p < .001$, power = 1.000, als praktisch adaptief functioneren, $F(1, 1034) = 52.534$, $p < .001$, power = 1.000. Op beide aspecten van het adaptief functioneren bleken meisjes hoger te functioneren. De univariate resultaten duiden daarnaast op een significant etnisch verschil in het conceptueel adaptief functioneren, $F(1, 1034) = 8.671$, $p = .003$, power = .837. Op deze subschaal behaalden de leerlingen met een autochtoon Nederlands etnische achtergrond een hoger gemiddelde dan de leerlingen met een niet-Nederlandse etnische achtergrond. In Tabel 5 en 6 zijn de gemiddelden en standaarddeviaties weergegeven.

Tot slot zijn de verschillen in de totale SAF-scores vergeleken voor de verschillende groepen met behulp van een univariate variantie-analyse (ANOVA) met de totale SAF-score als afhankelijke variabele en zowel geslacht als etniciteit als factor. De resultaten duiden wederom op zowel een sekseverschil, $F(1, 1034) = 46.280$, $p < .001$, power = 1.000, als een etnisch verschil, $F(1, 1034) = 5.439$, $p = .020$, power = .579, in het totale adaptieve functioneren. Het interactie-effect bleek niet significant. De gemiddelde scores in de afzonderlijke groepen toonden aan dat meisjes een hogere score hadden op het totale adaptieve functioneren dan jongens. Daarnaast bleken kinderen met een autochtoon Nederlands etnische achtergrond een hogere score te hebben op de totale SAF schaal dan kinderen met een niet-Nederlands etnische achtergrond (zie Tabel 6).

Tabel 6. Etnische verschillen in adaptief functioneren (subschalen en totale schaal).

	Autochtoon	Allochtoon
Conceptueel AF**	6.46 (.15)	5.90 (.12)
Sociaal AF	6.48 (.14)	6.21 (.11)
Praktisch AF	5.77 (.14)	5.66 (.11)
Totaal SAF*	18.72 (.35)	17.76 (.27)

Opm. AF = Adaptief Functioneren. ** = $p < .01$; * = $p < .05$.

Conclusie

Over het geheel genomen ondersteunen de resultaten van huidig onderzoek de betrouwbaarheid en validiteit van de Schaal Adaptief Functioneren (SAF). Zo bleek de interne consistentie en test-hertest betrouwbaarheid in orde van zowel de subschalen als de totale SAF. Daarnaast duiden de resultaten erop dat de SAF even betrouwbaar en valide is in de autochtoon Nederlands etnische groep als in de niet-autochtoon Nederlandse groep, getuige de resultaten van de betrouwbaarheids- en confirmatieve factoranalyses. Deze positieve resultaten zijn voor een ‘zuinig’ instrument (de SAF bestaat uit ‘slechts’ 15 items) mooie uitkomsten, vergelijk bijvoorbeeld resultaten van factoranalyses en betrouwbaarheidsanalyses van de alom bekende en veel gebruikte ‘Strengths and Difficulties Questionnaire’ (SDQ). Ook in huidig onderzoek bleek bijvoorbeeld de interne consistentie van de subschaal ‘problemen met leeftijdgenoten’ van de SDQ onder de maat (.51). Dit gegeven kan echter wel de marginaal significante (in plaats van volledig significante) correlatie tussen deze subschaal en de SAF verklaren. Met alle overige subschalen vertoonde de SAF tenslotte significante correlaties in de verwachte richting, hiermee de convergente validiteit van de SAF ondersteunend.

De resultaten met betrekking tot de convergente validiteit toonden verder aan dat in de totale SAF-score het sociale aspect het sterkst vertegenwoordigd lijkt (sterke samenhangen met de SDQ-subschaal pro sociaal gedrag en de subschaal ‘Interpersoonlijke relaties’ van het socialisatie domein van de ‘Vineland Adaptive Behavior Scales’; VABS). Desalniettemin vertoonde de totale SAF-score, volgens verwachting, ook vrij sterke significant positieve correlaties met beide CITO-scores. Over het geheel genomen bevestigden de gevonden correlaties de convergente validiteit van de SAF. Het is niet verwonderlijk dat de SAF een sterk vertegenwoordigd sociaal aspect bevat, aangezien de SAF is ontwikkeld met het idee dat kinderen met een LVB niet alleen tekortkomingen in het cognitief vermogen vertonen, maar ook in het sociaal (adaptief) vermogen. Het komt ook overeen met de meest recente definitie van het adaptief functioneren in de DSM-5, als kenmerk van een verstandelijke beperking, waarbij het gevoel van sociale verantwoordelijkheid specifiek wordt benadrukt (APA, 2013). Overige instrumenten, zoals de standaard intelligentietests (e.g., WISC-III), houden vaak (te) weinig rekening met de sociale aspecten en zijn (te) veel gericht op het cognitieve vermogen.

