

Wetenschappelijke verantwoording VISEON 2.0 Volginstrument voor sociaal-emotioneel functioneren

Hans Kuhlemeier, Hanneke Knoop, Herman van Boxtel,
Ilse Papenburg en Judith Hollenberg



Wetenschappelijke verantwoording

WISEON 2.0

Volginstrument voor sociaal-emotioneel functioneren

Hans Kuhlemeier
Hanneke Knoop
Herman van Boxtel
Ilse Papenburg
Judith Hollenberg

© Cito B.V. Arnhem (2016)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Inhoud

1	Inleiding	5
2	Uitgangspunten van de testconstructie	7
2.1	Meetpretentie	7
2.2	Doelgroep	10
2.3	Gebruiksdoel en functie	11
2.4	Theoretische inkadering	14
2.5	De keuze van domeinen en aspecten	21
2.5.1	Sociaal-emotioneel functioneren van individuele leerlingen	23
2.5.1.1	Zelfbeeld	25
2.5.1.2	Welbevinden	25
2.5.1.3	Sociaal gedrag	26
2.5.1.4	Werkhouding	26
2.5.2	Leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren	27
2.5.3	Doorlopende leerlijnen	28
3	Constructie en beschrijving van de vragenlijsten	31
3.1	Constructie van de leerkracht- en leerlinglijst	31
3.1.1	Uitgangspunten voor de keuze van uitspraken	31
3.1.2	De proefafnames	32
3.2	Opbouw, afname en scoring van de leerkracht- en leerlinglijst	33
3.2.1	Opbouw	33
3.2.2	Afname	34
3.2.3	Scoring	35
3.3	Beschrijvende inhoudelijke en statistische gegevens	35
4	Het normeringsonderzoek	43
4.1	Opzet en uitvoering van het normeringsonderzoek	43
4.2	Verschillen tussen afnamemodi en jaargroepen	46
4.2.1	Verschillen tussen afnamemodi	46
4.2.2	Verschillen tussen jaargroepen	52
4.2.3	Conclusies	55
4.3	De representativiteit van de normgroepen	55
4.3.1	De representativiteit van de normgroepen op schoolniveau	55
4.3.2	De representativiteit van de normgroepen op leerlingniveau	58
4.4	De normering	61
5	Betrouwbaarheid en meetnauwkeurigheid	67
5.1	Interne consistentie, betrouwbaarheidsintervallen en standaardmeetfout	67
5.2	Interbeoordelaarsbetrouwbaarheid	68
5.3	Test-hertestbetrouwbaarheid	71

6	Validiteit	75
6.1	Interne structuur	76
6.1.1	Uni-dimensionaliteit	76
6.1.2	Factorzuiverheid	77
6.1.3	Correlaties tussen de schalen	85
6.1.4	Invariantie over relevante groepen	88
6.2	Kwaliteit van de itemparameters: DIF-analyses	92
6.2.1	Afnamemodus	92
6.2.2	Jaargroep	94
6.2.3	Geslacht	94
6.3	Externe structuur	95
6.3.1	Relaties met VISEON (eerste generatie)	95
6.3.2	Relaties met de <i>Strength and Difficulties Questionnaire</i> (SDQ)	100
6.3.3	Relaties met de Kanjervragenlijst	102
6.3.4	Relaties met de Schoolvragenlijst (SVL)	106
6.4	Relevante verschillen tussen groepen	108
6.4.1	Geslacht	108
6.4.2	Stad versus platteland	110
6.4.3	Leerlinggewicht	111
7	Samenvatting	113
8	Literatuur	117
	Bijlage 1 Correlaties tussen de schalen van LK basis, LL basis en LL aanvullend	123

1 Inleiding

Deze wetenschappelijke verantwoording heeft betrekking op het digitale VolgInstrument Sociaal-Emotionele ONtwikkeling (verder VISEON). VISEON maakt sinds het schooljaar 2003-2004 deel uit van het Cito Volgstelsel primair en speciaal onderwijs. VISEON 2.0 bouwt in naam voort op zijn voorganger, maar is in vergelijking daarmee inhoudelijk sterk aangepast en verbreed. De eerste versie van VISEON was gebaseerd op een persoonlijkheidsmodel (het Big Five model) met bipolaire dimensies, waarbij overigens slechts vier van de vijf dimensies in schalen konden worden ondergebracht. Veel gebruikers plaatsten kanttekeningen bij een persoonlijkheidsmodel als basis voor een volgstelsel. Het bipolaire karakter van de dimensies riep daarnaast, in relatie tot de gebruikelijke niveau-indeling A-E, vragen op. Ook het feit dat deze dimensies zowel in de vorm van schalen voor de leerkrachtlijst als in de vorm van schalen voor de leerlinglijst parallel werden uitgewerkt (vanaf groep 5) kwam bij sommige gebruikers wat geforceerd over. Dit wil niet zeggen dat VISEON niet goed functioneerde. Ondanks alle genoemde bezwaren sloten de meeste schalen inhoudelijk goed aan bij wat leerkrachten relevante kindkenmerken vinden in relatie tot het functioneren op school en bij wat zij beschouwen als belangrijke 'sociale opbrengsten' van het onderwijs. In VISEON 2.0 is afscheid genomen van veel van de genoemde uitgangspunten, waaronder het persoonlijkheidsmodel, maar is tegelijkertijd getracht de inhoudelijke voordelen van VISEON te behouden en uit te breiden.

VISEON 2.0 biedt de mogelijkheid om zicht te krijgen op het sociaal-emotioneel functioneren van leerlingen en de kwaliteit van de leeromgeving. Het instrument sluit aan bij VISEON voor kleuters. VISEON 2.0 bestaat uit een vragenlijst voor de leerkracht van groep 3 tot en met groep 8 en een zelfbeoordelingslijst voor de leerlingen van groep 5 tot en met groep 8. Daarnaast bevat VISEON 2.0 aanvullende leerlinglijsten die naar behoefte ingezet kunnen worden (zie verder paragraaf 2.3).

VISEON 2.0 is gebaseerd op de gedachte dat leerkrachten en – vanaf groep 5 – ook leerlingen goede informanten zijn als het gaat om het sociaal-emotioneel functioneren. Elke informant heeft zijn eigen kijk op de verschillende aspecten van het sociaal-emotioneel functioneren. De observatielijst en de zelfbeoordelingslijst bestaan uit een aantal uitspraken of stellingen waarvan aangegeven kan worden in welke mate deze van toepassing zijn. Met behulp van de scores op de verschillende schalen van de leerkrachtlijst is het sociaal-emotioneel functioneren in de ogen van een leerkracht op een bepaald afnamemoment te bepalen. En het niveau van de leerling is te vergelijken met dat van een landelijke referentiegroep. Als een leerkracht elk leerjaar een of twee maal de leerkrachtlijst invult, is het mogelijk het sociaal-emotioneel functioneren te volgen in groep 3 tot en met groep 8. Vanaf eind groep 5 kan men leerlingen daarnaast de leerlinglijst laten invullen. De schalen van de leerlinglijst geven inzicht in het beeld dat een leerling van het eigen sociaal-emotioneel functioneren heeft en wat de leerling van de kwaliteit van de leeromgeving vindt. Als een leerling elk leerjaar de leerlinglijst een of twee maal invult, is het mogelijk de kwaliteit van de leeromgeving en het sociaal-emotioneel functioneren zoals de leerling deze zelf beleeft te volgen. Anders dan toetsen van Cito die schoolse vaardigheden in kaart brengen, bestaat er voor alle verschillende jaargroepen één leerkrachtlijst en één leerlinglijst. Elk jaar wordt dezelfde vragenlijst ingevuld.

De administratie- en afnameomgeving is gebaseerd op het Computerprogramma LOVS. Dit betekent dat de leerkrachtlijst en leerlinglijst digitaal worden afgenomen (ingevuld) door gebruik te maken van het Computerprogramma LOVS. Ook de planning, scoring en rapportage gebeuren via dit programma. Het Computerprogramma LOVS kan voor alle digitale toetsen van het Cito Volgstelsel primair en speciaal onderwijs worden gebruikt.

Samen met de inhoud van de leerkrachtlijst, de leerlinglijst, de aanvullende leerlinglijsten, de handleiding (Cito, 2015) die in de meest recente versie via Cito Portal op te vragen is en het Computerprogramma LOVS, bevat deze verantwoording alle informatie die nodig is voor een snelle en efficiënte beoordeling van

de kwaliteit van VISEON 2.0. Het genoemde materiaal maakt een beoordeling van het instrument mogelijk op de volgende aspecten:

- Uitgangspunten van de testconstructie
- De kwaliteit van het testmateriaal
- De kwaliteit van de handleiding
- Normen
- Betrouwbaarheid
- Validiteit

Het laatstgenoemde aspect betreft alleen begripsvaliditeit en geen criteriumvaliditeit. Omdat de onderdelen van het Cito Volgsysteem primair en speciaal onderwijs niet bedoeld zijn voor 'voorspellend gebruik' is criteriumvaliditeit niet van toepassing.

Het voorliggende document heeft met name betrekking op de uitgangspunten van de constructie (hoofdstuk 2), de daadwerkelijke constructie (hoofdstuk 3), de normen (hoofdstuk 4), de betrouwbaarheid en meetnauwkeurigheid (hoofdstuk 5) en de begripsvaliditeit (hoofdstuk 6) van VISEON 2.0. De kwaliteit van het testmateriaal en de handleiding is te bepalen door kennis te nemen van de desbetreffende onderdelen en het Computerprogramma LOVS.

2 Uitgangspunten van de testconstructie

2.1 Meetpretentie

Naast de cognitieve ontwikkeling heeft het onderwijs de opdracht de persoonlijke, sociale en morele ontwikkeling van leerlingen te bevorderen. VISEON 2.0 verhoogt het inzicht in de persoonlijke en sociale ontwikkeling van leerlingen door leerkrachten in staat te stellen het sociaal-emotionele functioneren in kaart te brengen en systematisch te volgen. De informatie die men met VISEON kan verkrijgen kan gelegd worden naast de resultaten van leervorderingstoetsen in het Cito Volgsysteem primair en speciaal onderwijs. Daarbij wordt een onderscheid gemaakt in het sociaal-emotioneel functioneren op individueel niveau enerzijds en de leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren anderzijds.

Het sociaal-emotioneel functioneren van individuele leerlingen

Het sociaal-emotioneel functioneren van individuele leerlingen omvat vier domeinen:

- *Zelfbeeld*: beleving van het eigen sociaal en cognitief functioneren
- *Welbevinden*: gedragsuitingen en ervaringen die een indicatie geven van de emotionele stabiliteit en ervaren veiligheid
- *Sociaal gedrag*: gedrag waarin leerlingen hun sociale vaardigheid (competentie) tot uitdrukking brengen
- *Werkhouding*: taak- en werkhouding in diverse situaties waarin men van de leerling verwacht dat deze een bepaalde mate van concentratie en (volgehouden) aandacht opbrengt.

Het is van belang om scherp onderscheid te maken tussen sociaal-emotioneel *functioneren* en sociaal-emotionele *ontwikkeling*. VISEON 2.0 focust sterk op concreet gedrag en functioneren. Deze worden deels bepaald door de ontwikkelingsprocessen en het ontwikkelingsniveau dat aan de basis ligt van dit gedrag, maar dat geldt zeker niet voor alle aspecten die in het instrument zijn ondergebracht in even sterke mate. Schoolbeleving bijvoorbeeld wordt niet alleen beïnvloed door het individuele ontwikkelingsverloop, maar ook door onder meer de kwaliteit van de leerkracht, de inrichting van de leeromgeving en ingrijpende gebeurtenissen in de thuissituatie. Daarnaast is het, voor zover ontwikkeling een rol speelt, met VISEON 2.0 niet mogelijk om deze ontwikkeling *af te beelden* in termen van bijvoorbeeld een vaardigheidsschaal zoals bij de toetsen voor leervorderingen binnen het Cito Volgsysteem primair en speciaal onderwijs het geval is. Wel kan men de kwaliteit van het sociaal-emotioneel functioneren bepalen door dit functioneren te *vergelijken met een representatieve referentiegroep* van leerlingen uit het reguliere basisonderwijs.

VISEON 2.0 bestaat uit een hoofdinstrument en aanvullende instrumenten. Het hoofdinstrument bevat een vragenlijst voor de leerkracht (verder: leerkrachtlijst) en een zelfbeoordelingslijst voor de leerling (verder: leerlinglijst). Mocht uit het hoofdinstrument blijken dat er behoefte is aan aanvullend onderzoek, dan kan de leerkracht vanaf groep 5 ervoor kiezen om aanvullende leerlinglijsten in te zetten.

De vier domeinen bestaan uit verschillende aspecten die met afzonderlijke schalen in het instrument vertegenwoordigd zijn, zoals sociaal zelfbeeld, motivatie voor leren en schoolbeleving (zie onderstaande tabellen voor het onderscheid tussen domeinen en aspecten). Een deel van de schalen wordt voorgelegd aan de leerkracht en een ander deel aan de leerling. Voor elk aspect is gekeken wie het meest geschikt is als bron van informatie, de leerkracht of de leerling.

In de gebruikershandleiding (Cito, 2015) wordt overigens niet over aspecten gesproken, maar over categorieën. De reden is dat VISEON 2.0 voor de afname en rapportage ervan is ondergebracht in het veel gebruikte Computerprogramma LOVS waarin vaardigheden (deelscores) in de rapportages niet met aspecten maar met categorieën worden aangeduid.

De leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren

De vragen over de leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren, worden alleen door de leerlingen beantwoord en in de vorm van een groepsoverzicht gerapporteerd.

Anders dan bij toetsen van Cito die schoolse vaardigheden in kaart brengen, bestaat er in VISEON 2.0 voor alle verschillende jaargroepen één leerkrachtlijst en één leerlinglijst. In elke jaargroep wordt dezelfde vragenlijst ingevuld. De vragen zijn zo geformuleerd dat zij betekenis krijgen in relatie tot het gedrag dat van leerlingen in een bepaalde jaargroep wordt verwacht. Van een leerling uit groep 8 verwacht men bijvoorbeeld bij het oplossen van een ruzie andere vaardigheden dan van een leerling uit groep 3. Bij het invullen van de lijsten, zowel door leerkracht als leerlingen, vormt dat gedrag het referentiekader. In dit opzicht is de opzet van VISEON 2.0 vergelijkbaar met die van zijn voorganger en die van veel vergelijkbare instrumenten zoals ZIEN! (Broer, Haverhals & De Bruin (2012), de SDQ als onderdeel van het volgsysteem Hart & Ziel (Diepenmaat, Van Eijnsden, Janssens, Loomans & Stone, 2014) en de Kanjervragenlijst (Vliek, Riet & Weide, 2012).

In onderstaand overzicht¹ zijn de schalen weergegeven die beschikbaar zijn in VISEON 2.0. De keuze van de domeinen en aspecten wordt verderop in dit hoofdstuk verantwoord. In de loop van de tijd zal het aanvullende gedeelte zo mogelijk nog uitgebreid worden met andere instrumenten voor leerlingen en leerkrachten.

VISEON 2.0 hoofdinstrument individuele leerlingen			
sociaal-emotioneel functioneren van individuele leerlingen			
	aspecten leerlinglijst	aspecten leerkrachtlijst	
domein zelfbeeld	sociaal zelfbeeld		
	cognitief zelfbeeld		
domein welbevinden	schoolbeleving		
domein sociaal gedrag	gepest worden		pro-sociaal gedrag
			pestgedrag
		storend gedrag	
		teruggetrokken gedrag	
domein werkhouding	motivatie voor leren	taakgericht gedrag	
		motivatie voor leren	
		zelfstandigheid	
		hyperactief/impulsief	

VISEON 2.0 hoofdinstrument leeromgeving	
leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren	
aspecten leerlinglijst: rapportage in groepsoverzicht	
domein leeromgeving	sociale leeromgeving
	onveilige leeromgeving
	onordelijke leeromgeving

¹ Dit overzicht is bepaald door de afnamevolgorde en is gelijk aan dat in de handleiding voor de leerkracht. Verderop zullen we een iets andere volgorde hanteren bij de presentatie van onderzoeksresultaten.

VISEON 2.0 aanvullende instrumenten*	
	<i>aspecten leerlinglijst</i>
domein zelfbeeld	zelfbeeld lezen
	zelfbeeld rekenen
domein welbevinden	spreekangst ²
	toetsangst
domein sociaal gedrag	pro-sociaal gedrag
	pestgedrag
	storend gedrag
domein werkhouding	taakgericht gedrag
	motivatie voor lezen
	motivatie voor rekenen

* Bij de aanvullende leerlinglijsten kan gekozen worden of één, enkele of alle domeinen worden afgenomen.

VISEON 2.0 geeft een beeld van het sociaal-emotioneel functioneren van individuele leerlingen en van groepen leerlingen als geheel. De scores van de leerlingen op de lijsten worden vergeleken met de scores van een landelijk representatieve groep leerlingen. Door het periodiek (laten) invullen van de lijsten worden veranderingen in het sociaal-emotioneel functioneren in kaart gebracht.

De unieke combinatie van leerkrachtlijst en leerlinglijst maakt het mogelijk tot een afgewogen en genuanceerd oordeel te komen, in gesprek te gaan en actie te ondernemen. Onderzoek laat zien dat de leerkracht bepaald gedrag dikwijls anders beoordeelt dan een leerling dat zelf doet (Vliek, Riet & Weide, 2012). Leerlingen en leerkrachten observeren, interpreteren en beoordelen gedrag vanuit een verschillend referentiekader. Daardoor zien ze deels andere dingen en hechten ze daaraan meer of minder waarde. Dat hangt onder meer af van de aard van het te observeren gedrag. Zo is externaliserend probleemgedrag, zoals agressief en storend gedrag in de klas, goed waarneembaar voor de leerkracht en heeft deze daar ook last van bij de ordehandhaving, terwijl internaliserend probleemgedrag, zoals piekeren en depressie, veel minder goed waarneembaar is voor de leerkracht, maar via introspectie juist weer toegankelijk voor de leerling zelf. Het is dus waardevol om ten aanzien van het gedrag van leerlingen verschillende informatiebronnen te benutten.

Zowel de leerkracht- als de leerlinglijst zijn bedoeld als signaleringsinstrumenten. Ze zijn bedoeld voor het handelen in de klas, dus niet om risicogedrag op te sporen. De aanvullende instrumenten kunnen gebruikt worden als uit het hoofdinstrument blijkt dat er behoefte is aan aanvullend onderzoek. De aanvullende instrumenten kunnen ook ingezet worden bij specifieke problemen (individueel dan wel op groepsniveau) of wanneer in het schoolplan is vastgelegd dat bepaalde onderwerpen extra aandacht behoeven (denk bijvoorbeeld aan pesten, agressie). De aanvullende instrumenten helpen om te bepalen waarop de school bij het maken van een plan van aanpak kan inzetten: gericht op de individuele leerling, op groepjes leerlingen of de klas, of op de school als geheel.

Voor diagnostiek op het terrein van de sociaal-emotionele ontwikkeling van kinderen is VISEON 2.0 niet geschikt. Onderzoek op dit terrein dient te gebeuren met gespecialiseerde instrumenten en door of onder supervisie van een psycholoog of orthopedagoog. VISEON 2.0 is slechts te gebruiken om te bepalen of het sociaal-emotioneel functioneren van een leerling in het algemeen of op specifieke onderdelen nadere aandacht behoeft of niet. Het instrument is uitdrukkelijk bestemd voor gebruik door de groepsleerkracht.

De meetpretentie van VISEON 2.0 heeft geen betrekking op het cognitieve functioneren. Wel stelt VISEON 2.0 de leerkracht in staat om de samenhang tussen (aspecten van) sociaal-emotioneel functioneren en het verloop van onderwijs- en leerprocessen in kaart te brengen. De resultaten op de leerkrachtlijst kunnen

² Bij de afname en in de handleiding voor de leerkracht wordt dit onderdeel aangeduid als 'Spreken in de klas'.

gelegd worden naast de uitkomsten van toetsen die een indruk geven van de cognitieve mogelijkheden en leerprestaties van de leerlingen. Binnen het Computerprogramma LOVS kan men hiervoor de cognitief gerichte toetsen uit het leerlingvolgsysteem gebruiken.