Naast een vergelijkbaarheid van de interne consistentie in de twee etnische groepen bleken er ook aanwijzingen te zijn voor meet invariante over de twee etnische groepen leerlingen (autochtoon Nederlandse etnische achtergrond en niet autochtoon Nederlandse etnische achtergrond), wat vertrouwen geeft in een vergelijking van gemiddelde SAF-scores van deze groepen leerlingen. Deze vergelijking liet zien dat kinderen met een autochtoon Nederlandse etnische achtergrond een hoger adaptief functioneren lieten zien (volgens de leerkrachten) dan kinderen met een niet-autochtoon Nederlandse etnische achtergrond. Deze resultaten zijn in lijn met de verwachtingen van de Gemeente Amsterdam. De verschillen leken volgens de SAF-scores overigens met name in het conceptuele adaptief functioneren te bestaan, oftewel, in de taalvaardigheden, lezen, schrijven, rekenen, redeneren, kennis en geheugen. Naar verwachting indiceert dit een groter risico op het bestaan van een LVB. Hierbij dient wel vermeld te worden dat de verschillen tussen meisjes en jongens in het adaptief vermogen groter leken dan de verschillen tussen de etnische groepen, zodat voorgaande conclusie dan ook met voorzichtigheid geïnterpreteerd moet worden. Vervolgonderzoek is nodig om de suggestie dat kinderen met een niet-autochtoon

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Nederlandse etnisch achtergrond een groter risico hebben op een LVB verder te onderbouwen.

In overeenstemming met de convergente validiteit resultaten bleek de SAF eveneens het sterkst samen te hangen met de oordelen op basis van de ‘Gouden Standaard’, de VALT. Het is een mooie bevinding dat de SAF het meeste overeenstemde met de VALT, aangezien bij de VALT eveneens breder naar het kind wordt gekeken, en niet slechts naar het cognitieve vermogen van een kind, wat een te beperkte aanpak is. Bij zowel de SAF als de VALT wordt het brede spectrum van het concept ‘adaptief functioneren’ verkend en daarmee wordt recht gedaan aan alle aspecten van het theoretisch model dat ten grondslag ligt aan het begrip LVB.

Wanneer wordt uitgegaan van de oordelen op basis van de VALT geven de resultaten van de ROC-analyses aan dat het discriminerend vermogen van de SAF redelijk goed is (AUC > .70). Een AUC-waarde van boven de 70 is redelijk goed, in ogenschouw nemende dat de SAF dient te worden toegepast als screeningsinstrument. Dat wil zeggen, de SAF dient niet gebruikt te worden als diagnostisch instrument voor LVB, maar als instrument om een risico op LVB te signaleren. Vervolgdiagnostiek (bijvoorbeeld met behulp van de ‘VALT’) is vervolgens nodig om vast te stellen of er bij een leerling daadwerkelijk sprake is van een functioneren op niveau van kinderen met een LVB.

Ondanks deze positieve eerste bevindingen zijn er nog enkele vragen waar toekomstig onderzoek zich op zou kunnen richten. Een eerste vraag is of er wellicht subtypen LVB bestaan. Resultaten van additionele analyses geven aan dat de kinderen met een LVB (volgens de ‘VALT’) met name verschilden van de kinderen zonder een LVB in het conceptuele domein van de SAF en (maar in minder mate) op het sociale domein van de SAF. Vervolgonderzoek zou gericht kunnen worden op het onderzoeken van de waarde van de SAF-subschalen voor de praktijk. Het zou interessant zijn om normscores voor de subschalen te ontwikkelen en te kijken naar de samenhangen tussen de scores op de afzonderlijke SAF-subschalen en de uitslagen op de afzonderlijke onderdelen van de ‘VALT’. Daarnaast zou vervolgonderzoek gericht op sekseverschillen in het adaptief functioneren gewenst zijn. Op basis hiervan kan informatie gegeven worden over eventuele sekse-specifieke normscores en sekse-specifieke afkappunten. Deze vormen van meer gedifferentieerde informatie zouden uiteindelijk gebruikt kunnen worden voor een specifieke, op het individu afgestemde vervolgzorg, oftewel, vervolgzorg ‘op maat’.