2.2 Doelgroep

De leerkrachtlijst van VISEON 2.0 is bestemd voor leerkrachten van groep 3 tot en met groep 8.

De leerlinglijst van VISEON 2.0 is geschikt voor leerlingen van groep 5 tot en met groep 8.

VISEON 2.0 is ontwikkeld en genormeerd voor leerlingen in het reguliere basisonderwijs, inclusief de zorgleerlingen op deze scholen. We hebben (nog) geen onderzoek gedaan met VISEON 2.0 in het speciaal onderwijs en speciaal basisonderwijs. De schalen van VISEON 2.0 zullen zeker ook bruikbaar zijn voor leerlingen in het speciaal (basis)onderwijs. De normen kunnen echter niet één-op-één worden toegepast. Dit hangt samen met de aard van het instrument die anders is dan die van leervorderingstoetsen.

Bij leervorderingstoetsen wordt de leerling in staat gesteld om aan de hand van opgaven *op objectieve wijze te tonen wat hij weet en kan*. Dat maakt het mogelijk om de toetsresultaten van leerlingen in het speciaal (basis)onderwijs langs dezelfde meetlat te houden als de toetsresultaten van leerlingen in het reguliere basisonderwijs.

Bij VISEON 2.0 is sprake van op observatie en interactie met de leerling gebaseerde rapportage door de leerkracht en zelfrapportage door de leerling. In de leerkrachtlijst *beoordeelt* de leerkracht het gedrag en het functioneren van de leerlingen. De ervaring leert dat de beoordelende leerkracht hierbij een relatief referentiekader hanteert dat hij steeds aanpast aan het gemiddelde niveau van de te beoordelen groep. Het gedrag van een leerling wordt dus beoordeeld in relatie tot het gedrag van zijn medeleerlingen in de groep. Daarbij houdt de leerkracht tevens rekening met gedrag dat typisch is voor een bepaalde leeftijd: wat kenmerkend is voor de gemiddelde ontwikkeling en het doorsnee functioneren in groep 3 ziet er heel anders uit dan de ontwikkeling en het functioneren dat men in groep 8 aantreft en mag verwachten. Gedrag dat opvalt in de ene groep, kan bovendien in een andere groep veelvuldig voorkomen en daarom niet opvallen. Meer concreet: in een groep rustige leerlingen zal een drukke leerling eerder opvallen dan in een groep die als geheel druk is. Iets soortgelijks kan bij *zelfrapportage* door de leerling aan de orde zijn. Ook de leerling zal in de regel bij de evaluatie van zijn eigen gedrag (het gedrag van medeleerlingen in) de groep als referentiekader gebruiken.

Er is dus – in tegenstelling tot de situatie bij leervorderingstoetsen – geen sprake van één objectieve meetlat die alle leerkrachten (en leerlingen) gebruiken om het gedrag mee te vergelijken. Dat betekent dat de uitkomsten van VISEON bij leerlingen in het speciaal (basis)onderwijs niet iets zeggen over hoe hun gedrag is vergeleken met de gemiddelde leerling in het basisonderwijs, maar over hun gedrag ten opzichte hun groepsgenoten in het speciaal (basis)onderwijs. Gezien de verwachting dat het karakteristieke leerlinggedrag in het speciaal (basis)onderwijs anders zal zijn dan in het reguliere basisonderwijs, kunnen we de normering gebaseerd op het basisonderwijs niet toepassen en interpreteren op de manier zoals dat gebeurt bij de leervorderingstoetsen, namelijk als vergelijking met een leerling in het regulier basisonderwijs.

Wel is het mogelijk om de scores van kinderen in het speciaal (basis)onderwijs te interpreteren binnen de eigen klas of groep. Daarbij kan men gebruikmaken van de verschillen in normscores binnen de groep: als een kind in vergelijking met groepsgenoten opvalt, kan men aan deze uitkomst wel degelijk waarde hechten. Dat geldt ook voor de interpretatie van normscores over meerdere afnamemomenten heen.

Als een kind dat bij een eerste afname van VISEON 2.0 met zijn normscore voor bijvoorbeeld storend gedrag erg opvalt, en dit gedrag bij latere afnames volgens de beoordelende leerkracht steeds minder laat zien, kan hieraan ook binnen het speciaal (basis)onderwijs de conclusie worden verbonden dat het kind een positieve ontwikkeling doormaakt.

VISEON 2.0 is geschikt om in te vullen door dyslectische leerlingen en leerlingen die anderszins beschikken over een beperkte leesvaardigheid. De vragen in de leerlinglijst kunnen desgewenst via de computer worden voorgelezen.

2.3 Gebruiksdoel en functie

Het doel van VISEON 2.0 is het in kaart brengen en volgen van het sociaal-emotioneel functioneren van leerlingen en de ervaren kwaliteit van de leeromgeving in groep 3 tot en met groep 8. De leerkrachtlijst is gebaseerd op de gedachte dat leerkrachten die werkzaam zijn in deze groepen praktijkdeskundigen zijn, die uitstekend in staat mogen worden geacht om het gedrag en de ontwikkeling van leerlingen te observeren en te beoordelen. Het instrument biedt hierbij steun in de vorm van een aantal uitspraken of stellingen (observatiepunten) aan de hand waarvan de leerkracht aan kan geven in welke mate deze van toepassing zijn op de leerling. De leerlinglijst is vergelijkbaar opgebouwd. Leerlingen geven voor een aantal uitspraken aan in hoeverre ze vinden dat deze op hen van toepassing zijn. De unieke combinatie van leerkrachtlijst en leerlinglijst maakt het mogelijk tot een afgewogen en genuanceerd oordeel te komen, in gesprek te gaan en actie te ondernemen. Zowel de leerkracht- als de leerlinglijst zijn bedoeld als signaleringsinstrumenten. Ze zijn bedoeld voor het handelen in de klas, dus niet om risicogedrag op te sporen.

De leerkracht- en leerlinglijsten van het hoofdinstrument van VISEON 2.0 bieden de leerkracht een breed overzicht van het sociaal-emotioneel functioneren. Mocht uit het hoofdinstrument blijken dat er behoefte is aan gerichtere informatie, dan kan de leerkracht ervoor kiezen om bij één of meerdere domeinen aanvullende leerlinglijsten te gebruiken. Er zijn uiteenlopende redenen om aanvullend onderzoek te doen. De aanvullende leerlinglijsten kunnen ingezet worden bij specifieke problemen (individueel dan wel op groepsniveau), wanneer in het schoolplan bepaalde onderwerpen extra aandacht krijgen of wanneer de scores op het hoofdinstrument anders zijn dan verwacht. De aanvullende leerlinglijsten geven waar nodig extra inzicht en houvast om te bepalen waarop de school gaat inzetten in het plan van aanpak. Dit kan zowel op het niveau van de leerling, als op het niveau van de groep, als op het niveau van de school zijn. Welke instrumenten gebruikt worden, bepaalt de leerkracht dus zelf in overleg met de intern begeleider, orthopedagoog of andere gedragsdeskundige binnen de school. Wanneer men een nauwkeuriger en gedetailleerder beeld van het functioneren en de ontwikkeling van een leerling wenselijk acht, is vervolgonderzoek noodzakelijk. Het kan nodig zijn om hiervoor een gedragswetenschapper in te schakelen.

De lijst is gestandaardiseerd, in zoverre dat zowel leerkrachten als leerlingen de lijsten invullen aan de hand van uitspraken die voor iedereen gelijk zijn. Het grote voordeel daarvan is dat leerlingen op deze manier, via de normering, vergeleken kunnen worden met de gegevens die verzameld zijn bij een grote en representatieve referentiegroep die de lijsten op dezelfde manier hebben ingevuld. In kaart brengen en volgen betekent hier, net als bij andere instrumenten in het Cito Volgsysteem primair en speciaal onderwijs, dat er na elke afname uitspraken kunnen worden gedaan over het functioneren van de leerling op dat moment.

In de rapportages van VISEON 2.0 worden de scores weergegeven als percentielen en als *signalen*. De percentielscores zijn weergegeven door middel van staafdiagrammen. Er zijn drie signalen in de kleuren groen voor niet afwijkend (N), oranje voor afwijkend (A) en rood voor zeer afwijkend (Z). Een signaal geeft aan of en in welke mate de score van een leerling afwijkt van leerlingen in de landelijke normgroep. Voor elk aspect ontvangt de leerkracht een apart signaal.

Signaal	Interpretatie
N	De score van de leerling wijkt niet af van de meeste andere leerlingen in de normgroep. 80% van de leerlingen behaalt de score N.
A	De score van de leerling wijkt af van de score van de meeste andere leerlingen in de normgroep. 15% van de leerlingen behaalt de score A.
Z	De score van de leerling wijkt zeer af van de score van de meeste andere leerlingen in de normgroep. 5% van de leerlingen behaalt de score Z.

De signalen geven aan of en in welke mate een leerling zich min of meer gedraagt zoals dat van een leerling van zijn of haar jaargroep mag worden verwacht. Daar kan dan vervolgens een passend

begeleidingstraject bij gezocht worden. Voor leerlingen die zeer afwijkend scoren in vergelijking met leerjaargenoten kan dit in de vorm van extra aandacht en stimulering die erop gericht zijn mogelijke achterstanden weg te werken. Bij zeer afwijkende scores kan ook de interpretatie passen dat er sprake is van een ontwikkelingsachterstand of dat de leerling door bepaalde factoren, omstandigheden of gebeurtenissen in zijn of haar ontwikkeling wordt bedreigd. In zo'n geval kan men de leerling voor nader onderzoek doorverwijzen naar een gedragswetenschapper. Dit betekent echter niet dat de leerling – alleen op grond van deze lage score – als probleemleerling moet worden aangemerkt. VISEON 2.0 is niet bedoeld als screenend instrument. Om het zo te kunnen gebruiken, zijn er onderzoeksgegevens nodig waaruit blijkt dat leerlingen die op grond van een of meer scores als risicoleerling zijn aangemerkt, ook daadwerkelijk problematisch zijn of later zijn geworden. Dit onderzoek is nog niet beschikbaar. Wel zijn er onderzoeksgegevens die erop wijzen dat het instrument in dit opzicht waardevol is (zie hiervoor hoofdstuk 6 over validiteit).

Het systematisch volgen van individuele leerlingen

Bij de meeste instrumenten (toetsen) in het Cito Volgsysteem primair en speciaal onderwijs kan men voor een bepaald kind of een groep kinderen de ontwikkeling van een vaardigheid afleiden uit de zogeheten vaardigheidsscore, een score op één en dezelfde uni-dimensionele vaardigheidsschaal. Vanwege de aard van het instrument en de wijze waarop de items zijn geformuleerd, kon een dergelijke vaardigheidsschaal bij VISEON 2.0 niet worden geconstrueerd.

Dit houdt in dat men het functioneren van een kind alleen kan volgen aan de hand van de percentielen en signalen (niet afwijkend, afwijkend en zeer afwijkend). Heeft een leerling bijvoorbeeld een jaar geleden nog een ruwe score gekregen die resulteerde in het signaal 'afwijkend' en krijgt dezelfde leerling nu op basis van zijn ruwe score het signaal 'zeer afwijkend', dan kan men alleen in relatieve zin constateren dat zijn niveau van functioneren erop achteruit is gegaan. Uiteraard moet de leerkracht daarbij ook het verschil tussen de percentielscores van verschillende afnames in de overwegingen betrekken die hij/zij kan afleiden uit het verschil tussen de lengte van de staven in opeenvolgende leerlingoverzichten. Voor leerlingen op de grens tussen twee opeenvolgende signalen kan een verschil van een percentieelpunt tussen twee afnames er immers toe leiden dat het signaal verandert van Niet afwijkend naar Afwijkend of van Afwijkend naar Zeer afwijkend (en vice versa). Zie hiertoe ook paragraaf 4.4 in het hoofdstuk over de normering van VISEON 2.0.

Leerlingrapportage

In de leerlingrapportage staan signalen weergegeven voor de onderscheiden aspecten van het sociaal-emotioneel functioneren. De aspecten bij het domein leeromgeving staan niet op de leerlingrapportage, deze staan alleen op het groepsrapport. We rapporteren geen totaalscore voor sociaal-emotioneel functioneren, omdat een totaalscore geen zinvolle informatie geeft voor het handelen. Daarnaast rapporteren we om dezelfde reden ook per domein geen totaalscore.

Hiernaast is een voorbeeld opgenomen van een leerlingrapportage. In de handleiding voor de gebruiker (Cito, 2015) wordt deze rapportage om technische redenen categorieënanalyse genoemd. Wanneer er geen signaal staat, betekent dit dat een leerling de vragen bij dit aspect niet allemaal heeft ingevuld.

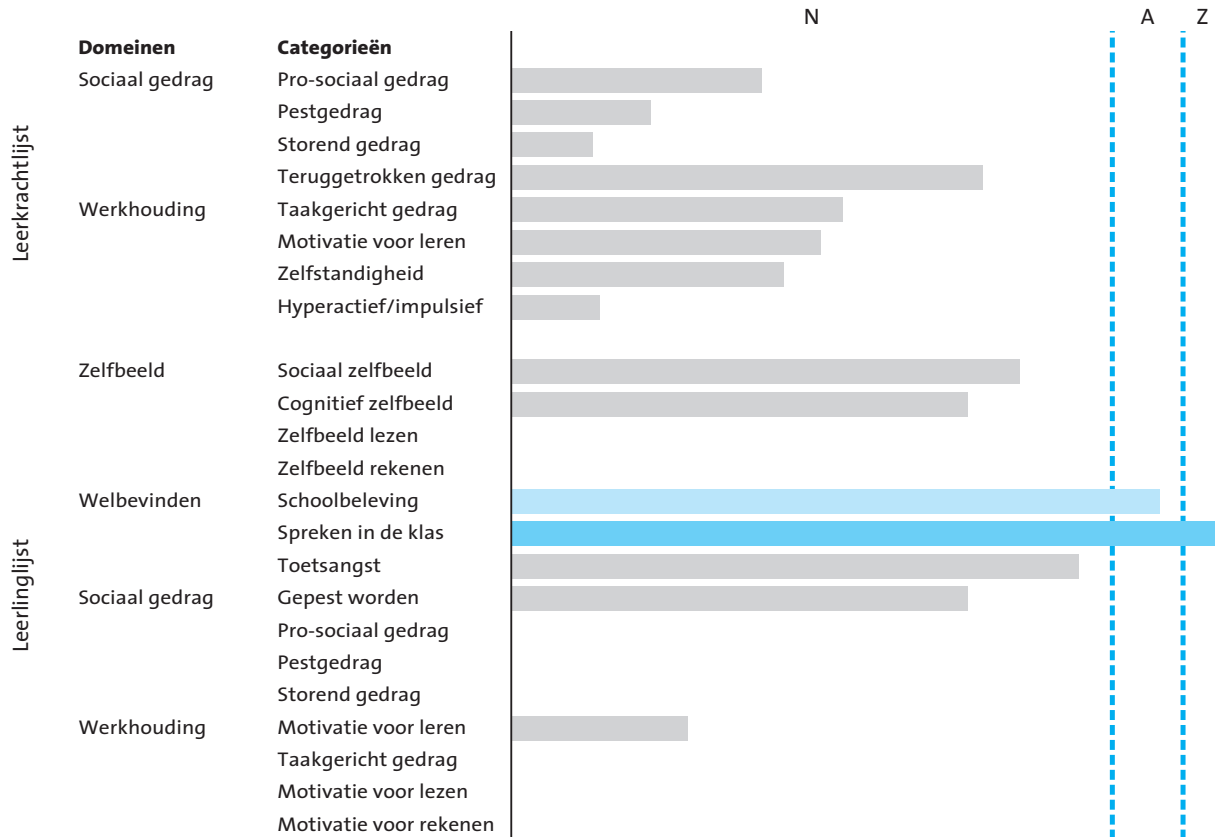
Figuur 1 Categorieënanalyse leerling (voorbeeld)

Cito - DEMO versie Arnhem



Categorieënanalyse leerling

Leerling: **Tijn**
Groep: **7**



Voorbeeld

Het rapport is een voorbeeldrapport van de fictieve leerling Tijn. Tijn zit in groep 7. De leerkracht had Tijn niet gesignaleerd als een jongen die extra aandacht op het gebied van sociaal-emotioneel functioneren vraagt. Uit de leerlinglijst blijkt dat Tijn een afwijkende score (A) heeft op het aspect schoolbeleving. Dit betekent dat hij zich niet altijd prettig voelt op school. De leerkracht neemt een aantal aanvullende instrumenten af om meer zicht te krijgen op waarom Tijn dit zo ervaart. Hierbij valt de zeer afwijkende score (Z) bij *Spreken in de klas* op. De leerkracht bespreekt dit met Tijn en maakt met hem een plan van aanpak op welke manier hij op een veilige manier hiermee kan oefenen. De leerkracht nodigt de ouders van Tijn uit om samen met Tijn te bespreken wat Tijn nodig heeft om zich prettiger te voelen in de groep.

Het voorbeeld van Tijn maakt duidelijk welke gegevens VISEON 2.0 oplevert op leerlingniveau. Door het minimaal één keer per jaar invullen van de leerkrachtilijst en de leerlinglijst (vanaf groep 5) en het afnemen van aanvullende instrumenten, is het sociaal-emotioneel functioneren van een leerling over de tijd te volgen. Dit maakt het mogelijk om bij een ongewenste ontwikkeling tijdig in te grijpen. Bovendien stelt het leerkrachten in staat om te bepalen of er sprake is van een zekere consistentie in het gedrag van een leerling op de verschillende aspecten. Tot slot kan een leerkracht aan de hand van deze gegevens het sociaal-emotioneel functioneren van een leerling afzetten tegen zijn of haar prestaties op de diverse leergebieden.

Behalve informatie op het niveau van de individuele leerlingen (in het leerlingrapport) kan het Computerprogramma LOVS ook informatie verschaffen in de vorm van groepsoverzichten. We gaan daar in dit verband niet verder op in; voor meer informatie over alle rapportagemogelijkheden verwijzen we naar de handleiding.

2.4 Theoretische inkadering

Bij de ontwikkeling van VISEON 2.0 was een belangrijk praktisch uitgangspunt dat leerkrachten goed in staat zijn het sociaal-emotioneel functioneren van hun leerlingen te beoordelen aan de hand van het concrete en direct waarneembare gedrag van de leerling op school. Dit op voorwaarde dat zij voldoende tijd hebben gehad om de leerling te leren kennen en de leerling voldoende tijd heeft gehad om te wennen aan de groep en de leerkracht. Daarnaast wordt met VISEON 2.0 ook het beeld dat een leerling van zichzelf heeft in kaart gebracht. Hoewel de leerling als informant over zijn eigen sociaal-emotioneel functioneren beperkingen kent, is het waardevol om deze informatie te verzamelen. Op de eerste plaats zijn er eigenschappen waarover men de leerling (weliswaar pas vanaf een bepaalde leeftijd) bij uitstek zelf aan het woord moet laten. Het zelfbeeld is daar een belangrijk voorbeeld van. Zelfrapportage is hierbij een gebruikelijk middel (vergelijk Harter, CBSK). Op de tweede plaats wilden we in VISEON ook aandacht besteden aan het klassenklimaat (als 'som' van de evaluatie van de sociale, veilige en ordelijke leeromgeving door alle klasgenoten afzonderlijk). En ten slotte is het waardevol om op een aantal kenmerken of eigenschappen de eigen inschatting door de leerling zelf te leggen naast die van de leerkracht.

Het is van belang dat leerkrachten in het primair onderwijs aandacht besteden aan het sociaal-emotioneel functioneren van hun leerlingen. In de eerste plaats omdat de kerndoelen voor het primair onderwijs duidelijk maken dat het onderwijs ook gericht moet zijn op leergebiedoverstijgende inhouden (Besluit kerndoelen primair onderwijs; Ministerie van OCW, 1998) en het sociaal-emotioneel functioneren van leerlingen hier deel van uitmaakt. In de tweede plaats, omdat er sprake is van een duidelijke wisselwerking tussen de sociaal-emotionele ontwikkeling van een leerling en zijn of haar ontwikkeling op andere terreinen. Problemen op het sociaal-emotionele vlak kunnen het onderwijsleerproces negatief beïnvloeden. Maar evenzeer is het mogelijk dat leerproblemen een storende invloed hebben op de ontwikkeling van een leerling op het sociaal-emotionele vlak.

Bij de constructie van VISEON 2.0 stond een aantal theoretische overwegingen centraal, waarbij we ter onderbouwing van de keuze van domeinen en aspecten zowel vanuit een ontwikkelingspsychologisch als een onderwijskundig perspectief vertrokken. Vanuit het *ontwikkelingspsychologisch perspectief* waren vragen belangrijk als: wat weten we over de sociaal-emotionele ontwikkeling en welke factoren en processen liggen eraan ten grondslag? Daarbij hadden we geen specifieke ontwikkelingspsychologische theorie op het oog, wel meer algemene ontwikkelingsprincipes waarbij uiteindelijk het begrip ontwikkelingsopgave of ontwikkelingstaak voor ons doel goede diensten heeft bewezen. Daarnaast hebben we ons laten leiden door een *onderwijskundig perspectief*. Daarom hebben we ons georiënteerd op de (communale) onderwijsdoelen die voor deze vorm van educatie zijn geformuleerd. Het hoeft geen verbazing te wekken dat communale onderwijsdoelen en ontwikkelingstaken in grote mate convergeren: onderwijsdoelen zijn immers niet onafhankelijk van de normale ontwikkelingsmogelijkheden van kinderen op te stellen.

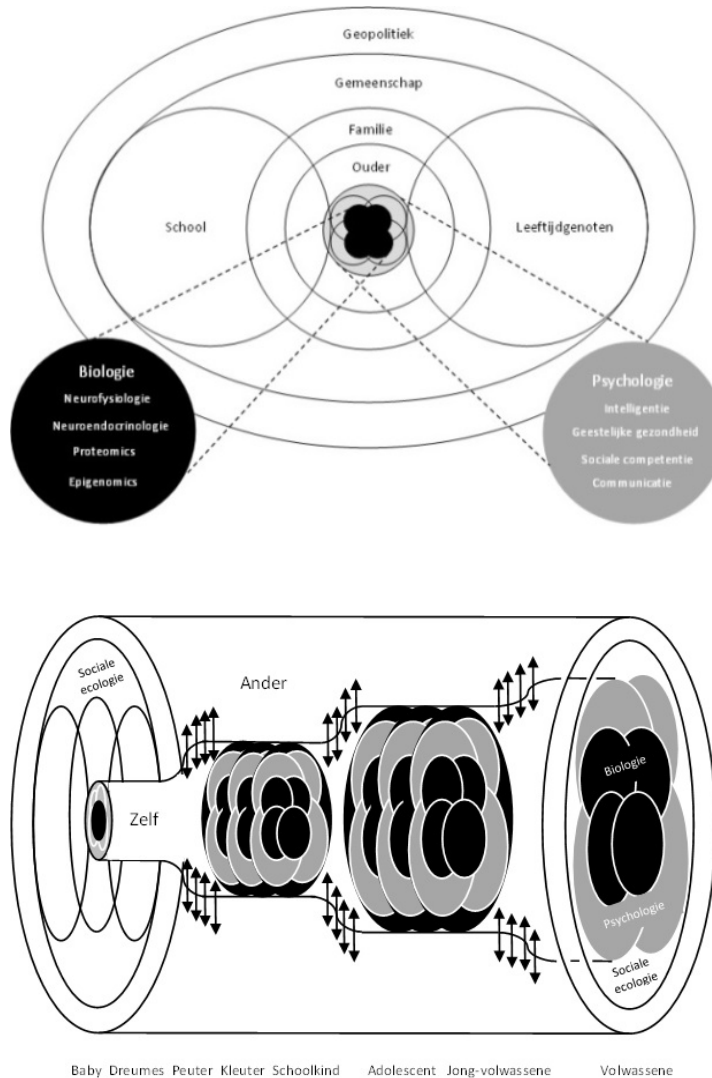
Het ontwikkelingspsychologische perspectief

Een groot deel van de ontwikkelingspsychologische theorievorming houdt zich bezig met de vraag of en hoe ontwikkeling zich voltrekt onder invloed van genetisch-biologische factoren (*nature*) enerzijds en de sociale en culturele omgeving (*nurture*) anderzijds. Daarbij gaat het vooral om de vraag hoe *nature* en *nurture* samen, in hun onderlinge samenspel, de ontwikkeling bepalen. Sameroff (2010) geeft een goed overzicht van de complexe processen waarin dit samenspel tot uitdrukking komt. Hij schetst vier verschillende 'modellen' die in de ontwikkeling tegelijkertijd aan de orde zijn:

1. *Het individueel persoonlijk veranderingsmodel*. Het model dat men hier kiest, geeft aan hoe men aankijkt tegen persoonlijke verandering. Veronderstelt men de aanwezigheid van relatief stabiele

trekken die nauwelijks aan verandering onderhevig zijn, is er sprake van (kwantificeerbare) groei of voltrekt de verandering zich in stadia die steeds resulteren in nieuw en kwalitatief verschillend gedrag?

Figuur 2 Geïntegreerde ontwikkelingspsychologische theorie in cross-sectioneel (boven) en longitudinaal (onder) perspectief (afbeeldingen ontleend aan Sameroff, 2010)



2. **Het contextueel model.** Gedrag en ontwikkeling kunnen niet los worden gezien van de sociale context. In de klassieke analyse van Bronfenbrenner (1977) zoals geciteerd in Sameroff (2010) kan men in de sociale context verschillende lagen onderscheiden (micro-, meso- en macrosysteem, exo- en chronosysteem). Kenmerkend is dat het zich ontwikkelende individu met de leeftijd in steeds breder worden sociale systemen functioneert (ouders – gezin – school – peergroep et cetera). Hoe dat gebeurt is cultureel-maatschappelijk en historisch bepaald. Voor de school als sociale context kan men op zijn minst stellen dat het kind een groot deel van zijn wakkere uren op school doorbrengt en er in zijn ontwikkeling en functioneren in zekere mate door wordt bepaald.
3. **Het regulatief model.** Hierbij wordt het zich ontwikkelende kind opgevat vanuit een dynamische systeembenadering, waarin (in afnemende mate) door anderen gereguleerde en (in toenemende mate) door het kind zelf gereguleerde elementen zijn te onderscheiden. Dit systeem heeft een zogeheten transactioneel karakter: niet alleen wordt het kind beïnvloed door zijn omgeving, het kind beïnvloedt ook zelf die omgeving, verandert deze daardoor en zet daarmee tot op zekere hoogte de eigen ontwikkeling naar zijn hand.

4. *Het representatiemodel.* Ten slotte is het van belang aandacht te schenken aan de interne representatie van de buitenwereld en van de ervaringen die het kind opdoet. Afhankelijk van de specifieke theorie spreken we hier over cognitieve schema's over de buitenwereld en interne werkmodellen als afspiegeling van de sociale relaties waarmee het kind ervaring opdoet. Andere termen die in dit opzicht relevant zijn, zijn zelfbeeld, competentiebeleving, zelfwaardering en, belangrijker naarmate men ouder wordt, het eigen, autobiografische levensverhaal ('narratief').

Sameroff integreert deze vier invalshoeken, respectievelijk modellen tot één geïntegreerde ontwikkelingspsychologische theorie die in cross-sectioneel en longitudinaal perspectief is af te beelden zoals in figuur 2.

Een geïntegreerde theorie als deze is vooral verhelderend als men tot doel heeft de belangrijkste factoren en processen in beeld te krijgen die bepalen hoe ontwikkeling zich voltrekt. Bij zeer jonge kinderen is voor deze ontwikkeling vooral het hoge tempo kenmerkend. Hun gedrag en functioneren wordt in hoge mate bepaald door de snelle en massieve veranderingen die zich voltrekken in het brein (voor een overzicht zie bijvoorbeeld Johnson, 2011). Bij kinderen in de basisschoolleeftijd ligt het tempo van de ontwikkeling beduidend lager, maar het is nog altijd hoog vergeleken met de structurele veranderingen op latere leeftijd. Deze en andere genetisch gestuurde biologische veranderingen geven aanleiding tot psychologische ontwikkelingen binnen verschillende domeinen (bijvoorbeeld intelligentie, sociale competentie, identiteit), waarbij het kenmerkend is dat deze zich voltrekken in steeds breder wordende sociale contexten (gezin, school, buurt, maatschappij). De nadruk op processen en verklaringen kan weliswaar verhelderen waarom kinderen – gemiddeld gesproken – binnen een specifieke culturele en historisch bepaalde context in staat zijn tot bepaald gedrag (en eerder niet), toch biedt een procesmatige 'formatieve' aanpak weinig aanknopingspunten als men zicht wil krijgen op de typische ontwikkelingsmijlpalen die men voor een bepaalde leeftijdsrange zou willen beschrijven. Daarom is het handig om nog een vijfde invalshoek toe te voegen, namelijk het 'summatieve' perspectief dat ontwikkelingsopgaven of -mijlpalen kunnen bieden. In de rest van deze paragraaf lichten we het concept ontwikkelingsopgaven nader toe.

Theorieën over ontwikkelingsopgaven of -taken ('developmental tasks') vinden hun oorsprong in het midden van de vorige eeuw. De termen worden soms zonder betekenisverschil door elkaar heen gebruikt, soms is sprake van onderscheid, waarbij het begrip 'ontwikkelingsopgave' wordt gehanteerd om een belangrijk ontwikkelingspsychologisch thema aan te duiden (bijvoorbeeld het (kunnen) aangaan van adequate sociale relaties met andere kinderen), terwijl men het begrip 'ontwikkelingstaak' reserveert voor een aantal concretere deeltaken die daarin zijn te onderscheiden. Wij zullen in het navolgende vooral het begrip ontwikkelingsopgave in zijn brede betekenis gebruiken.

De term ontwikkelingsopgave werd geïntroduceerd door Havighurst (1948, maar zie ook Havighurst, 1972). Ontwikkelingsopgaven zijn te omschrijven als "psychologische opdrachten waarmee elk mens in de loop van zijn ontwikkeling in een min of meer vaste volgorde geconfronteerd wordt" (Meij, 2011, p. 4). Ontwikkelingsopgaven zijn aan diverse bronnen te ontleen. De belangrijkste zijn de biologische en psychologische rijpingsprocessen zoals leren lopen en spreken en het verkrijgen van controle over blaas en darmen. Maar daarnaast worden ontwikkelingsopgaven ook bepaald door persoonlijke waarden (bijvoorbeeld het maken van adequate keuzes met betrekking tot opleiding en beroep) en door culturele en maatschappelijke druk (bijvoorbeeld leren lezen, kunnen functioneren in een democratische samenleving). Het begrip ontwikkelingsopgave is bezig aan een revival (vergelijk bijvoorbeeld Riksen-Walraven, 1989; Meij, 2011; Broer, Haverhals & De Bruin, 2012). Dat is wellicht te verklaren door het feit dat ontwikkelingsopgaven gemakkelijk zijn te 'flankeren' met bijbehorende opvoedingsopgaven.

Het is opmerkelijk dat er in de literatuur op verschillende manieren tegen het concept ontwikkelingsopgave wordt aangekeken. Soms wordt vooral benadrukt wat het kind op een bepaald moment in de tijd wordt geacht te kunnen. Dat wordt dan benoemd als een concreet te realiseren ontwikkelingsmijlpaal, die het kind al dan niet bereikt heeft, waarbij men de indruk wekt een dichotomie te kunnen hanteren: wel gerealiseerd versus niet gerealiseerd. Aan de andere kant wordt soms juist het proces benadrukt: een ontwikkelingstaak heeft betrekking op het zich verwerven van bepaalde vaardigheden en het kind is daar in een bepaalde

ontwikkelingsfase volop mee bezig. Het accent ligt dan dus vooral op “work in progress”. Men heeft bij dit gebruik van het concept minder de neiging om in alles-of-niets termen te spreken en denkt vooral in termen van ‘mate waarin’. Mogelijk hangt de wijze waarop men het concept ontwikkelingstaak hanteert samen met het gemak waarmee men een concrete en objectieve mijlpaal kan beschrijven. We zien dat dit vooral bij de vroege motorische en spraak-taalontwikkeling goed mogelijk is: de baby kan zich bijvoorbeeld wel of niet van de buik op de rug omrollen en vice versa. Of, ander voorbeeld: het kind kan wel of niet (al) spreken in twee-woord-zinnen. Het al dan niet beheersen van een vaardigheid is bij deze voorbeelden vrij goed in objectief waarneembare termen te beschrijven en te observeren. Onze indruk is dat dit bij de sociaal-emotionele ontwikkeling veel minder goed mogelijk is: geaccepteerd worden door leeftijdgenoten bijvoorbeeld (vergelijk Meij in het citaat hieronder) is niet een alles-of-niets kwestie, maar iets wat het kind in bepaalde mate realiseert. Bovendien is het een taak die men op verschillende leeftijden in dezelfde termen onder woorden kan brengen, maar waarbij het soms om heel verschillende competenties gaat.

De globale ontwikkelingsopgave voor 4- tot 12-jarigen zoals beschreven door Meij luidt als volgt (2011, p. 5): “In deze periode neemt de autonomie verder toe. Dit uit zich ook in het snel toenemend vermogen van het kind om voor zichzelf te zorgen (zichzelf wassen en aan- en uitkleden, eten). Verder begint het vermogen tot decentratie zich te ontwikkelen (andermans perspectief leren zien). Tot slot wordt het kind in deze periode geconfronteerd met de maatschappelijke eis om te leren lezen, schrijven en rekenen. Daarnaast moet het zich in de schoolse situatie ook een taakhouding eigen maken om zich gedurende (steeds) langere tijd te concentreren op schoolse taken. In de omgang met de leerkracht is het noodzakelijk om de leerling-rol aan te nemen. Naast de schoolse vaardigheden (lezen, schrijven, rekenen), is ook het vermogen om wederkerige relaties op te bouwen met vriendjes/vriendinnetjes en het geaccepteerd worden door de groep een belangrijke ontwikkelingstaak in deze periode. In de omgang met leeftijdgenoten moet het kind leren zijn egocentrische houding steeds meer te laten varen.”

We hebben afgezien van een nadere specificatie van deze globale ontwikkelingsopgaven als basis voor de aspecten, respectievelijk schalen die in VISEON 2.0 zijn opgenomen. We hebben deze keuzes vooral gelegitimeerd vanuit een onderwijskundig perspectief (zie hieronder). Daarbij gelden de volgende overwegingen:

- VISEON 2.0 is geconstrueerd met het oog op gebruik in het onderwijs. Daarom wilden we vooral focussen op het sociaal-emotioneel functioneren *in de context van de school*.
- Leerdoelen in het onderwijs (kerndoelen, tussendoelen, streef- of ontwikkelingsdoelen) lopen grotendeels parallel met de omschrijving van ontwikkelingstaken. Doelen zijn immers alleen zinvol tot uitgangspunt te nemen voor het onderwijs wanneer zij aansluiten bij het ontwikkelingsniveau van de kinderen voor wie dat onderwijs is bedoeld. Daarbij is het vooral een academische (en lastig te beantwoorden) vraag, of het hier nu gaat om ontwikkeling of over leren. Hetzelfde geldt min of meer parallel voor de vraag in welke mate we kunnen spreken van ‘sociale opbrengsten van het onderwijs’: voor sommige leerdoelen zal de invloed van de school op het realiseren ervan groter zijn dan voor andere.
- Niet alle schalen in VISEON 2.0 veronderstellen noodzakelijkerwijs of in even sterke mate de ontwikkeling van vaardigheden als basis voor het functioneren. Daarnaast is ‘ontwikkeling’ lastig rechtstreeks vast te stellen. De meeste schalen die in VISEON zijn ondergebracht representeren aspecten van de kwaliteit van individueel sociaal-emotioneel functioneren, waarbij verondersteld wordt dat deze kwaliteit afhankelijk is van een complex van factoren dat voor elk gemeten aspect een specifieke samenstelling kent. Zo veronderstelt bijvoorbeeld een adequate taakhouding een bepaalde vaardigheidsontwikkeling bij de leerling die hem in staat stelt zich langere tijd op een bepaalde taak te concentreren zonder zich te laten afleiden door de omgeving. Tegelijkertijd zijn echter ook factoren als groepsgrootte, groepssamenstelling (zitten er veel drukke, orde-verstorende kinderen in de groep) en wijze van orde handhaven door de leerkracht van belang. Hoe een specifieke leerling op een gegeven afnamemoment op een bepaald aspect functioneert, is dus deels afhankelijk van zijn individuele vaardigheidsontwikkeling en deels van andere factoren, waaronder de aard van de context (leeromgeving).

We zijn ervan uitgegaan, dat de keuze van domeinen en aspecten het beste te verantwoorden is vanuit de leerdoelen voor het onderwijs, waarmee we het ontwikkelingspsychologische perspectief voldoende verdisconteerd achten. In de volgende sectie gaan we op deze leerdoelen uitgebreider in.

Onderwijskundig perspectief

Gewoonlijk worden opbrengsten van het onderwijs niet omschreven in termen van ontwikkelingsopgaven, maar in termen van onderwijsdoelen (kerndoelen, tussendoelen). De SLO biedt een overzicht van de kern- en tussendoelen zoals deze kunnen worden onderscheiden in het basisonderwijs. Ook de CED-Groep heeft een leerlijn ontwikkeld ten aanzien van de sociaal-emotionele ontwikkeling. Hieronder staan bij de verschillende domeinen en aspecten uit VISEON 2.0 voorbeelden van leerdoelen uit deze leerlijnen gegeven. In de figuur hiernaast staat een overzicht van de verschillende leerdoelen op het gebied van sociaal-emotioneel functioneren.

	Aspecten VISEON 2.0	Leerdoelen groep 3/4	Leerdoelen groep 5/6	Leerdoelen groep 7/8
domein zelfbeeld*	<ul style="list-style-type: none"> – sociaal zelfbeeld – cognitief zelfbeeld – zelfbeeld lezen – zelfbeeld rekenen 	Ontdekken de eigen mogelijkheden, kenmerken, wensen, gevoelens, beperkingen en voorkeuren.	Ontdekken dat veel eigenschappen en vaardigheden niet vastliggen, maar ontwikkeld kunnen worden en ontwikkelen een zelfbewuste houding.	Ontdekken dat veel eigenschappen en vaardigheden niet vastliggen, maar ontwikkeld kunnen worden en ontwikkelen een zelfbewuste houding.
domein welbevinden*	<ul style="list-style-type: none"> – schoolbeleving – spreekangst – toetsangst 	Gaan bewust om met gevoelens, wensen en opvattingen en kunnen deze voor anderen begrijpelijk uiten.	Gaan bewust om met gevoelens, wensen en opvattingen en kunnen deze voor anderen begrijpelijk uiten.	Gaan bewust om met gevoelens, wensen en opvattingen en kunnen deze voor anderen begrijpelijk uiten.
domein sociaal gedrag*	<ul style="list-style-type: none"> – pro-sociaal gedrag – pestgedrag – storend gedrag – teruggetrokken gedrag – gepest worden 	Zien in hoe conflictsituaties in de eigen omgeving ontstaan en welke oplossingen mogelijk zijn.	Kunnen omgaan met conflictsituaties in de eigen omgeving door voor alle partijen passende oplossingen te vinden.	Kunnen omgaan met conflictsituaties in de eigen omgeving door voor alle partijen passende oplossingen te vinden.
		Zien in dat conflicten veroorzaakt worden door tegenstellingen in belangen, opvattingen en gevoelens.	Kunnen samenwerken in een groep	Kunnen samenwerken in een groep.
		Ontdekken dat iedereen deel uitmaakt van verschillende groepen waarin verschillende rollen worden vervuld.	Kunnen zich inleven in gevoelens, wensen en opvattingen van anderen.	Geven van de diverse leefeenheden kenmerken, overeenkomsten en verschillen aan.
		Staan open voor gevoelens, wensen en opvattingen van anderen.	Houden rekening met de ander en staan open voor de opvattingen van anderen.	Houden rekening met de ander en staan open voor de opvattingen van anderen.
		Kunnen zich inleven in gevoelens, wensen en opvattingen van anderen.	Onderkennen vooroordelen en discriminatie.	Houden rekening met verschillen en overeenkomsten tussen etnische groepen.
		Zien in dat regels door mensen zijn bedacht en ook door mensen kunnen worden veranderd of afgeschaft.	Ontdekken de relatie tussen rol en verwachtingspatroon en de wijzigingen in beide onder invloed van tijd en (sub)cultuur.	Noemen uitingsvormen van discriminatie en weten waardoor discriminatie kan ontstaan.
				Zien in hoe sekse-rollen ontstaan en worden aangeleerd.