Tenslotte zijn er meer gegevens nodig voor de definitieve bepaling van de afkappunten van de SAF. De huidig beschikbare gegevens zijn gebruikt voor de bepaling van voorlopige afkappunten (Totale SAF-score < 12: vervolgonderzoek geïndiceerd; 12 – 13: twijfelgebied; 14 en hoger: in principe geen vervolgonderzoek geïndiceerd tenzij er andere aanwijzingen zijn), maar meer gegevens waarbij een koppeling wordt gemaakt tussen SAF-scores en het wel of niet bestaan van een LVB zijn nodig om hier gefundeerde uitspraken over te kunnen doen en om de definitieve afkappunten vast te stellen. Deze gegevens zullen naar verwachting komend schooljaar verzameld worden (zie voor de meest recente gegevens omtrent de afkappunten: <http://www.uva.nl/profiel/i.b.wissink>). Daarnaast zijn eveneens gegevens verzameld over het adaptief functioneren bij de ouders van de 45 leerlingen die uitgebreid getest zijn (Wissink, Moonen, Zand Scholten, & Bindels, 2014). Aangezien we voor de screening in het huidige onderzoek slechts zijn uitgegaan van leerkrachtrapportages

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

is het interessant om in de toekomst te onderzoeken in hoeverre deze samenhangen met rapportages van de ouders. Hierbij zal het informatief zijn om ook te kijken naar de verbanden tussen de ouderrapportages over het adaptief functioneren en de VALT uitslagen.

Op dit moment kan echter geconcludeerd worden dat de resultaten van huidig onderzoek de betrouwbaarheid en validiteit van de nieuw ontwikkelde Schaal Adaptief Functioneren (SAF) bevestigen. Met de SAF wordt het ‘adaptief functioneren’ gemeten, gedefinieerd als ‘intellectueel functioneren in het dagelijks leven, oftewel, het vermogen om onafhankelijk te kunnen functioneren met een gevoel van sociale verantwoordelijkheid’. Voorbeelden van adaptief functioneren die worden genoemd zijn begrip van tijd en van de waarde van geld, perceptie en interpretatie van sociale signalen en persoonlijke zorg (zoals eten en aankleden). Een kind dat functioneert op het niveau van een LVB zal gebrekkige algemene mentale vermogens hebben die het adaptief functioneren beïnvloeden (APA, 2013). De resultaten duiden erop dat de SAF een zuinig screeningsinstrument is waarmee op betrouwbare en valide manier het adaptief functioneren gemeten kan worden bij zowel leerlingen met een autochtoon Nederlandse etnische achtergrond als bij leerlingen met een niet-autochtoon Nederlandse achtergrond, en waarbij een lage SAF-score een verhoogd risico op het bestaan van een LVB indiceert. In geval van een lage SAF-score is vervolgonderzoek de aangewezen volgende stap.

De SAF kan ingevuld worden door de leerkracht (of een ander persoon die de leerling zeer goed kent en zicht heeft op hoe deze leerling functioneert) voor leerlingen vanaf groep 5 en kan op die manier gebruikt worden om het mogelijk bestaan van een LVB vroegtijdig te signaleren. In een in 2014 verschenen rapport van het Expertisecentrum voor jeugd, samenleving en ontwikkeling (Greeven, 2014) werd al geconcludeerd dat dit van cruciaal belang is ter voorkoming van een cumulatie van problemen op latere leeftijd. Het rapport duidde echter ook op een tekort aan aandacht voor LVB bij scholen. Alhoewel de verantwoordelijkheid voor vroegtijdige signalering van LVB niet alleen bij de scholen zou liggen (maar ook bij sociale wijkteams en/of jeugdteams), zou vroegtijdige signalering wel in de eerste plaats vragen om alertheid en scherpte bij de scholen. De hier ontwikkelde SAF, in combinatie met de beschreven ‘VALT’ (zie ook: Moonen & Wissink, 2015), kan scholen en leerkrachten handvatten bieden om een dergelijke vroegtijdige signalering van een functioneren op het niveau van een LVB bij kinderen mogelijk te maken, zodat deze kinderen tijdig de juiste hulp kunnen krijgen en op die manier een verdere positieve ontwikkeling ondersteund wordt.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Referenties