	Aspecten VISEON 2.0	Leerdoelen groep 3/4	Leerdoelen groep 5/6	Leerdoelen groep 7/8
domein werkhouding**	<ul style="list-style-type: none"> – taakgericht gedrag – motivatie voor leren – zelfstandigheid – hyperactief/impulsief – taakgericht gedrag – motivatie voor lezen – motivatie voor rekenen 	Kiest uit twee manieren om een taak aan te pakken de beste manier	Controleert zijn eigen taak en corrigeert waar nodig	Stelt prioriteiten wanneer er meerdere dingen tegelijk moeten gebeuren
		Kijkt hoe medeleerlingen een taak aanpakken en neemt de strategie over	Stelt zichzelf bij een bekende taak een tijdsdoel (dan moet ik het af hebben)	Geeft vooraf aan bij welk resultaat hij tevreden is met zijn werk
		Weet wanneer hij hulp nodig heeft	Werkt een uur zelfstandig aan een taak op een dag	Maakt een plan voor de aanpak van een klein project (volgorde van taken, duur van taken, tijdsindeling, benodigdheden)
		Zet door en past oplossingen toe bij een taak die hij moeilijk of eng vindt	Houdt zelf de tijd in de gaten om op tijd het werk af te krijgen	Zet zich langere tijd in voor een taak die hij niet leuk vindt of die niet lukt
		Hanteert de afgesproken regels om problemen op te lossen (bv. hulp van medeleerling of leerkracht, op zoeken)	Begint aan een volgende taak waarvan hij weet dat hij deze zelfstandig mag en kan doen	Beoordeelt of hij een uitgevoerde taak goed had voorbereid en uitgevoerd en verbindt hier consequenties aan voor de volgende keer
		Werkt 15-30 minuten zelfstandig aan een taak	Werkt netjes en/of snel afhankelijk van de eisen die aan de taakuitvoering worden gesteld	Beoordeelt of hij de juiste prioriteiten heeft gesteld en geeft aan welke gevolgen dit heeft voor een volgende planning
		Maakt een langdurige opdracht af als daar af en toe aanwijzingen voor worden gegeven	Reageert zonder boos te worden op kritiek op zijn werk	
		Zoekt een fout op als de leerkracht vraagt nog eens goed te kijken	Stelt vorderingen bij zichzelf vast (dit ging beter dan vorige keer)	

* Doelen afkomstig van <http://sociaalemotioneel.slo.nl/thema/algemeen/leerlijnen/>

** Doelen zijn een selectie van leerdoelen uit de leerlijn Leren leren van de CED-Groep. De doelen zijn slechts voorbeelden, de volledige leerlijn is te downloaden via www.cedgroep.nl

2.5 De keuze van domeinen en aspecten

Na de beschrijving van deze algemene en brede theoretische achtergronden van VISEON 2.0, gaan we in deze paragraaf in op de overwegingen die geleid hebben tot de uiteindelijke keuze voor de opzet en structuur van het instrument. Vervolgens zoomen we per domein verder in op de daarin ondergebrachte aspecten, geven een conceptuele definitie daarvan en bespreken we onze beweegredenen om deze op te nemen in VISEON 2.0.

In het voorafgaande is globaal toegelicht hoe de inhoud en structuur van VISEON 2.0 is terug te voeren op een ontwikkelingspsychologisch en onderwijskundig perspectief. Daarmee is nog niet in alle opzichten duidelijk waarom sommige domeinen en aspecten wel en andere niet in het instrument zijn opgenomen.

De keuze daarvan werd bepaald door een groot aantal, deels pragmatische overwegingen.

Op de eerste plaats hebben wij bij de keuze van relevante aspecten gebruikgemaakt van de domeinbeschrijving die is gemaakt voor de Peiling Sociale Competentie in jaargroep 8 (Kuhlemeier, Van Boxtel & Van Til, 2012). Deze vrij uitputtende inventarisatie leverde een veelheid van zogeheten 'indicatoren' op die onmogelijk alle in een handzame leerkracht- en leerlinglijst konden worden ondergebracht. Er moesten dus keuzes worden gemaakt.

Om een beredeneerde reductie tot stand te brengen werd in eerste instantie een drietal criteria toegepast:

- In hoeverre verwijst een indicator naar de kerndoelen van het onderwijs (zie ook paragraaf 2.4)?
- Hoe belangrijk is een indicator in termen van het toezichtkader van de Inspectie van het onderwijs?
- In hoeverre is een indicator relevant voor de leerprestaties en voor de vraag of een leerling erin slaagt zijn cognitieve mogelijkheden te realiseren?

Daarnaast speelde ook de continuïteit in relatie tot andere Cito-instrumenten (de observatielijsten voor baby's en dreumesen, respectievelijk peuters en VISEON voor kleuters) en de eerste versie van VISEON een belangrijke rol:

- Komt een indicator voor in de bestaande versie van VISEON (voor leerkracht- en leerlinglijst afzonderlijk)?
- Is er aansluiting te vinden bij de Cito Observatielijsten voor jonge kinderen, in het bijzonder bij VISEON voor Kleuters?

Belangrijk is ook dat we zoveel mogelijk wilden aansluiten bij wetgeving. In de WPO (Wet op het primair onderwijs) is vastgelegd dat het onderwijs zich in ieder geval (ook) moet bezighouden met de emotionele ontwikkeling en het verwerven van sociale vaardigheden. Leerlingen moeten op deze aspecten een ononderbroken leer- en ontwikkelingsproces kunnen doorlopen. Leerlingen die individuele zorg en begeleiding nodig hebben, moet deze geboden worden, waarbij de voortgang in de ontwikkeling (van alle kinderen) moet worden geregistreerd. De kerndoelen (34, 35, 37 en 38) die op deze aspecten betrekking hebben zijn echter betrekkelijk globaal geformuleerd en boden weinig echt concrete aanknopingspunten. Iets meer houvast boden de onderwijsdoelen die in de Preambule op de kerndoelen in het sociale domein worden opgesomd als van belang *voor alle leergebieden* (zie eerder in dit hoofdstuk). Ten slotte is het nog van belang dat scholen op grond van de Wet 'Actief Burgerschap en Sociale Integratie' van 2006 actief burgerschap en integratie dienen te bevorderen.

In deze eerste analyse werd besloten om de brede domeinen burgerschap en morele ontwikkeling niet in VISEON op te nemen. Voor *burgerschap* is een uitgebreid en gedetailleerd instrumentarium ontwikkeld door de groep rondom Ten Dam (vergelijk Ten Dam, Geijssel, Ledoux, Reurman, Keunen & Visser, 2010; Ten Dam, Geijssel, Reurman & Ledoux, 2010; Ten Dam, Volman, Westerbeek, Wolfram & Ledoux, 2003). VISEON 2.0 zou weinig kunnen toevoegen aan wat al in dit specifieke instrument is uitgewerkt.

Morele ontwikkeling is lastig in een instrument zoals we dat met VISEON 2.0 voor ogen hadden te operationaliseren. Door deze domeinen te laten vallen, beperkt VISEON zich dus, als instrument om sociale opbrengsten in kaart te brengen, tot het domein van de sociale competentie in engere zin, inclusief aspecten van school- en klasbeleving (school- en klasklimaat).

Vervolgens gingen we na welke elementen van sociale competentie in engere zin in het formele toezichtkader van de inspectie belangrijk zijn. Daarin wordt op de eerste plaats gesteld dat de sociale competenties van leerlingen op het niveau moeten liggen dat verwacht mag worden. De school moet een actief onderwijsaanbod hebben om deze competenties te ontwikkelen. Het toezichtkader bevatte een opsommende beschrijving van de aspecten van sociale competentie die voor het onderwijs relevant worden geacht. Uit de meetpretenties van instrumenten die de inspectie toelaatbaar vindt om sociale competentie mee te evalueren blijkt dat deze vrij sterk uiteenlopen. De nadruk lijkt echter te liggen op wat men zou kunnen benoemen als de interpersoonlijke competenties. Daarnaast is het opvallend dat de Inspectie veel aandacht vraagt voor veiligheid op school en de beleving daarvan. Leerlingen (en personeel) moeten zich veilig (kunnen) voelen op school. Ten slotte vraagt de Inspectie aandacht voor motivatie en werkhouding als eigenschappen die van belang zijn voor de wijze waarop leerlingen hun schoolloopbaan vervolgen.

Op grond van deze analyses van het toezichtkader werd besloten om in VISEON 2.0 in het hoofdinstrument in ieder geval ruimte vrij te maken voor:

1. De evaluatie van het interpersoonlijk ('sociaal') functioneren van leerlingen als indicator van hun sociale competentie en ontwikkeling.
2. De evaluatie van intrapersoonlijke kenmerken die belangrijk zijn voor de leerling als persoon en het leren op school.
3. De evaluatie van het school- en klasklimaat in termen van een sociale, veilige en ordelijke leeromgeving, ook in termen van de beleving van de school als belangrijk aspect van het persoonlijk welbevinden van de leerling.

- Ad 1 Met betrekking tot de vraag welke aspecten van het interpersoonlijk functioneren van belang zijn en op de basisschool tevens een goede indicator vormen van de kwaliteit van dat functioneren werd vertrokken vanuit de literatuur over sociometrische status van kinderen op school (Van Boxtel, 1992). Voor een positieve sociometrische status zijn vooral pro-sociale vaardigheden van belang, evenals de afwezigheid van sociaal relevant probleemgedrag. Ten aanzien van dit probleemgedrag valt onderscheid te maken tussen (ver)storend, agressief en pestgedrag enerzijds en overdreven sociaal teruggetrokken gedrag anderzijds. Ook speelt een positief sociaal zelfbeeld hier een belangrijke rol.
- Ad 2. Uit de Peiling Sociale Competentie (Kuhlemeier, Van Boxtel & Van Til, 2012) en een literatuurstudie met betrekking tot de kindkenmerken die van invloed zijn op het leren op school (Vijverberg, 2009), kwam naar voren dat met name een aantal aspecten van de taak- en werkhouding (zoals aandacht en concentratievermogen, (intrinsieke) motivatie voor school en leren, een positieve taakopvatting en gewetensvolle, gedisciplineerde werkhouding) een aanzienlijk deel van de variantie in leerprestaties kunnen verklaren. Ook speelt een positief zelfbeeld als leerling ('academic self concept') hierbij een belangrijke rol.
- Ad 3. Het betreft hier geen individuele leerlingkenmerken (ook al wordt de individuele leerling als informant gebruikt), maar kenmerken van de leeromgeving als sociaal, veilig en ordelijk. Ook de schoolbeleving van de leerlingen in de groep (als belangrijk aspect van het persoonlijk welbevinden van deze leerlingen) is op het niveau van de groep te interpreteren als een indicator voor de kwaliteit van de leeromgeving.

In de twee genoemde clusters van indicatoren op het niveau van de individuele leerling tekenden zich de twee domeinen af die goed bleken aan te sluiten bij de dimensies 'sociaal gedrag' en 'speelwerkhouding' zoals deze in VISEON voor kleuters werden geoperationaliseerd. Omdat we in VISEON 2.0 de gelegenheid hadden om (vanaf groep 5) de leerling ook zelf aan het woord te laten, besloten we de aspecten zelfbeeld en welbevinden die hierboven al als relevant naar voren kwamen met name in de leerlinglijst verder uit te werken. Het zijn immers vooral deze aspecten waarover men bij voorkeur de leerling zelf zou willen bevragen. Het domein 'welbevinden' (schoolbeleving, aangevuld met spreekangst en toetsangst) werd hiermee het derde domein dat aansluit bij een schaal in VISEON voor kleuters met hetzelfde label. Met het vierde domein, 'zelfbeeld' (sociaal zelfbeeld en zelfbeeld als leerling, tevens gespecificeerd naar de leerstofdomeinen rekenen en lezen), werd VISEON 2.0 gecompleteerd.

In de volgende subparagrafen gaan we in meer detail in op elk van de vier onderscheiden domeinen die betrekking hebben op sociaal-emotioneel functioneren van individuele leerlingen (paragraaf 2.5.1), het domein leeromgeving als voorwaarde voor het sociaal-emotioneel functioneren (paragraaf 2.5.2) en de wijze waarop men VISEON 2.0 kan situeren ten opzichte van een 'doorlopende leerlijn' (paragraaf 2.5.2).

2.5.1 Sociaal-emotioneel functioneren van individuele leerlingen

Zoals eerder beschreven bestaat het eerste aspect, het sociaal-emotioneel functioneren op individueel niveau, uit vier domeinen:

- *Zelfbeeld*: beleving van het eigen sociaal en cognitief functioneren
- *Welbevinden*: gedragsuitingen en ervaringen die een indicatie geven van de emotionele stabiliteit en ervaren veiligheid
- *Sociaal gedrag*: gedrag waarin leerlingen hun sociale vaardigheid (competentie) tot uitdrukking brengen
- *Werkhouding*: taak- en werkhouding in diverse situaties waarin men van de leerling verwacht dat hij/zij een bepaalde mate van concentratie en (volgehouden) aandacht opbrengt.

Deze vier domeinen bestaan uit verschillende aspecten. Een aantal aspecten wordt op leerkrachtniveau bevraagd en andere aspecten op leerlingniveau. Voor elk aspect is gekeken wie het meest geschikt is om de informatie te geven; de leerkracht of de leerling.

Het schema op pagina 24 geeft een overzicht van de domeinen wat betreft het sociaal-emotioneel functioneren van individuele leerlingen in VISEON 2.0. Naast een beschrijving van de inhoud van de domeinen, staat beschreven welke aspecten in de leerkrachtlijst zitten, welke in de leerlinglijst zitten en waarvoor de aanvullende instrumenten ingezet kunnen worden (bijvoorbeeld voor het verkrijgen van specifiekere informatie of ter vergelijking).

In aansluiting op het schema gaan we wat uitgebreider in op elk van de domeinen.

	Beschrijving	Wat brengt de leerkrachtlijst in kaart?	Wat brengt de leerlinglijst in kaart?	Aanvullende instrumenten leerlinglijst
domein zelfbeeld	Een positief en realistisch zelfbeeld is een wenselijk kenmerk van een gezonde persoonlijke en sociale ontwikkeling. Zelfbeeld hangt samen met zelfwaardering, zelfvertrouwen en zelfstandigheid. Deze eigenschappen behoren tot de persoonlijke opbrengsten van onderwijs en worden gezien als voorwaardelijk voor de ontwikkeling van sociale competentie.	-	Hoe ervaart een leerling zijn functioneren in sociaal en cognitief opzicht?	Het cognitieve zelfbeeld kan nog specifiek in kaart worden gebracht met de twee aanvullende instrumenten van dit domein: zelfbeeld lezen en zelfbeeld rekenen.
domein welbevinden	Het domein welbevinden is een weergave van hoe de leerling zich over het algemeen voelt op school. Binnen dit domein staat de beleving van de school als veilige leeromgeving centraal. Een veilige leeromgeving is voorwaardelijk voor een positief welbevinden. Wanneer een leerling zich niet prettig voelt op school, is een leerling niet goed in staat om te functioneren en te leren.	-	In hoeverre voelt een leerling zich prettig en veilig op school?	Informatie over beperkende factoren wat betreft welbevinden: sociale angst (spreekangst) en toetsangst.
domein sociaal gedrag	Een belangrijke uiting van sociaal gedrag is het pro-sociaal handelen. Pro-sociaal gedrag komt tot uiting in het samenwerken en samenspelen in en buiten de klas. De leerkracht heeft hier een goed beeld van. Binnen het hoofdinstrument wordt pro-sociaal gedrag dan ook op leerkrachtniveau bevraagd. Naast sociale competentie (pro-sociaal gedrag) besteedt VISEON 2.0 ook aandacht aan probleemgedrag. Een deel van de kinderen in het basisonderwijs heeft namelijk te lijden onder externaliserende of internaliserende problematiek die hen hindert hun sociale vaardigheden verder te ontwikkelen en in praktijk te brengen.	In hoeverre vertoont een leerling pestgedrag, storend gedrag, en/of teruggetrokken gedrag?	In hoeverre wordt een leerling gepest?	Informatie over het pro-sociaal gedrag, pestgedrag en storend gedrag zoals waargenomen door de leerling. De resultaten van deze categorieën legt u naast de resultaten op de leerkrachtschalen.
domein werkhouding	Taakgericht werken is een krachtige voorspeller van leren in het cognitieve domein. Zoals eerder genoemd bij het domein zelfbeeld, is zelfstandigheid voorwaardelijk voor de ontwikkeling van sociale competentie. Daarnaast is motivatie voor leren een belangrijke indicator van het sociaal-emotionele functioneren op school en in de klas.	In hoeverre vertoont een leerling taakgericht gedrag, motivatie voor leren en zelfstandigheid?	In hoeverre is een leerling gemotiveerd voor leren?	De motivatie kan nog specifiek in kaart worden gebracht met motivatie voor lezen en motivatie voor rekenen. Daarnaast kan informatie verzameld worden over taakgericht gedrag. De resultaten van deze categorie legt u naast de resultaten op de leerkrachtschaal.

2.5.1.1 Zelfbeeld

Een positief en realistisch zelfbeeld wordt universeel als een wenselijk kenmerk van een gezonde persoonlijke en sociale ontwikkeling beschouwd (Branden, 1994). In termen van Sameroffs (2010) ontwikkelingspsychologische invalshoeken is het zelfbeeld een van de manieren waarop het kind in toenemende mate een interne representatie opbouwt van de buitenwereld en van de ervaringen die het opdoet (het intern representatiemodel van ontwikkeling). Het zelfbeeld vormt de kern van iemands identiteit. Het hangt samen met zelfwaardering, zelfvertrouwen en zelfstandigheid.

Deze eigenschappen behoren tot de persoonlijke opbrengsten van onderwijs en worden tegenwoordig gezien als voorwaardelijk voor de ontwikkeling van sociale competentie (Ledoux, Meijer, Van der Veen & Breetveld, 2013). Onderwijs in sociale competentie is vaak sterk gericht op het bevorderen van het zelfbeeld, het zelfvertrouwen en de zelfstandigheid van de leerlingen (Ten Dam & Volman, 1999). Onderzoek heeft laten zien dat een negatief zelfbeeld aanleiding kan geven tot internaliserend en externaliserend probleemgedrag en tot lagere schoolprestaties kan leiden (Haney & Durlak, 1998; Hattie, 1992). Uit een recente peiling in groep 8 van het basisonderwijs komt naar voren dat er maar weinig leerlingen zijn met een negatief beeld van hun algemene en sociale competenties (Kuhlemeier, Van Boxtel & van Til, 2012). Een duidelijke uitzondering op deze positieve uitkomsten vormt het beeld van de eigen competentie in lezen en rekenen. De onderzoekers vonden een duidelijke tweedeling in een groep lezers en rekenaars met een negatief zelfbeeld (en een lage motivatie) en een groep met een positief zelfbeeld (en een hoge motivatie). Binnen VISEON 2.0 maken we onderscheid tussen sociaal zelfbeeld en cognitief zelfbeeld en wordt in kaart gebracht hoe het kind zijn eigen sociaal en cognitief functioneren waarneemt. Dit onderscheid sluit aan bij het werk van Harter met de CBSK (1999), Shavelson en Bolus (1982) en Marsh (1990). Onderzoek naar de samenhang tussen academisch zelfconcept en schoolprestaties laat zien dat academisch zelfconcept tussen de 16% en 36% kan verklaren van het schools presteren (Spinath & Spinath, 2005). Domein-specifieke zelfperceptie van de eigen capaciteiten (taal en rekenen) draagt volgens Steinmayr en Spinath (2009) bij aan het voorspellen van nieuwe prestaties. Het cognitieve zelfbeeld kan daarom nog specifiek in kaart worden gebracht met de twee aanvullende instrumenten van dit domein: zelfbeeld lezen en zelfbeeld rekenen.