- Altman, D. G. (1999). *Practical statistics for medical research*. New York, NY: Chapman & Hall/CRC Press.
- American Psychiatric Association (2013). *Factsheet Intellectual Disability*. Verkregen via: www.psychiatry.org/...5/DSM-5-Intellectual-Disability-Fact-Sheet.pdf
- Cheung, G.W., & Rensvold, R.B. (1999). Cross-cultural comparisons using non-invariant measurement items. *Applied Behavioral Science Review*, 6, 93-110.
- Dolan, M., & Doyle, M. (2000). Violence risk prediction. Clinical and actuarial measures and the role of the Psychopathy Checklist. *British Journal of Psychiatry*, 177, 303-311.
- Driessen, G. (2005). *De cross-sectionele en longitudinale ontwikkeling van de intelligentietest scores in PRIMA*. Nijmegen: ITS.
- Duits, N., & Bartels, J.A.C. (2011). *Jeugdpsychiatrie en recht: Wetgeving, zorgveld en praktijk*. Assen: Van Gorcum.
- Dumont, R., & Faro, C. (1993). A WISC-III short form for learning-disabled students. *Psychology in the Schools*, 30, 12-19.
- GGD Amsterdam-UvA (2008). *Zware bagage: Psychische problemen en verstandelijke beperkingen bij allochtone kinderen in Amsterdam*. Amsterdam: GGD-UvA.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Greeven, H. (2014). *Rapportage onderzoek naar jongeren met een licht verstandelijke beperking*. Gouda: JSO Expertisecentrum voor jeugd, samenleving en ontwikkeling.
- Kenny, D.A. (2014). *Measuring model fit*. Verkregen via: <http://davidakenny.net/cm/fit.htm>.
- Koskelainen, M., Sourander, A., & Vauras, M. (2001). Self-reported Strengths and Difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry*, 10, 180-18.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130-149.
- McKenzie, K., & Megson, P. (2012). Screening for Intellectual Disability in children: A review of the literature. *Journal of Applied Research in Intellectual Disabilities*, 25, 80-87.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

- Moonen, X., & Ponsioen, A. (2013). *Onderzoeksvoorstel GGD*. Amsterdam: Universiteit van Amsterdam.
- Moonen, X., & Verstegen, D. (2006). LVG-jeugd met ernstige gedragsproblematiek in de verbinding van praktijk en wetgeving. *Onderzoek & Praktijk. Tijdschrift voor de LVG-zorg, 1*, 23-28.
- Moonen, X. & Wissink, I. (2015). *Signalering van kinderen die functioneren op het niveau van een LVB in het basisonderwijs*. Utrecht: Landelijk Kenniscentrum LVB.
- Ponsioen, A. (2010). *Een kind met mogelijkheden: Een andere kijk op LVG-kinderen*. Houten: Bohn Stafleu van Loghum.
- Ras, M., Woittiez, I., Van Kempen, H., & Sadiraj, K. (2010). *Steeds meer verstandelijk gehandicapten? Ontwikkelingen in vraag en gebruik van zorg voor verstandelijk gehandicapten 1998-2008*. Den Haag: Sociaal en Cultureel Planbureau (SCP).
- Resing, W.C.M., Bleichrodt, N., Drenth, P.J.D.D., & Zaal, J.N. (2012). *Revisie Amsterdamse Kinder Intelligentie Test-2 (RAKIT-2). Gebruikershandleiding*. Amsterdam: Pearson.
- Schouten, R., Lekkerkerker, L., & Konijn, C. (2009). *Ontwikkeling van en onderzoek naar het instrument Beoordeling Sociaal Aanpassingsvermogen: Eindrapport*. Utrecht: Nederlandse Jeugdinstuut (NJI).
- Shapiro, J.H. (1999). The interpretation of diagnostics tests. *Statistical Methods in Medical Research, 8*, 113-134.
- Web center for Social Research Methods (2014). *Convergent & discriminant validity*. Verkregen via: <http://www.socialresearchmethods.net/kb/convdisc.php>.
- Wissink, I.B., Moonen, X., Van Vugt, E.S., Stams, G.J., & Vergeer, M. (2012). *Seksueel grensoverschrijdend gedrag en misbruik bij kinderen en jongeren met een (licht) verstandelijke beperking. Deelonderzoek 5B Commissie-Samson*. Amsterdam: Boom.
- Wissink, I.B., Moonen, X., Zand Scholten, A., & Bindels, A. (2014). *Vragenlijst Adaptief Functioneren*. Amsterdam: Universiteit van Amsterdam.
- Wissink, I.B., Van Vugt, E.S., Moonen, X.M.H., Stams, G.J.J.M., & Hendriks, J. (2014). Sexual abuse involving children with an Intellectual Disability (ID): A narrative review. *Research in Developmental Disabilities, 36*, 20-35.
- Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2006). *Vineland-II: Vineland Adaptive Behavior Scales (Second Edition)*. San Antonio: Pearson.
- Vandenberg, R.J., & Lance, C.E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70.

De ontwikkeling en validering van een LVB-screeningsinstrument voor toepassing in het basisonderwijs

Figuur 1. Factormodel Confirmatieve Factor Analyse (CFA) Schaal Adaptief Functioneren (SAF).