2.5.1.2 Welbevinden

Aan de ene kant wordt een positieve beleving van de school gezien als een indicator van een gezonde persoonlijke en sociale ontwikkeling. Een kind dat lekker in zijn vel zit, zal zich over het algemeen goed voelen, ook als het op school is, en de school doorgaans als positief ervaren. Schoolbeleving zoals het in VISEON 2.0 is ondergebracht, is daarom in eerste instantie op te vatten als een kenmerk van het kind. Aan de andere kant wijst een positieve beleving van de school op een gunstig school- en klasklimaat. Het school- en klasklimaat is een onderdeel van het vormingsaanbod. Als voorwaarde voor het kunnen realiseren van cognitieve en sociale doelen wordt het verderop (paragraaf 2.4.2) besproken. De wijze waarop de kinderen – als groep – de school beleven, sluit hier echter bij aan: wanneer veel kinderen hun leeromgeving als sociaal, veilig en ordelijk ervaren, is te verwachten dat er bij veel kinderen tevens sprake is van een positieve schoolbeleving.

Sociale angst is medebepalend voor het welbevinden van leerlingen en is in conceptueel opzicht nauw verbonden met negatieve faalangst. Sociale angst kan worden omschreven als de verlamme en belemmerende angst om te falen in situaties die een intellectuele, sociale of fysieke prestatie vereisen. Uit onderzoek komt naar voren dat sociale angst het sociale functioneren ernstig kan belemmeren en vaak gepaard gaat met sociaal teruggetrokken gedrag en gepest worden. Indien onbehandeld, kan sociale angst ertoe leiden dat kinderen op school onder hun mogelijkheden presteren en minder uit hun carrière halen dan mogelijk zou zijn geweest. Twee voor basisschoolleerlingen belangrijke situaties die aanleiding geven tot sociale angst zijn de situatie waarin de leerling getoetst wordt (Zeidner, 1998) en de situatie waarin hij/zij moet spreken voor een grotere groep (Kuhlemeier, Van den Bergh & Rijlaarsdam, 2002). Beide situaties komen veelvuldig voor in het voortgezet onderwijs en de latere beroepspraktijk. Het is daarom belangrijk dat scholen aandacht schenken aan het vroegtijdig signaleren en zo nodig (laten) behandelen van toets- en spreekangst. Vandaar dat VISEON 2.0 in de

vorm van de aanvullende instrumenten ook aandacht besteedt aan faalangst in relatie tot toetsen en het spreken voor de klas als geheel en in taakgerichte groepen.

2.5.1.3 Sociaal gedrag

De Preambule bij de kerndoelen voor het basisonderwijs noemt zeven onderwijsdoelen die voor alle leergebieden van belang zijn (Greven & Letschert, 2006). Twee daarvan hebben betrekking op sociaal gedrag: respectvol luisteren en kritiseren van anderen en respectvol en verantwoordelijk omgaan met elkaar. Sociale competentie is een onderdeel van de sociale opbrengsten dat in het onderwijs en het toezicht van de Inspectie een centrale plaats inneemt. Aan sociale competentie wordt doorgaans een persoonlijke, interpersoonlijke en maatschappelijke component onderscheiden (o.a. Ten Dam & Volman, 1999; Ten Dam e.a., 2003). Met betrekking tot de interpersoonlijke component wordt het vermogen tot pro-sociaal handelen unaniem gezien als de kern van sociale competentie (o.a. Ledoux, Meijer, Van der Veen & Breetveld, 2013). Het ontwikkelen van pro-sociale vaardigheden is hiermee een wezenlijk onderdeel van de pedagogische opdracht van de school. In de dagelijkse onderwijspraktijk heeft aandacht voor sociale competentie echter niet uitsluitend een positieve achtergrond (Dijkstra, 2012). Naast sociale competentie (pro-sociaal gedrag) besteedt VISEON 2.0 ook aandacht aan probleemgedrag. Een deel van de kinderen in het basisonderwijs heeft namelijk te lijden onder externaliserende of internaliserende problematiek die hen hindert hun sociale vaardigheden verder te ontwikkelen en in praktijk te brengen. Leerkrachten besteden een niet onaanzienlijk deel van hun tijd en energie aan de begeleiding van 'lastige' of anderszins problematische leerlingen (Nijntjes & Van Diepen, 2008). Van scholen wordt verwacht dat zij een beleid hebben waarin gedragsproblematiek vroegtijdig gesignaleerd en behandeld wordt (Inspectie van het Onderwijs, 2010). Het is mede om deze reden dat VISEON 2.0 naast pro-sociaal gedrag ook aandacht schenkt aan het verstorende kind en het pestende kind. Naast de genoemde vormen van externaliserend probleemgedrag, kan ook overdreven teruggetrokken, verlegen gedrag op een gebrek aan sociale competentie duiden. Dit type gedrag komt bij kinderen veel voor (Rubin, Coplan & Bowker, 2009). Sociaal teruggetrokken kinderen eisen veel aandacht van de leerkracht op (bijvoorbeeld Coplan & Prakasch, 2003), gedragen zich sterk afhankelijk van de leerkracht (Ladd & Burgess, 1999; Rudasill e.a., 2006) en hebben een verhoogd risico op diverse sociaal-emotionele problemen, waaronder een negatief zelfbeeld, problemen met leeftijdgenoten en schoolproblemen (voor een review, zie Rubin, Coplan & Bowker, 2009). Sociaal teruggetrokken gedrag is niet per definitie problematisch; dat wordt het in de regel pas als het disfunctioneel is. Bij de disfunctionele vormen ervan kan men actief en passief teruggetrokken gedrag onderscheiden. In het geval van actieve terugtrekking is het kind veel alleen omdat andere kinderen hem of haar afwijzen, buitensluiten of pesten, vaak als een reactie op het gedrag van het teruggetrokken kind zelf (zie bijvoorbeeld Rubin e.a., 2006). In het geval van passieve terugtrekking isoleert het kind zichzelf en ligt de oorzaak veel meer in een angst om in sociale relaties te falen, een gepercipieerd gebrek aan sociale vaardigheden en relaties of een negatief zelfbeeld (Rubin & Asendorpf, 1993).

Waar we aannemen dat het aspect *gepest worden* het best bevraagd kan worden bij de leerling, veronderstellen we dat de leerkracht de beste informant is als het gaat om de verschillende aspecten binnen het domein sociaal gedrag (pro-sociaal gedrag, pestgedrag, storend gedrag en teruggetrokken gedrag). Leerlingen hebben veelal nog geen realistische kijk op het effect van hun gedrag op anderen (de groep en de leerkracht). Echter, met behulp van de aanvullende instrumenten kan de leerkracht ook de perceptie van de leerling op deze aspecten in kaart brengen (met uitzondering van teruggetrokken gedrag).

2.5.1.4 Werkhouding

De Preambule bij de kerndoelen voor het basisonderwijs noemt zeven onderwijsdoelen die voor alle leergebieden van belang zijn (Greven & Letschert, 2006). Twee daarvan hebben betrekking op taakgericht gedrag: het bevorderen van een goede werkhouding en het leren reflecteren op het eigen handelen en leren. Taakgericht werken is een krachtige voorspeller van leren in het cognitieve domein. Leerlingen die moeite hebben met het richten van hun aandacht blijken minder goede presteerders en zelfs soms onderpresteerders te zijn (verklaarde variantie tussen 8 en 18%) ten opzichte van leerlingen die in staat zijn om aandachtig, zorgvuldig en zelf-gereguleerd te werken (Barry, Lyman & Grofer

Klinger, 2002). Vandaar dat wij in VISEON 2.0 ook aandacht besteden aan de mate waarin de leerling in staat is tot aandachtig, zorgvuldig en zelf-gereguleerd werken.

Daarnaast behoort motivatie voor leren tot de persoonlijke opbrengsten van onderwijs en is het een belangrijke indicator van het sociaal-emotionele functioneren op school en in de klas (Mönks & Knoers, 2009). Een aantal aspecten van motivatie draagt volgens Steinmayr en Spinath (2009) bij aan de voorspelling van de prestaties op school. Als leerlingen plezier hebben in rekenen en taal, draagt dit bij aan goede prestaties. Motivatie is weliswaar niet de enige voorwaarde voor verbeterde leerprestaties, maar wel een noodzakelijke voorwaarde (Boekaerts & Simons, 1995). Resultaten van onderzoek naar de samenhang tussen algemene motivatie voor school en schoolprestaties zijn niet eenduidig in termen van verklaarde variantie. Wellicht heeft dit te maken met de veelal indirecte verbanden. Leerlingen die geloven in hun eigen kunnen en intrinsieke interesse hebben in de leerstof, willen meer tijd en inzet investeren in het ontwikkelen van cognitieve strategieën en metacognitieve vaardigheden. Deze investering leidt uiteindelijk tot betere leerresultaten (Boekaerts & Simons, 1995). Intrinsiek gemotiveerde leerlingen voeren taken uit omdat zij daar plezier aan beleven, zonder dat zij zich daartoe gedwongen voelen en zonder dat daar een externe beloning of een straf tegenover staat (Deci & Ryan, 1985, 1991, 2000). In het toezicht op de kwaliteit van het onderwijs is motivatie een belangrijk aandachtspunt (Inspectie van het Onderwijs, 2010). Leerkrachten besteden in de dagelijkse lespraktijk veel tijd en energie aan het bevorderen van de motivatie van hun leerlingen. Uit onderzoek blijkt dat intrinsiek gemotiveerde leerlingen zich meer voor school en leren inzetten en hogere leerprestaties behalen dan niet of extrinsiek gemotiveerde leerlingen. Tegelijkertijd laat onderzoek zien dat een aanzienlijk deel van de basisschoolleerlingen te kampen heeft met een gebrekkige motivatie voor de vakgebieden lezen en rekenen (Kuhlemeier, Van Boxtel & Van Til, 2012; Kuhlemeier, 2014). De aspecten taakgericht gedrag, zelfstandigheid en hyperactief/impulsief worden in VISEON 2.0 op leerkrachtniveau bevestigd, omdat we ook hier veronderstellen dat de leerkracht de beste informant is. Motivatie voor leren wordt op zowel leerkracht- als leerlingniveau bevestigd. Met behulp van de aanvullende instrumenten kan ook taakgericht gedrag op leerlingniveau in kaart worden gebracht. De motivatie voor leren kan nog specifiek in kaart worden gebracht met de twee aanvullende instrumenten op leerlingniveau van dit domein: motivatie voor lezen en motivatie voor rekenen. Waar algemene motivatie voor school niet eenduidig en altijd rechtstreeks samenhangt met schoolprestaties, hangt domein-specifieke motivatie sterk samen met schools presteren (Boekaerts & Simons, 1995).

2.5.2 Leeromgeving van de groep als voorwaarde voor sociaal-emotioneel functioneren

Het contextueel model zoals omschreven door Sameroff (2010) vestigt de aandacht op het belang van de sociale context voor de ontwikkeling. Het rechtvaardigt de toevoeging van (de evaluatie van) de school als leeromgeving aan VISEON 2.0. Dit wordt bevestigd op leerlingniveau. Een kind vanaf groep 5 is op een leeftijd gekomen dat het zelf uitspraken kan doen over (de kwaliteit van) die leeromgeving. In onderstaand overzicht staat een beschrijving van dit domein, inclusief de vraag waarop de score antwoord geeft.

	Beschrijving	Wat brengt de leerlinglijst in kaart?
domein leeromgeving	De school moet een plaats zijn, waar leerlingen zich veilig voelen, zodat ze kunnen en durven te leren. Het school- en klasklimaat is een onderdeel van het vormingsaanbod. Het scheppen van een sociale, veilige en ordelijke leeromgeving is immers een belangrijke randvoorwaarde voor het bereiken van sociale (sociale competentie) en cognitieve doelstellingen door de leerlingen. VISEON 2.0 beoogt daarom ook aandacht te schenken aan een sociale, veilige en ordelijke leeromgeving als voorwaarde voor leren in het sociale en cognitieve domein. Eerder is een positieve beleving van de school, de leerkracht en klasgenoten onder 'welbevinden' beschreven als een indicator van een gezonde persoonlijke en sociale ontwikkeling van het kind. Wanneer veel kinderen zo'n positieve beleving laten zien is dat op groepsniveau een indicator van een positief school- en klasklimaat.	In hoeverre ervaart een groep de omgeving als sociaal, veilig en ordelijk?

Zoals eerder gezegd is het school- en klasklimaat een onderdeel van het vormingsaanbod. Het scheppen van een sociale, veilige en ordelijke leeromgeving is immers een belangrijke randvoorwaarde voor het bereiken van sociale (sociale competentie) en cognitieve doelstellingen (o.a. Ten Dam & Volman, 1999; Goudena, 1999). Het klimaat en de veiligheid op school en in de klas spelen een centrale rol in het toezicht van de Inspectie op de kwaliteit van het onderwijs (Inspectie van het Onderwijs, 2010, 2011). VISEON 2.0 beoogt daarom ook aandacht te schenken aan een sociale, veilige en ordelijke leeromgeving als voorwaarde voor leren in het sociale en cognitieve domein. De aspecten bij het domein leeromgeving worden alleen in het groepsoverzicht getoond. De resultaten voor dit domein geven immers vooral informatie over de groep als totaal, niet over het functioneren van individuele leerlingen. Leerlingen evalueren in deze aspecten immers primair de leeromgeving, niet zichzelf.

2.5.3 Doorlopende leerlijnen

In het Cito Volgsysteem primair en speciaal onderwijs voor de groepen 3 tot en met 8, waarvan VISEON 2.0 deel uitmaakt, wordt zoveel mogelijk gewerkt met doorlopende leerlijnen als basis voor de toetsen in dit volgsysteem. In de regel gaat het om vaardigheden zoals rekenen-wiskunde, technisch lezen, spelling, begrijpend lezen en beheersing van de woordenschat. Deze vaardigheden laten over de leerjaren heen een stelselmatige groei zien, waardoor de leervorderingen over een groot deel van de basisschoolperiode voor een beperkt aantal leerlijnen in de vorm van een score op een vaardigheidsschaal te volgen zijn. Bij de constructie van toetsen voor deze leerlijnen worden einddoelen (kerndoelen) en tussendoelen gehanteerd die houvast bieden bij het operationaliseren van de vaardigheidsgroei.

VISEON 2.0 maakt deel uit van een reeks instrumenten die tezamen tot doel hebben om het sociaal-emotioneel functioneren van kinderen van zeer verschillende leeftijd in kaart te kunnen brengen aan de hand van zoveel mogelijk dezelfde domeinen en aspecten. De volgende vier instrumenten met hun naar leeftijd verschillende doelgroepen zijn te onderscheiden:

Instrument	Doelgroep
VISEON 2.0	Leerlingen in groep 3 tot en met groep 8 (<i>leerkrachtlijst</i>) en leerlingen in groep 5 tot en met groep 8 (<i>leerlinglijst</i>) van het primair onderwijs
VISEON voor kleuters	Leerlingen in groep 1 en 2 van het primair onderwijs
Observatielijst (voor) peuters	Kinderen van 2 en 3 jaar oud op kinderdagverblijven of peuterspeelzalen
Observatielijst (voor) baby's en dreumesen	Kinderen van 6 maanden tot 2 jaar op kinderdagverblijven

Omdat VISEON voor kleuters en VISEON 2.0 deel uitmaken van hetzelfde volgsysteem en qua doelgroep op elkaar aansluiten, is vooral hun onderlinge afstemming van belang.

In onderstaand overzicht staat op welke wijze de dimensies uit VISEON voor kleuters overlappen met de domeinen uit VISEON 2.0:

	VISEON voor kleuters	VISEON 2.0
domein zelfbeeld		x
domein welbevinden	x	x
domein sociaal gedrag	x	x
domein werkhouding	x	x
domein leeromgeving		x

Drie domeinen sluiten aan bij de dimensies die in VISEON voor kleuters konden worden benoemd als 'welbevinden', 'sociaal gedrag' en 'werkhouding'. Het verschil is dat de domeinen in VISEON 2.0 ieder konden worden uitgewerkt in verschillende relevante aspecten (zie de overzichten eerder in dit hoofdstuk). Daarnaast kon het aantal domeinen voor de groepen 5 tot en met 8 worden uitgebreid met de domeinen 'zelfbeeld' en 'leeromgeving'. Dit werd mogelijk door de introductie van de leerlinglijst voor deze groepen. Beide domeinen zijn in termen van zelfrepresentatie en sociale context belangrijke ontwikkelingspsychologische aspecten. Ze konden worden toegevoegd omdat leerlingen vanaf groep 5 als belangrijke informant over zichzelf en hun eigen leeromgeving gelden. Dat laatste geldt ook voor de meeste aspecten uit 'welbevinden', 'sociaal gedrag' en 'werkhouding'; voor deze aspecten is het mogelijk de observatie van de leerkracht en de beoordeling door de leerling zelf naast elkaar te leggen.

Het is belangrijk om te onderkennen dat toenemende differentiatie in ontwikkeling en functioneren (en daarmee ook het aantal aspecten in het instrument) samengaat met het toenemen van de leeftijd. Zo onderscheiden we in de observatielijst voor de allerjongste leeftijdsgroep ('Baby's en dreumesen') maar één schaal ('sociaal-emotioneel functioneren'), bij de peuters en kleuters drie schalen waarin dat sociaal-emotioneel functioneren tot uitdrukking komt (de lijsten voor peuters en kleuters kennen dezelfde drie aspecten) en is in VISEON 2.0 sprake van een nog grotere differentiatie in schalen. Er is dus sprake van dezelfde 'doorlopende leerlijn(en)', maar toenemende gedragsdifferentiatie maakt een meer gedifferentieerde meting mogelijk en noodzakelijk. Kortom, we hebben getracht een reeks instrumenten te construeren waarmee het sociaal-emotioneel functioneren van kinderen van zeer verschillende leeftijden in kaart kan worden gebracht aan de hand van zoveel mogelijk dezelfde, zij het steeds meer gedifferentieerde dimensies. Dit wil echter nog niet zeggen dat deze instrumenten geschikt zijn om de individuele ontwikkeling of vaardigheidsgroei rechtstreeks af te beelden. Wél is het bij elke observatielijst mogelijk de (zeer) afwijkende scores van een kind op de onderscheiden schalen bij elke afname op te vatten als signaal: dit functioneren is dermate zwak dat ingrijpen op zijn plaats is.

3 Constructie en beschrijving van de vragenlijsten

Zoals aangegeven in het vorige hoofdstuk beoogt VISEON 2.0 in totaal 26 aspecten van het sociaal-emotionele functioneren van leerlingen te meten. Daarvan worden er acht bevraagd in het basisgedeelte van de leerkrachtvragenlijst (LK basis), acht in het basisgedeelte van de leerlinglijst (LL basis) en tien in het aanvullend gedeelte van de leerlinglijst (LL aanvullend). In dit hoofdstuk beschrijven we eerst de manier waarop de lijsten voor leerkrachten en leerlingen ontwikkeld zijn (paragraaf 3.1). Vervolgens bespreken we de opbouw, afname en scoring van de leerkracht- en leerlinglijsten (paragraaf 3.2). Tot slot presenteren het instrument in zijn uiteindelijke samenstelling, samen met enkele beschrijvende statistische gegevens per schaal en per uitspraak (paragraaf 3.3). Het normeringsonderzoek is het onderwerp van hoofdstuk 4.

3.1 Constructie van de leerkracht- en leerlinglijst

In eerste instantie is een omvangrijke set uitspraken over het sociaal-emotioneel functioneren van leerlingen samengesteld. Niet al deze uitspraken zijn speciaal voor VISEON 2.0 ontwikkeld. Een aanzienlijk deel is onverkort overgenomen uit de Peiling Sociale Competentie (Kuhlemeier, Van Boxtel & Van Til, 2012) en een eerdere versie van VISEON (Kamphuis e.a., 2004). Daarnaast zijn geheel nieuwe uitspraken geformuleerd waarvan nog geen psychometrische eigenschappen bekend waren. Deze nieuwe uitspraken zijn geformuleerd door drie onderzoekers: twee psychologen en een onderwijskundige.

Bij de ontwikkeling van VISEON 2.0 zijn we niet over een nacht ijs gegaan. Ten behoeve van de constructie van de definitieve schalen zijn twee proefafnames en een normeringsonderzoek uitgevoerd. In de beide proefafnames is het concept-instrumentarium onderzocht op onder meer begrijpelijkheid, hanteerbaarheid (o.a. afnametijd), betrouwbaarheid en validiteit. In het vervolg van deze paragraaf beschrijven we de uitgangspunten voor de keuze van uitspraken (paragraaf 3.1.1) en doen we verslag van de beide proefafnames (paragraaf 3.1.2).

3.1.1 Uitgangspunten voor de keuze van uitspraken

Aan de selectie van uitspraken voor VISEON 2.0 lagen de volgende uitgangspunten ten grondslag:

- **Begrijpelijkheid voor de doelgroep**
De uitspraken dienen voor leerkracht en leerling begrijpelijk en niet kwetsend te zijn en door hen geïnterpreteerd te worden zoals bedoeld door de ontwikkelaar. Daartoe zijn de uitspraken in kleinschalige settings met leerkrachten en leerlingen besproken en waar nodig op basis van de resultaten geherformuleerd dan wel uit de pool van uitspraken verwijderd.
- **Afnametijd van de lijst als geheel**
De omvang van de vragenlijsten voor de leerkracht en de leerlingen mag niet te groot zijn. De drie onderdelen van het instrumentarium – LK basis, LL basis en LL aanvullend – mogen leerkracht en leerling niet al te zeer belasten en moeten dus snel in te vullen zijn. Op basis van behoeften- en gebruikersonderzoek bij de leerkrachten die aan de proefafnames deelnamen, zou de uiteindelijke leerlinglijst tachtig uitspraken mogen bevatten (range 50 – 116). Bij een vastgestelde gemiddelde afnametijd van vier uitspraken per minuut zou dat mogelijkheden bieden voor in totaal twintig minuten afnametijd. Rekening houdend met de range werd de wenselijke invultijd voor een afname bij leerlingen bepaald op ongeveer vijftien minuten met ruimte voor ongeveer zestig uitspraken (per onderdeel). Op soortgelijke wijze hebben de leerkrachten aangegeven dat de leerkrachtlijst uit 61 vragen zou mogen bestaan (range 50-91). Bij een vastgestelde gemiddelde invultijd van zes vragen per minuut komen we dan voor de leerkrachtlijst uit op ongeveer zestig vragen met een invultijd van ongeveer tien minuten per leerling.

- Aantal uitspraken per schaal
Het aantal uitspraken per schaal mag niet groter zijn dan nodig. Op basis van eerder onderzoek (Kuhlemeier, Van Boxtel & Van Til, 2012) werd ingeschat dat vijf à acht uitspraken per schaal zouden volstaan om een goede betrouwbaarheid te garanderen.
- Homogeniteit
Bij de formulering en selectie van uitspraken is gestreefd naar homogene schalen zonder afbreuk te doen aan de inhoudelijke diversiteit. Het doel was uni-dimensionele schalen te ontwikkelen met een hoog percentage verklaarde variantie. Daartoe zijn uitspraken met een relatief lage item-restcorrelatie verwijderd (tenzij er sterke inhoudelijke argumenten waren om de desbetreffende uitspraak toch te handhaven, wat overigens zelden voorkwam). Als extra check zijn de uitspraken per schaal aan een exploratieve factoranalyse onderworpen waarbij uitspraken die op een tweede factor laadden in beginsel verwijderd werden.
- Factorzuiverheid van de meting van sociaal-emotioneel functioneren
De aspecten van het sociaal-emotioneel functioneren moeten factor-zuiver gemeten worden. Daartoe zijn exploratieve factoranalyses (principale componentenanalyse met varimaxrotatie) uitgevoerd op de uitspraken van meerdere schalen tegelijkertijd. Uitspraken die hoger op een andere factor laadden dan op de 'eigen' factor, werden daarbij verwijderd. Dit was slechts zelden nodig. Daarnaast zijn de correlaties tussen de aldus samengestelde schalen berekend waarbij (te) hoog correlerende schalen nogmaals kritisch bekeken werden op ongewenste inhoudelijke overlap.
- Inhoudelijke overwegingen
De selectie van uitspraken dient plaats te vinden op basis van psychometrische en inhoudelijke overwegingen. Soms bleken op basis van de item-restcorrelaties meerdere uitspraken in aanmerking te komen voor selectie, maar bleek de inhoudelijke overlap met een of meer andere uitspraken groot. In dit soort gevallen werden 'look-a-likes' vermeden.
Tegelijkertijd is de dekking van elk aspect van het sociaal-emotioneel functioneren breed gehouden. Soms werd een uitspraak met wat minder uitmuntende empirische eigenschappen toch gekozen omdat een bepaald aspect in onze ogen niet in de schaal mocht ontbreken. Een andere inhoudelijke overweging betrof de mate waarin het taalgebruik als passend voor de doelgroep werd beoordeeld.

3.1.2 De proefafnames

Aan het normeringsonderzoek gingen twee proefafnames vooraf. De eerste proefafname is gehouden in het voorjaar van 2014. Leerkrachten vulden een vragenlijst in voor 211 leerlingen en 315 leerlingen vulden een leerlinglijst in. Behalve psychometrische gegevens zijn in de eerste proefafname ook kwalitatieve gegevens verzameld. Zo is leerkrachten gevraagd naar hun mening over onder meer de ervaren bruikbaarheid van de schalen voor de onderwijspraktijk en de formulering van de uitspraken. Ook is het volledige instrumentarium aan enkele ter zake deskundigen ter becommentariëring voorgelegd, waaronder een methodeauteur en enkele praktijkdeskundigen. Op basis van onder meer itemkarakteristieken, betrouwbaarheidsanalyse, factoranalyse en correlatieanalyse zijn uit de omvangrijke set uitspraken 24 concept-schalen samengesteld. Daarnaast bleek het nodig om voor Sociaal zelfbeeld en Cognitief zelfbeeld twee geheel nieuwe schalen te ontwikkelen. De aldus ontwikkelde 26 schalen vormden de input voor de tweede proefafname.

In de tweede proefafname zijn de 26 concept-schalen nogmaals aan leerkrachten en leerlingen voorgelegd. De tweede proefafname is gehouden in het najaar van 2014. Leerkrachten vulden de leerkrachtvragenlijst in voor 437 leerlingen terwijl 501 leerlingen de leerlinglijst invulden. De tweede proefafname diende als check op de schalen zoals deze op basis van de eerste proefafname waren samengesteld. Daarnaast had de tweede proefafname tot doel de definitieve schalen voor het normeringsonderzoek samen te stellen. De uitkomsten van de tweede proefafname gaven aanleiding tot enkele kleine wijzigingen. Zo werden uit de leerkrachtschalen Pestgedrag en Storend gedrag enkele minder goed functionerende uitspraken verwijderd terwijl aan de leerlingschaal Cognitief zelfbeeld twee uitspraken werden toegevoegd.

3.2 Opbouw, afname en scoring van de leerkracht- en leerlinglijst

Naar aanleiding van het normeringsonderzoek zijn nogmaals enkele uitspraken verwijderd terwijl er geen uitspraken geherformuleerd of toegevoegd hoefden te worden (zie ook hoofdstuk 4). In deze paragraaf beschrijven we de opbouw, afname en scoring van het uiteindelijk samengestelde instrument.

3.2.1 Opbouw

De uiteindelijk uitgegeven versie van VISEON 2.0 bevat in totaal 174 uitspraken verdeeld over 26 schalen. Daarvan behoren er acht tot het basisgedeelte van de leerkrachtlijst (LK basis), acht tot het basisgedeelte van de leerlinglijst (LL basis) en tien tot het aanvullende gedeelte van de leerlinglijst (LL aanvullend). LK basis bevat 55 uitspraken, LL basis eveneens 55 uitspraken en LL aanvullend 64 uitspraken. De uiteindelijk gekozen uitspraken van LK basis, LL basis en LL aanvullend zijn onverkort weergegeven in respectievelijk tabel 3.3, 3.4 en 3.5. Vanwege de omvang zijn deze tabellen aan het einde van dit hoofdstuk geplaatst.

Op basis van theoretische overwegingen (zie hoofdstuk 2) zijn de 26 schalen van VISEON 2.0 ingedeeld in vijf domeinen van sociaal-emotioneel functioneren. In de handleiding (Cito, 2015) zijn deze domeinen als volgt omschreven:

- *Sociaal gedrag*: gedrag waarin leerlingen hun sociale vaardigheid (competentie) tot uitdrukking brengen;
- *Werkhouding*: taak- en werkhouding in diverse situaties waarin men van de leerling verwacht dat het een bepaalde mate van concentratie en (volgehouden) aandacht opbrengt;
- *Zelfbeeld*: beleving van het eigen sociaal en cognitief functioneren;
- *Welbevinden*: gedragsuitingen en ervaringen die een indicatie geven in hoeverre de leerling zich prettig voelt op school;
- *Leeromgeving als voorwaarde voor sociaal-emotioneel functioneren*: de mate waarin leerlingen hun leeromgeving als sociaal, veilig en ordelijk ervaren.

Tabel 3.1 laat zien hoe de 174 uitspraken zijn verdeeld over de drie onderdelen van VISEON 2.0, de vijf domeinen en de 26 aspecten van sociaal-emotioneel functioneren. Voor een verantwoording van de indeling van de schalen in de vijf domeinen wordt verwezen naar hoofdstuk 2. In de loop van de tijd zal de aanvullende leerlinglijst naar verwachting nog uitgebreid worden met nieuwe schalen.

Tabel 3.1 Verdeling van de uitspraken van VISEON 2.0 over de 26 schalen van LK basis, LL basis en LL aanvullend en de vijf domeinen van sociaal-emotioneel functioneren

Schaal	Domein				
	Sociaal gedrag	Werkhouding	Zelfbeeld	Welbevinden	Leeromgeving
LK basis					
pro-sociaal gedrag	8				
pestgedrag	6				
storend gedrag	6				
teruggetrokken gedrag	7				
taakgericht gedrag		7			
motivatie voor leren		7			
zelfstandigheid		7			
hyperactief/impulsief		7			
LL basis					
sociaal zelfbeeld			6		
cognitief zelfbeeld			7		
schoolbeleving				7	
gepest worden	7				
motivatie voor leren		7			
sociale leeromgeving					7
onveilige leeromgeving					7
onordelijke leeromgeving					7
LL aanvullend					
pro-sociaal gedrag	7				
pestgedrag	7				
storend gedrag	7				
taakgericht gedrag		7			
motivatie voor lezen		5			
motivatie voor rekenen		5			
zelfbeeld lezen			5		
zelfbeeld rekenen			5		
spreekangst				8	
toetsangst				8	

3.2.2 Afname

VISEON 2.0 is een volledig computergestuurd instrument. Om het te kunnen gebruiken dienen het Computerprogramma LOVS en VISEON 2.0 geïnstalleerd te zijn (zie gebruikershandleiding). De lijsten voor leerkracht en leerlingen zijn uitsluitend digitaal af te nemen. Via het computerprogramma kan de leerkracht afnames plannen, vragenlijsten invullen en rapportages opvragen.

Met behulp van de module 'toetsen plannen' kan de leerkracht in het Computerprogramma LOVS de afname van de vragenlijsten inplannen. Uitgebreide informatie over het plannen van de vragenlijsten staat in de handleiding van het Computerprogramma LOVS (Cito, z.j.).

Bij het plannen van de afname kiest de leerkracht of hij/zij LK basis en/of LL basis wil afnemen. Deze twee onderdelen van VISEON 2.0 worden integraal afgenomen. De leerkracht kiest zelf voor welke leerlingen hij/zij de leerkrachtlijst invult en geeft aan welke leerlingen LL basis maken.

In de handleiding wordt de leerkracht geadviseerd aanvullende informatie te verzamelen als de resultaten van LK basis of LL basis daartoe aanleiding geven. Bij LL aanvullend kan de leerkracht daarom kiezen voor

afname van één of meer van de modules Sociaal gedrag, Werkhouding, Zelfbeeld en Welbevinden. Voor een bespreking van de uiteenlopende redenen om aanvullend onderzoek te doen verwijzen wij naar de handleiding (Cito, 2015).

WISEON 2.0 kan in beginsel op elk willekeurig moment worden afgenomen. In de handleiding wordt de leerkracht geadviseerd de lijsten in november af te nemen en zo nodig in april een tweede afname te plannen. De leerkracht wordt aangeraden de lijst alleen in te vullen voor leerlingen die langer dan twee maanden in de groep zitten. Dan heeft de leerkracht voldoende tijd gehad om zich een beeld te vormen van de leerling en heeft de leerling genoeg tijd gehad om te wennen aan zijn nieuwe omgeving.

Het invullen van LK basis kost de leerkracht in de regel niet meer dan tien minuten per leerling. Het invullen van LL basis vergt eveneens ongeveer tien minuten. De tijd die nodig is voor het invullen van LL aanvullend neemt afhankelijk van het aantal domeinen dat de leerkracht selecteert drie tot vijftien minuten in beslag. Hiermee blijven we ruimschoots binnen de grenzen die leerkrachten in de proefafnames hebben aangegeven (zie paragraaf 3.1.1).

3.2.3 Scoring

De scoring van het instrument is volledig geautomatiseerd. Per uitspraak worden de waarden 1, 2, 3 en 4 toegekend aan de gekozen antwoordmogelijkheden. Er zijn afhankelijk van de aard van de gemeten eigenschap twee schaaltypen in WISEON 2.0 vertegenwoordigd: “klopt helemaal niet – klopt niet – klopt – klopt helemaal” en “nooit – een enkele les – meerdere lessen – (bijna) alle lessen”. Bij de zogeheten ‘positieve schalen’ zoals Pro-sociaal gedrag staat 1 voor functioneren dat als relatief ongunstig wordt beoordeeld en 4 voor functioneren dat als gunstig wordt gezien. Bij de zogeheten ‘negatieve’ schalen zoals Pestgedrag is de betekenis van de antwoordcategorieën precies omgekeerd. Hier indiceert een hogere score een minder gunstig niveau van sociaal-emotioneel functioneren. De antwoorden van de leerkracht of leerling worden per schaal opgeteld tot ruwe scores, die op hun beurt weer worden omgezet in percentielscores en niveauscores, in WISEON 2.0 signalen genoemd. Daarbij merken we op dat de gerapporteerde percentiel- en niveauscores zo berekend zijn dat een hogere score bij alle schalen een minder gunstige uitslag indiceert (zie hoofdstuk 4 over de normering).

3.3 Beschrijvende inhoudelijke en statistische gegevens

In deze paragraaf presenteren we de schalen en uitspraken van het instrument in zijn uiteindelijke samenstelling. Daarnaast bespreken we zowel de uitspraken als de schalen in termen van enkele statistische descriptoren. Uitspraken en descriptoren zijn per schaal weergegeven in de tabellen 3.3, 3.4 en 3.5 voor respectievelijk LK basis, LL basis en LL aanvullend. Vanwege de omvang zijn deze tabellen opgenomen aan het einde van dit hoofdstuk. Achtereenvolgens zijn weergegeven het aantal leerlingen (N), de item-restcorrelatie (Rir), de gemiddelde score (Gem), de standaarddeviatie (Stddev) en de procentuele verdeling van de antwoorden over de vier antwoordmogelijkheden. De Rir geeft een realistisch beeld van de samenhang tussen de itemscore en de somscore op de schaal waartoe het item behoort omdat de itemscore zelf niet in de somscore is meegenomen. Daardoor is de Rir in de praktijk lager dan de Rit. Voor Rir bestaan echter geen normen, waar dat voor Rit wel het geval is (vergelijk het COTAN Beoordelingssysteem (Evers, Lucassen, Meijer & Sijtsma, 2010).

De item-restcorrelaties kunnen we als volgt samenvatten:

- Voor LK basis (55 uitspraken) varieert de Rir van 0,57 tot 0,87 met een gemiddelde van 0,74;
- Voor LL basis (55 uitspraken) varieert de Rir van 0,43 tot 0,70 met een gemiddelde van 0,58;
- Voor LK aanvullend (64 uitspraken) varieert de Rir van 0,24 tot 0,89 met een gemiddelde van 0,64.

Passen we de COTAN-normen voor de Rit toe op onze (lagere) Rir's, dan heeft slechts één uitspraak een Rir die lager is dan de ondergrenswaarde die de COTAN geeft om een uitspraak als goed te beoordelen (waarbij

we opmerken dat het oordeel over de kwaliteit van die ene dissonante uitspraak nog steeds 'voldoende' is). De zeer hoge item-restcorrelaties zijn deels te danken aan het gegeven dat het hier niet gaat om dichotome goed-foutscores (zoals bij toetsitems gebruikelijk is), maar om vier-puntschalen. Desalniettemin kunnen we concluderen dat er sprake is van in psychometrisch opzicht goede uitspraken zonder dat de uitspraken blijken te geven van storende overlap (zie verder ook hoofdstuk 5 over de betrouwbaarheid). Dat er geen sprake is van storende overlap blijkt uit de uitspraken zelf.

In de tabellen 3.3, 3.4 en 3.5 is ook de procentuele verdeling van de antwoorden over de vier antwoordmogelijkheden gepresenteerd. De verdeling is – in overeenstemming met onze verwachtingen – voor sommige uitspraken scheef. Dit geldt met name voor de uitspraken van enkele 'negatieve' schalen waarbij een hogere score een minder gunstig niveau van sociaal-emotioneel functioneren indiceert. Vooral daar worden de beide positieve antwoordmogelijkheden duidelijk vaker gekozen dan de beide negatieve antwoordmogelijkheden.

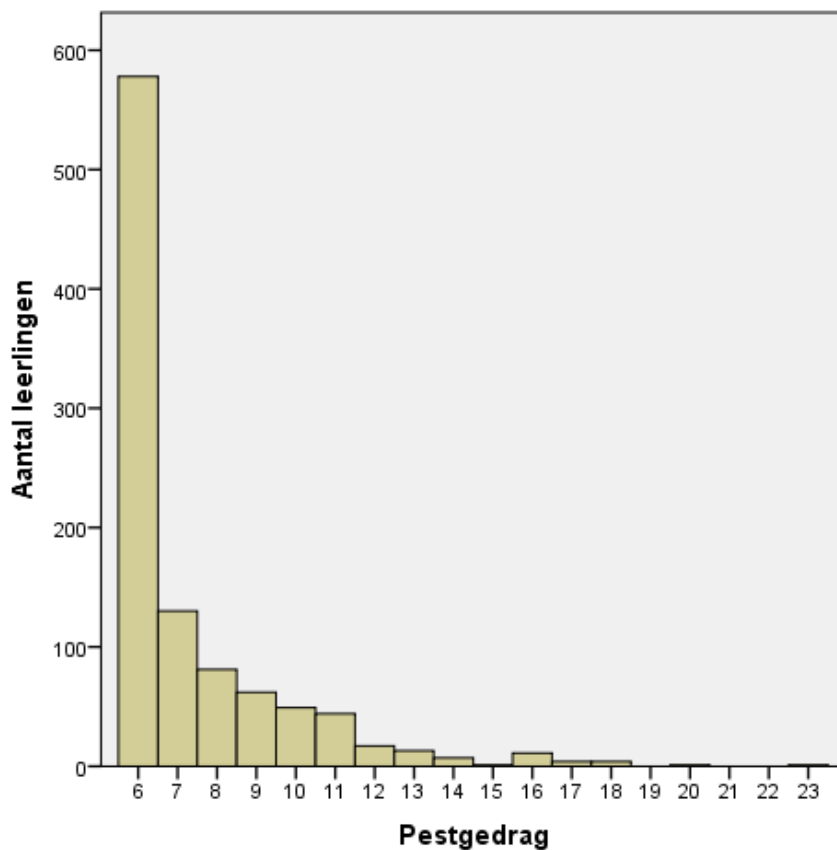
De beschrijvende statistische gegevens per schaal zijn weergegeven in tabel 3.2. Achtereenvolgens zijn weergegeven het aantal leerlingen (N), de laagst waargenomen somscore (Min), de hoogst waargenomen somscore (Max), de gemiddelde somscore (Gem), de standaarddeviatie (Stddev), de standaardmeetfout (Sem), de scheefheid en de kurtosis (gepiektheid). Voor de bespreking van de standaardmeetfouten wordt verwezen naar paragraaf 5.1 uit het hoofdstuk over de betrouwbaarheid en meetnauwkeurigheid.

Tabel 3.2 Enkele beschrijvende statistische gegevens per schaal

	N	Min	Max	Gem	Stddev	Sem	Scheefheid	Kurtosis
LK basis								
pro-sociaal gedrag	1001	8	32	24,09	4,46	1,18	-0,12	0,20
pestgedrag	1003	6	23	7,38	2,32	0,80	2,35	6,67
storend gedrag	1004	6	24	9,91	3,58	1,29	1,09	1,10
teruggetrokken gedrag	1005	7	27	13,06	4,07	1,35	0,26	-0,36
taakgericht gedrag	1004	7	28	19,99	4,47	1,26	-0,23	-0,06
motivatie voor leren	1001	7	28	21,63	4,17	1,18	-0,43	0,17
zelfstandigheid	1002	7	28	19,39	4,21	1,19	0,17	-0,30
hyperactief/impulsief	654	7	28	12,05	4,74	1,18	0,69	-0,11
LL basis								
sociaal zelfbeeld	1883	7	24	20,26	2,78	1,21	-0,92	1,48
cognitief zelfbeeld	1862	7	28	23,01	3,06	1,30	-0,56	0,69
schoolbeleving	1868	7	28	24,39	3,39	1,36	-1,45	3,22
gepest worden	1882	7	27	10,27	3,15	1,34	1,43	2,72
motivatie voor leren	1879	7	28	21,94	3,35	1,34	-0,27	0,06
sociale leeromgeving	1877	9	28	22,05	3,25	1,30	-0,11	-0,03
onveilige leeromgeving	1888	7	28	12,46	3,72	1,49	1,01	1,20
onordelijke leeromgeving	1885	7	28	13,80	3,68	1,60	0,85	0,82
LL aanvullend								
pro-sociaal gedrag	577	7	28	21,42	3,21	1,44	-0,28	0,92
pestgedrag	569	7	26	9,15	2,37	1,01	1,77	5,41
storend gedrag	574	7	28	10,86	3,42	1,49	1,17	2,32
taakgericht gedrag	577	10	28	21,93	3,31	1,44	-0,30	0,39
motivatie voor lezen	587	5	20	14,24	4,14	1,37	-0,37	-0,68
motivatie voor rekenen	592	5	20	13,18	4,37	1,16	-0,23	-0,79
zelfbeeld lezen	587	5	20	15,22	3,24	1,41	-0,59	0,17
zelfbeeld rekenen	590	5	20	14,40	3,67	1,27	-0,34	-0,39
spreekangst	577	8	32	15,63	5,36	1,93	0,52	-0,07
toetsangst	576	8	32	14,48	6,24	1,76	1,19	0,57

Zoals we op basis van de procentuele verdelingen van de antwoorden over de antwoordmogelijkheden al konden vermoeden (zie tabel 3.1, 3.2 en 3.3) is de verdeling bij enkele schalen uitzonderlijke scheef. Ter illustratie toont figuur 3.1 de extreem scheve en gepiekte frequentieverdeling voor de schaal Pestgedrag uit LK basis. In overeenstemming met eerder onderzoek (o.a. Kuhlemeier, Van Boxtel & Van Til, 2012) maakt de meerderheid van de leerlingen zich volgens hun leerkracht nooit schuldig aan de bevroegde vormen van pestgedrag. Voor deze leerlingen hebben de leerkrachten bij alle zes uitspraken voor de antwoordmogelijkheid 'Nooit' gekozen. Daarentegen zijn er maar zeer weinig leerlingen die volgens hun leerkracht vaak pestgedrag vertonen. De extreme waarde van de kurtosis lijkt vooral toe te schrijven aan het naar verhouding wat grotere aantal leerlingen met een somscore van zestien (zie figuur 3.1) en hoeft ons dus niet al te zeer te verontrusten.

Figuur 3.1 Frequentieverdeling voor de schaal Pestgedrag uit LK basis



Bij de 'negatieve' schalen Pestgedrag, Storend gedrag en Gepest worden is de scheefheid van de verdeling onlosmakelijk verbonden met het gegeven dat deze relatief ernstige vormen van probleemgedrag in het regulier basisonderwijs (gelukkig) weinig voorkomen. Uiteraard hadden wij een gelijkmatigere verdeling kunnen realiseren door de uitspraken wat gematigder te formuleren. Zo had vervanging van de uitspraak 'Treitert of pest andere kinderen' uit de schaal Pestgedrag door 'Plaagt andere kinderen' zonder twijfel tot een minder scheve verdeling geleid. Op grond van inhoudelijke overwegingen hebben wij hier echter van afgezien.

Tabel 3.3 De uitspraken en enkele beschrijvende statistische gegevens van LK basis

	N	Rir	Gem	Stddev	Verdeling over de antwoordmogelijkheden			
					1	2	3	4
Pro-sociaal gedrag								
kan goed dingen delen	1006	0,69	3,19	0,63	0%	11%	58%	31%
bemiddelt bij een ruzie tussen andere kinderen	1007	0,72	2,82	0,74	3%	28%	52%	17%
kan goed met andere kinderen samenwerken in groepjes	1006	0,78	3,06	0,70	2%	15%	57%	26%
kan andere kinderen goed aanvoelen	1007	0,80	2,89	0,73	3%	23%	55%	19%
troost andere kinderen als zij pijn of verdriet hebben	1006	0,73	3,07	0,68	1%	15%	58%	25%
kan ruzies met andere kinderen zelf oplossen	1004	0,73	2,84	0,71	3%	25%	56%	16%
kan zich aanpassen aan andere kinderen	1006	0,79	2,99	0,69	2%	17%	60%	21%
helpt andere kinderen	1006	0,77	3,22	0,59	1%	7%	62%	30%
Pestgedrag								
is onredelijk boos of driftig	1006	0,68	1,33	0,58	72%	23%	5%	0%
maakt ruzie zonder aanleiding	1007	0,78	1,26	0,52	77%	19%	3%	0%
treitert of pest andere kinderen	1005	0,74	1,29	0,56	76%	19%	5%	0%
slaait of schopt andere kinderen	1007	0,68	1,15	0,42	87%	11%	2%	0%
scheldt andere kinderen uit	1007	0,77	1,29	0,54	75%	21%	3%	0%
bedreigt een ander kind	1006	0,61	1,06	0,26	95%	4%	1%	0%
Storend gedrag								
stoort de les	1006	0,76	1,65	0,79	52%	34%	11%	3%
houdt zich aan de gedragsregels	1007	0,61	1,47	0,77	66%	23%	7%	3%
doet wat de leerkracht zegt	1007	0,59	1,41	0,67	67%	27%	4%	2%
houdt andere kinderen van hun werk	1006	0,72	1,85	0,77	34%	50%	12%	4%
bemoeit zich ongevraagd met het werk van andere kinderen	1006	0,62	1,66	0,77	49%	38%	10%	3%
kan tijdens de les zijn/haar mond niet houden	1007	0,68	1,85	0,86	39%	42%	13%	6%
Teruggetrokken gedrag								
heeft veel emotionele steun nodig van de leerkracht	1007	0,74	2,05	0,84	29%	42%	25%	4%
gedraagt zich te afhankelijk van de leerkracht	1006	0,68	1,78	0,71	36%	50%	12%	2%
heeft uitzonderlijk weinig zelfvertrouwen	1007	0,77	1,92	0,73	28%	54%	15%	3%
is snel in tranen	1007	0,64	1,74	0,74	42%	43%	13%	2%
maakt een bijzonder onzekere indruk	1006	0,77	1,90	0,75	31%	51%	16%	3%
is bang om contacten met andere kinderen aan te gaan	1007	0,57	1,64	0,66	45%	48%	6%	1%
maakt zich snel zorgen	1007	0,68	2,03	0,78	26%	49%	22%	3%
Taakgericht gedrag								
checkt bij een taak of hij/zij alles goed gedaan heeft	1006	0,79	2,72	0,78	6%	31%	49%	14%
controleert zelf of de taak volledig is uitgevoerd	1006	0,81	2,79	0,79	6%	26%	51%	17%
herstelt uit zichzelf fouten tijdens het uitvoeren van een taak	1006	0,75	2,81	0,73	4%	25%	56%	15%
werkt zorgvuldig	1006	0,81	2,85	0,78	5%	25%	50%	19%
kan de aandacht goed bij spel of werk houden	1006	0,74	2,94	0,78	4%	21%	52%	23%
let goed op in de groep	1006	0,71	3,02	0,74	3%	16%	56%	25%
werkt netjes	1004	0,68	2,88	0,81	6%	23%	50%	22%
Motivatie voor leren								
leest met plezier	1006	0,60	3,07	0,76	3%	17%	50%	30%
toont weinig interesse voor school	1007	0,72	3,26	0,73	2%	12%	46%	41%
rekent met plezier	1005	0,64	3,05	0,71	2%	17%	55%	26%
moet tot leren worden aangezet	1007	0,76	3,00	0,85	5%	22%	42%	31%
heeft meestal zin om taken uit te voeren	1004	0,83	3,08	0,68	2%	14%	59%	25%
is gemotiveerd om een taak af te ronden	1007	0,80	3,08	0,70	2%	15%	57%	27%
leert met plezier	1007	0,86	3,10	0,69	1%	15%	56%	27%
Zelfstandigheid								
weet precies wat hij/zij wil	1007	0,64	3,07	0,68	1%	16%	57%	25%
kan goed voor zichzelf opkomen	1006	0,68	2,99	0,70	1%	21%	55%	23%
neemt het initiatief	1007	0,80	2,82	0,74	3%	29%	51%	17%
komt zelfverzekerd over	1006	0,79	2,78	0,71	2%	31%	52%	15%
toont veel zelfvertrouwen	1005	0,78	2,74	0,73	3%	36%	47%	15%
neemt bij groepswerk anderen op sleeptouw	1007	0,76	2,52	0,80	7%	45%	36%	12%
voert de regie bij groepswerk	1006	0,75	2,46	0,80	9%	47%	34%	11%
Hyperactief/impulsief								
kan tijdens de les geen moment stilzitten	657	0,87	1,71	0,80	47%	38%	11%	4%
is voortdurend in beweging	657	0,87	1,70	0,81	48%	37%	10%	4%
is rusteloos en overactief	657	0,86	1,61	0,76	53%	36%	8%	3%
praat en kletst aan één stuk door, van de hak op de tak	656	0,80	1,69	0,74	45%	41%	11%	2%
handelt impulsief	656	0,79	1,86	0,81	38%	40%	19%	3%
kan niet wachten tot hij/zij aan de beurt is	657	0,76	1,76	0,78	43%	40%	15%	2%
geeft al antwoord voordat de leerkracht de vraag heeft afgemaakt	656	0,77	1,71	0,76	45%	41%	11%	2%

Tabel 3.4 De uitspraken en enkele beschrijvende statistische gegevens van LL basis

	N	Rir	Gem	Stddev	Verdeling over de antwoordmogelijkheden			
					1	2	3	4
Sociaal zelfbeeld								
Ik ben tevreden over het aantal vrienden/vriendinnen dat ik heb	1944	0,54	3,61	0,60	1%	3%	30%	66%
Ik ben tevreden over hoe ik met andere kinderen kan opschieten	1935	0,54	3,33	0,60	1%	4%	56%	39%
Ik ben tevreden over hoe aardig andere kinderen mij vinden	1943	0,66	3,30	0,65	1%	6%	53%	39%
Ik ben tevreden over hoe vaak ik met andere kinderen afspreek	1940	0,43	3,38	0,72	2%	7%	41%	50%
Ik ben tevreden over hoe leuk leeftijdsgenoten mij vinden	1939	0,63	3,33	0,63	1%	5%	53%	40%
Ik ben er tevreden over hoe graag kinderen mij mogen	1943	0,67	3,31	0,66	1%	7%	51%	41%
Cognitief zelfbeeld								
Ik ben meestal tevreden over mijn werk op school	1941	0,57	3,23	0,61	1%	7%	61%	32%
Ik ben blij met hoe ik het doe op school	1934	0,65	3,40	0,62	0%	6%	47%	47%
Dingen kunnen altijd beter, maar ik vind dat ik de dingen op school goed doe	1937	0,56	3,32	0,62	1%	6%	54%	39%
Ik ben gelukkig met mijn manier van werken op school	1932	0,55	3,35	0,64	1%	7%	49%	43%
Als ik mijn best wil doen op school, dan lukt me dat ook	1941	0,50	3,25	0,67	1%	10%	52%	37%
Ik vind mezelf een goede leerling	1940	0,59	3,26	0,64	1%	7%	56%	35%
Ik ben tevreden met hoe ik het doe tijdens de les	1934	0,56	3,20	0,61	1%	9%	61%	30%
Schoolbeleving								
Ik vind het leuk op deze school	1946	0,68	3,51	0,65	1%	4%	37%	58%
Ik voel me ongelukkig op deze school	1940	0,49	3,54	0,79	4%	6%	21%	68%
De sfeer bij ons op school is goed	1944	0,56	3,32	0,64	1%	6%	53%	40%
Ik voel me thuis op deze school	1941	0,64	3,23	0,79	4%	11%	43%	42%
Ik heb een hekel aan deze school	1940	0,64	3,72	0,57	1%	2%	19%	77%
Ik hoor er op deze school niet echt bij	1938	0,52	3,61	0,67	2%	5%	23%	70%
Ik voel me veilig op deze school	1926	0,62	3,44	0,67	2%	5%	41%	52%
Gepest worden								
Andere kinderen lachen mij uit	1949	0,52	1,52	0,60	53%	43%	3%	1%
Een ander kind heeft mij uitgescholden	1945	0,63	1,70	0,71	42%	48%	7%	2%
Een ander kind maakt mij belachelijk	1937	0,57	1,42	0,65	65%	29%	5%	1%
Een ander kind heeft mij gepest	1935	0,57	1,46	0,72	65%	27%	5%	3%
Een ander kind heeft mij geschopt of geslagen (met opzet)	1942	0,63	1,48	0,66	59%	35%	5%	1%
Een ander kind heeft met mij gevochten (de ander is begonnen)	1939	0,49	1,44	0,66	63%	30%	4%	2%
Een ander kind heeft mij met opzet verwond (blauwe plek of erger)	1942	0,54	1,28	0,53	76%	22%	2%	1%
Motivatie voor leren								
Als ik leer, werk ik zo hard mogelijk	1948	0,56	3,19	0,65	1%	10%	57%	31%
Als ik aan iets begin, maak ik het af	1936	0,54	3,13	0,63	0%	13%	60%	26%
Als ik leer, doe ik mijn uiterste best	1937	0,57	3,37	0,62	0%	7%	49%	44%
Als ik leer, werk ik hard door, zelfs als de leerstof moeilijk is	1935	0,66	3,15	0,68	1%	13%	56%	30%
Ik ben een echte doorzetter, ook als het tegenzit	1946	0,63	3,02	0,71	2%	18%	55%	25%
Ik geef de moed niet op, ook al is de opdracht nog zo moeilijk	1947	0,64	3,10	0,69	1%	16%	55%	28%
Ik werk net zo lang aan een probleem tot ik het heb opgelost	1944	0,51	2,97	0,76	2%	23%	50%	25%
Sociale leeromgeving								
De kinderen in de groep zijn vrienden of vriendinnen van elkaar	1940	0,47	3,35	0,62	1%	5%	52%	42%
In mijn groep wordt niemand buitengesloten	1933	0,48	2,78	0,84	5%	33%	41%	21%
De kinderen in de groep gaan leuk met elkaar om	1940	0,68	3,14	0,68	1%	12%	57%	29%
De leerlingen in de groep zijn aardig voor elkaar	1929	0,70	3,11	0,64	1%	13%	60%	26%
De kinderen in de groep helpen elkaar	1933	0,56	3,28	0,57	0%	5%	61%	34%
De kinderen in de groep kunnen goed met elkaar opschieten	1929	0,66	3,16	0,61	1%	10%	63%	27%
De kinderen kunnen goed met elkaar spelen en samenwerken	1928	0,61	3,24	0,61	0%	8%	59%	33%
Onveilige leeromgeving								
Een kind roddelt over een ander kind	1943	0,49	2,17	0,82	19%	54%	19%	8%
Een kind lacht een ander kind uit	1929	0,65	1,98	0,77	26%	54%	15%	5%
Een kind scheldt een ander kind uit	1936	0,69	1,92	0,80	31%	50%	14%	5%
Een kind krabt of knijpt een ander kind	1932	0,55	1,54	0,69	55%	37%	5%	2%
Een kind schopt of slaat een ander kind	1936	0,65	1,77	0,73	38%	50%	9%	3%

	N	Rir	Gem	Stddev	Verdeling over de antwoordmogelijkheden			
					1	2	3	4
Een kind beschadigt of vernielt iets van een ander kind	1937	0,52	1,43	0,63	63%	32%	4%	1%
Een kind treitert of pest een ander kind	1939	0,64	1,66	0,76	49%	39%	9%	3%
Onordelijke leeromgeving								
Er is veel onrust en lawaai in de klas	1937	0,57	2,65	0,82	4%	46%	32%	18%
De leerlingen doen niet wat de juf of meester zegt	1936	0,55	1,89	0,73	29%	56%	11%	4%
De juf of meester krijgt de klas maar niet stil	1936	0,57	1,75	0,82	44%	41%	11%	4%
De leerlingen kunnen niet goed werken	1936	0,52	1,64	0,69	46%	46%	6%	2%
De leerlingen gaan pas aan het werk als de les al lang begonnen is	1935	0,47	1,59	0,76	54%	36%	6%	4%
Leerlingen kletsen en klieren als ze in groepjes werken	1931	0,51	2,24	0,81	14%	57%	19%	9%
De leerlingen verstoren de les	1938	0,61	2,04	0,75	21%	59%	15%	5%

Tabel 3.5 De uitspraken en enkele beschrijvende statistische gegevens van LL aanvullend

	N	Rir	Gem	Stddev	Verdeling over de antwoordmogelijkheden			
					1	2	3	4
Pro-sociaal gedrag								
Ik kan goed dingen delen met anderen	609	0,41	3,19	0,66	2%	7%	61%	30%
Ik help andere kinderen om hun ruzie op te lossen	608	0,51	2,82	0,76	6%	21%	57%	15%
Ik help andere kinderen als ze iets niet leuk vinden	612	0,59	2,92	0,69	4%	16%	64%	16%
Ik troost andere kinderen als zij pijn of verdriet hebben	613	0,56	3,19	0,70	2%	10%	54%	33%
Ik kan ruzies met andere kinderen zelf oplossen	618	0,46	2,89	0,77	5%	20%	56%	19%
Ik help andere kinderen	616	0,61	3,34	0,53	0%	1%	62%	36%
Ik kom op voor andere kinderen	616	0,61	3,07	0,66	2%	12%	62%	24%
Pestgedrag								
Ik maak ruzie met een ander kind	614	0,40	1,73	0,59	33%	62%	4%	1%
Ik lach andere kinderen uit	611	0,55	1,31	0,50	71%	28%	1%	0%
Ik heb een ander kind met opzet verwond (blauwe plek of erger)	611	0,56	1,14	0,38	87%	12%	1%	0%
Ik treiter of pest een ander kind	611	0,58	1,17	0,42	85%	15%	0%	0%
Ik sla of schop een ander kind	612	0,68	1,31	0,51	71%	28%	1%	0%
Ik vecht met een ander kind (ik ben zelf begonnen)	612	0,65	1,17	0,42	84%	16%	0%	0%
Ik scheld een ander kind uit	614	0,66	1,35	0,55	68%	29%	2%	0%
Storend gedrag								
Ik stoer de les	616	0,64	1,55	0,63	51%	45%	3%	2%
Ik houd me aan de gedragsregels in de klas	613	0,24	1,65	0,96	63%	17%	14%	7%
Ik houd andere kinderen van hun werk	611	0,59	1,59	0,70	51%	42%	5%	2%
Ik kan tijdens de les mijn mond niet houden	615	0,54	1,71	0,80	46%	43%	6%	5%
Ik stoer andere kinderen bij hun werk	613	0,72	1,50	0,61	55%	41%	3%	1%
Ik praat er doorheen als de juf of meester aan het woord is	616	0,66	1,52	0,63	55%	40%	4%	1%
Ik val andere kinderen lastig tijdens de les	614	0,68	1,40	0,60	65%	32%	2%	1%
Taakgericht gedrag								
Ik let goed op bij de lessen	613	0,57	3,30	0,60	1%	4%	58%	37%
Als de juf of meester een opdracht geeft, begin ik meteen	608	0,54	3,17	0,68	1%	11%	56%	32%
Ik werk slordig	616	0,43	3,12	0,82	4%	17%	43%	37%
Als we in groepjes werken, let ik goed op	612	0,55	3,27	0,59	1%	5%	60%	34%
Ik werk netjes	616	0,60	3,10	0,74	3%	14%	53%	30%
Als andere kinderen kletsen, werk ik rustig door	616	0,57	2,87	0,76	4%	25%	52%	19%
Als de leerkracht de klas uitloopt, werk ik rustig door	616	0,57	3,09	0,70	1%	16%	55%	28%
Motivatie voor lezen								
Ik vind lezen op school leuk	619	0,80	2,94	0,95	10%	19%	38%	32%
Ik lees graag thuis	616	0,59	2,78	1,00	13%	24%	35%	28%
Ik vind lezen op school saai	613	0,77	2,92	1,02	13%	18%	34%	35%
Als we gaan lezen op school, heb ik daar zin in	614	0,79	2,74	0,98	13%	25%	37%	25%
Ik lees omdat ik dat leuk vind	612	0,78	2,87	0,96	12%	18%	41%	29%
Motivatie voor rekenen								
Ik vind rekenen op school leuk	615	0,83	2,87	0,98	12%	19%	39%	30%
Ik vind het leuk om thuis te rekenen	615	0,66	2,19	1,04	31%	33%	21%	15%
Ik reken graag	615	0,89	2,70	1,02	17%	21%	38%	25%
Ik los graag sommen op	617	0,80	2,79	0,95	13%	20%	44%	24%
Als we gaan rekenen op school, heb ik daar zin in	618	0,87	2,66	1,00	17%	23%	38%	22%
Zelfbeeld lezen								
Ik begrijp de moeilijkste teksten	617	0,48	2,67	0,88	10%	29%	43%	17%
Ik lees goed	614	0,65	3,23	0,74	3%	9%	50%	38%
Ik lees langzamer dan de andere leerlingen in mijn klas	609	0,63	3,09	0,86	4%	20%	39%	37%
Ik lees niet zo goed als de meeste andere kinderen in mijn klas	615	0,60	2,96	0,95	9%	21%	36%	34%
Ik vind lezen makkelijk	615	0,69	3,27	0,84	5%	10%	37%	47%
Zelfbeeld rekenen								
Ik ben goed in rekenen	621	0,76	3,01	0,84	6%	16%	48%	30%
Als ik iets uitreken, heb ik daar meer tijd voor nodig dan andere kinderen in mijn klas	617	0,68	2,83	0,91	9%	26%	40%	26%
Als ik iets moet uitrekenen, heb ik daar meestal geen enkele moeite mee	614	0,68	2,83	0,87	7%	26%	44%	23%
Ik reken niet zo goed als de meeste andere kinderen in mijn klas	615	0,64	2,84	0,95	9%	26%	36%	29%
Ik vind rekenen makkelijk	611	0,79	2,89	0,91	8%	23%	41%	28%

	N	Rir	Gem	Stddev	Verdeling over de antwoordmogelijkheden			
					1	2	3	4
Spreekangst								
Als ik voor de hele klas iets moet vertellen, raak ik vaak in paniek	616	0,65	2,04	0,93	33%	39%	19%	9%
Ik vind het eng als ik in een groepje iets moet vertellen	615	0,61	1,64	0,79	52%	36%	8%	4%
Als ik in de klas iets hardop moet voorlezen, voel ik me erg gespannen	616	0,57	2,30	1,04	28%	29%	28%	15%
Ik ben vaak bang dat andere kinderen mij uittachen als ik in een groepje iets moet zeggen	619	0,65	1,82	0,93	47%	32%	14%	7%
Als de juf of meester onverwachts mijn naam noemt, krijg ik het meestal heel erg warm	617	0,65	2,15	0,99	32%	32%	25%	11%
Als de juf of meester mij opeens een vraag stelt, ben ik bang dat ik niet uit mijn woorden kan komen	616	0,67	1,97	0,92	37%	36%	20%	7%
Ik word helemaal zenuwachtig als ik eraan denk dat de hele klas naar mij kijkt	614	0,72	2,19	1,01	31%	31%	26%	12%
Als ik hoor dat we iets in groepjes moeten bespreken, raak ik vaak helemaal in paniek	613	0,56	1,51	0,72	59%	33%	6%	2%
Toetsangst								
Als ik een toets maak, ben ik bang om fouten te maken	611	0,70	2,13	1,00	31%	39%	17%	13%
Voordat ik een toets maak, heb ik vaak last van buikpijn	617	0,66	1,54	0,88	66%	21%	7%	6%
Als ik met een toets bezig ben, ben ik zo gespannen dat ik niet goed kan nadenken	616	0,74	1,68	0,96	57%	26%	8%	9%
Als ik een toets maak, voel ik me zenuwachtig	616	0,75	1,98	1,02	41%	34%	13%	13%
Als ik weet dat ik morgen een toets heb, maak ik me heel erg druk	614	0,74	1,77	1,00	53%	29%	7%	11%
Als ik een toets maak, denk ik vaak dat ik het niet goed doe	616	0,75	1,93	0,99	41%	38%	10%	12%
Het lijkt dat mijn hart extra hard klopt wanneer ik een toets maak	614	0,77	1,67	0,98	59%	24%	7%	10%
Bij het maken van een toets denk ik vaak dat het me niet zal lukken	614	0,77	1,78	0,93	48%	34%	8%	9%

4 Het normeringsonderzoek

In dit hoofdstuk bespreken we de normering van VISEON 2.0. In paragraaf 4.1 gaan we in op de opzet en de uitvoering van het normeringsonderzoek dat in april 2015 heeft plaatsgevonden. Het normeringsonderzoek werd gecombineerd met een onderzoek naar de betrouwbaarheid en de validiteit. De vragenlijsten zijn zowel op papier als digitaal via de computer aangeboden. In het normeringsonderzoek zijn gegevens verzameld over het sociaal-emotioneel functioneren van leerlingen in de jaargroepen 3, 4, 5, 6, 7 en 8 van het regulier basisonderwijs. In paragraaf 4.2 bespreken we de vraag of het geoorloofd is om de via verschillende afnamemodi en bij verschillende jaargroepen verzamelde gegevens tot één normgroep samen te voegen. In paragraaf 4.3 besteden we aandacht aan de representativiteit van de normeringsteekproef ten opzichte van de landelijke populatie. In paragraaf 4.4 ten slotte bespreken we de eigenlijke normering: de omzetting van de ruwe scores in percentielscores en niveauscores ofwel signalen.

4.1 Opzet en uitvoering van het normeringsonderzoek

VISEON 2.0 bestaat uit drie onderdelen die de gebruiker los van elkaar kan afnemen:

- de acht schalen van de basislijst voor leerkrachten met 55 uitspraken (LK basis);
- de acht schalen van de basislijst voor leerlingen met 55 uitspraken (LL basis);
- de tien schalen van de aanvullende lijst voor leerlingen met 64 uitspraken (LL aanvullend).

LK basis en LL basis zijn zowel digitaal als op papier aan leerkrachten en leerlingen voorgelegd, terwijl LL aanvullend alleen digitaal bij leerlingen is afgenomen.

De papieren afname

Een papieren versie van LK basis en LL basis is voorgelegd aan leerkrachten en leerlingen die deelnamen aan het normeringsonderzoek van de Entreetoets in jaargroep 7. Voor het gecombineerde normeringsonderzoek is een steekproef van 1502 scholen getrokken uit het bestand van scholen voor regulier basisonderwijs en voor deelname aan het onderzoek benaderd. Van de benaderde scholen gaven zich 206 (14%) op voor deelname. Van deze 206 scholen kregen er 67 ook boekjes voor VISEON 2.0 toegewezen. De leerkrachten is gevraagd de vragenlijst voor minimaal vijf, op basis van toeval door Cito geselecteerde leerlingen uit groep 7 in te vullen. De leerlinglijst was bestemd voor alle leerlingen uit groep 7. Uiteindelijk zijn op 57 scholen (4%) één of meer leerkrachtlijsten ingevuld, terwijl 60 scholen (4%) de leerlingen één of meer leerlinglijsten lieten invullen. De papieren afname vond plaats in april 2015. De schaal Hyperactief/impulsief van LK basis is alleen digitaal afgenomen (terwijl voor de overige zeven schalen zowel papieren als digitale afnamegegevens beschikbaar zijn). De reden is dat deze schaal oorspronkelijk voor het aanvullend gedeelte van VISEON 2.0 was bedoeld, maar uiteindelijk toch in LK basis is opgenomen.

De digitale afname

Voor de digitale afname van LK basis, LL basis en LL aanvullend is een steekproef van 200 scholen getrokken uit het bestand van scholen die in 2013 of 2014 de voorgaande versie van VISEON gebruikt hadden. Om zo goed mogelijk te kunnen generaliseren naar de populatie van basisschoolleerlingen in Nederland, zijn de 200 scholen getrokken met een trekkingskans evenredig aan het aantal leerlingen waarvoor de school een rapportage had aangevraagd. Hierbij is het aantal aangevraagde rapportages opgevat als een proxy voor schoolgrootte. De leerkrachten is gevraagd de digitale leerkrachtlijst voor tien, op basis van toeval door Cito geselecteerde leerlingen uit hun groep in te vullen. LL basis was bestemd voor alle leerlingen in de desbetreffende jaargroep (klas). Ondanks intensief rappelleren zijn uiteindelijk op achttien scholen (9%) één of meer leerkrachtlijsten ingevuld, terwijl zestien scholen (8%) de leerlingen één of meer leerlinglijsten lieten invullen. Bij de digitale afname zijn alle 26 schalen aan de respondenten voorgelegd. De digitale afname vond plaats in april 2015.

Aantal voor normering beschikbare vragenlijsten

Tabel 4.1 toont per afnamemodus (papier versus digitaal) per type respondent (leerkracht versus leerling) het aantal scholen en het aantal voor normering beschikbare lijsten.

Tabel 4.1 Aantal scholen en aantal voor normering beschikbare vragenlijsten per type respondent per afnamemodus

Type respondent	Afnamemodus	Aantal scholen	Aantal lijsten
Leerkracht	Digitaal	18	657
Leerling	Digitaal	16	624
Leerkracht	Papier	57	350
Leerling	Papier	60	1346
Totaal		151	2977

De verdeling van het aantal voor normering beschikbare vragenlijsten naar type respondent, afnamemodus en jaargroep is weergegeven in tabel 4.2. Wellicht ten overvloede wijzen wij erop dat de leerlinglijsten pas vanaf groep 5 worden afgenomen.

Tabel 4.2 Aantal groepen en aantal voor normering beschikbare vragenlijsten per type respondent per afnamemodus

Type respondent	Afnamemodus	Jaargroep	Aantal groepen	Aantal lijsten		
Leerkracht	Digitaal	3	10	89		
		4	11	114		
		5	12	121		
		6	11	101		
		7	12	153		
		8	8	79		
		Leerling	Digitaal	5	11	175
				6	11	161
7	13			168		
8	8			120		
Leerkracht	Papier	7	57	350		
Leerling	Papier	7	60	1346		
Totaal				2977		

Voor een groot deel van de 2977 leerlingen zijn zowel leerkracht- als leerlinglijsten beschikbaar. Daardoor is het aantal beschikbare vragenlijsten groter dan het aantal verschillende leerlingen dat aan het normeringsonderzoek deelnam. In totaal zijn van 2244 unieke leerlingen gegevens over het sociaal-emotioneel functioneren verkregen.

Aantal vragenlijsten per onderdeel van VISEON 2.0

Het aantal voor normering beschikbare leerlingen was niet voor elk onderdeel van het instrumentarium gelijk. De normering van de acht schalen van LK basis is gebaseerd op 1007 leerlingen. Daarvan deden er 350 mee aan de papieren afname en 657 aan de digitale afname.

De normering van de acht schalen van LL basis is gebaseerd op gegevens van 1970 leerlingen. Daarvan hadden er 1346 deelgenomen aan de papieren afname en 624 aan de digitale afname.

De responsgroep voor de normering van de tien schalen van LL aanvullend omvat de 624 leerlingen die aan de digitale afname deelnamen.

Afzonderlijke normgroepen per onderdeel van VISEON 2.0

De representativiteit van de respons is afzonderlijk bepaald voor LK basis, LL basis en LL aanvullend. Alvorens de resultaten te bespreken (zie paragraaf 4.3.1) leggen we eerst uit waarom voor afzonderlijke normering gekozen is. Voorbereidende analyses lieten zien dat de responsgroep van scholen die deelnam aan het normeringsonderzoek van de Entreeoets landelijk representatief is voor wat betreft regio, verstedelijking, leerlinggewicht, schoolgrootte en denominatie. Van de veel kleinere responsgroep van scholen die aan de digitale afname deelnam, bleek de representativiteit minder goed gewaarborgd (wat gezien het kleine aantal scholen niet verwonderlijk is). De reden om voor afzonderlijke normering van LK basis, LL basis en LL aanvullend te kiezen is dat het aandeel van de minder representatieve digitale respons sterk van onderdeel tot onderdeel verschilt. Van de scholen in de responsgroep voor LK basis heeft 29% deelgenomen aan de digitale afname tegenover 76% in de responsgroep van scholen voor LL basis. Van de leerlingen in de responsgroepen LK basis, LL basis en LL aanvullend heeft respectievelijk 59%, 32% en 100% deelgenomen aan de digitale afname. Het verschillende aandeel van de minder representatieve digitale respons is de reden dat wij de representativiteit voor elk van de drie onderdelen van VISEON 2.0 afzonderlijk bepaald hebben.

Ontbrekende waarnemingen

Een deel van de respondenten heeft niet alle vragen van de hun voorgelegde lijst ingevuld. Tabel 4.3 toont voor elk van de 26 schalen het percentage ontbrekende waarnemingen.

Tabel 4.3 Percentages ontbrekende waarnemingen bij de schalen van LK basis, LL basis en LL aanvullend voor de papieren en digitale afname

	LK basis (N = 1007)	LK papier (N = 350)	LK digitaal (N = 657)	LL basis (N = 1970)	LL papier (N = 1346)	LL digitaal (N = 624)
LK basis						
pro-sociaal gedrag	0,6	1,4	0,2			
pestgedrag	0,4	0,3	0,5			
storend gedrag	0,3	0,3	0,3			
teruggetrokken gedrag	0,2	0,3	0,2			
taakgericht gedrag	0,3	0,0	0,5			
motivatie voor leren	0,6	1,1	0,3			
zelfstandigheid	0,5	0,6	0,5			
hyperactief/impulsief	--	--	0,5			
LL basis						
sociaal zelfbeeld				4,4	1,8	10,1
cognitief zelfbeeld				5,5	2,3	12,3
schoolbeleving				5,2	3,6	8,7
gepest worden				4,5	3,1	7,4
motivatie voor leren				4,6	2,5	9,1
sociale leeromgeving				4,7	3,5	7,4
onveilige leeromgeving				4,2	3,3	6,1
onordelijke leeromgeving				4,3	2,5	8,2
LL aanvullend						
pro-sociaal gedrag						7,5
pestgedrag						8,8
storend gedrag						8,0
taakgericht gedrag						7,5
motivatie voor lezen						5,9
motivatie voor rekenen						5,1
zelfbeeld lezen						5,9
zelfbeeld rekenen						5,4
spreken in de klas						7,5
toetsangst						7,7

--: niet op papier afgenomen

Uit tabel 4.3 blijkt dat er bij de acht schalen van LK basis vrijwel geen waarnemingen ontbreken. Kennelijk hebben de leerkrachten de schriftelijke vragen zorgvuldig beantwoord en heeft de digitale afname bij hen goed gewerkt.

Bij de acht schalen van LL basis zijn de percentages ontbrekende waarnemingen duidelijk hoger dan bij LK basis. Vergelijking van de percentages van de papieren en digitale afname van de leerlingenlijst laat zien dat er bij de digitale afname van LL basis sprake is van een aanzienlijk verlies aan informatie. Dit wekt verbazing aangezien ontbrekende waarnemingen bij een digitale afname normaliter niet voorkomen. Immers, de computer signaleert onmiddellijk dat de leerling een vraag heeft overgeslagen, waarna de leerling pas naar het volgende scherm kan als hij de vraag alsnog beantwoord heeft. De oorzaak bleek een technische calamiteit met de interface waardoor er regelmatig één of meer (maar niet alle) scores van een schaal verloren gingen. Afhankelijk van de schaal betreft het 6 tot 12% van de antwoorden. De percentages ontbrekende waarnemingen bij de tien schalen van LL aanvullend geven een vergelijkbaar beeld te zien. Hier ontbreken 5 à 9 % van de antwoorden.

Uiteraard zijn wij nagegaan in hoeverre deze ontbrekende waarnemingen van invloed zijn op de schattingen van het sociaal-emotioneel functioneren. Zoals uiteengezet in paragraaf 4.2 geven onze analyses geen aanleiding tot de veronderstelling dat er sprake is geweest van een systematische vertekening. De ontbrekende waarnemingen mogen als random worden beschouwd, waarmee de integriteit van de digitale gegevens gewaarborgd is.

4.2 Verschillen tussen afnamemodi en jaargroepen

In het normeringsonderzoek zijn de vragenlijsten zowel digitaal als op papier afgenomen bij leerlingen van verschillende jaargroepen. In deze paragraaf bespreken we de vraag of het geoorloofd is om de via verschillende afnamemodi en bij verschillende jaargroepen verzamelde gegevens tot één normgroep samen te voegen. Samenvoeging tot één normgroep voor allen is alleen geoorloofd als er geen betekenisvolle verschillen zijn tussen de papieren en digitale afname en tussen de onderscheiden jaargroepen (groep 3 tot en met 8). In deze paragraaf doen wij verslag van een analyse van de verschillen tussen afnamemodi (paragraaf 4.2.1) en tussen jaargroepen (paragraaf 4.2.2). Daarbij kijken we niet alleen naar verschillen in het niveau van het sociaal-emotioneel functioneren, maar ook naar verschillen qua betrouwbaarheid en *Differential Item Functioning* (DIF).

Ook stellen we de vraag in hoeverre de op alle beschikbare gegevens bepaalde normen in de subgroepen die deelnamen aan de papieren en digitale afname tot dezelfde uitkomsten leiden.

4.2.1 Verschillen tussen afnamemodi

Uit een recent uitgevoerde inhoudsanalyse van dertien reviews en meta-analyses komt naar voren dat de afnamemodus voor de hoogte van de scores op meerkeuzetoetsen en gestandaardiseerde vragenlijsten niet of nauwelijks van belang is (Kuhlemeier, Straat & Van der Molen, 2015). Desalniettemin zijn wij nagegaan in hoeverre er verschillen zijn tussen de gegevens van de papieren afname en de digitale afname van VISEON 2.0.

Gemiddelden en standaarddeviaties per afnamemodus

De gemiddelden en standaarddeviaties van de papieren en digitale afnames van het normeringsonderzoek zijn weergegeven in tabel 4.4. Voor de schaal Hyperactief/impulsief van LK basis en de tien schalen van LL aanvullend zijn alleen digitale afnamegegevens beschikbaar zodat vergelijking van de papieren afname hier niet mogelijk is. Daarom vermeldt tabel 4.4 ook de gemiddelden en standaarddeviaties voor de papieren afname in de tweede proefafname. In de tweede proefafname vulden leerkrachten de leerkrachtvragenlijst in voor 437 leerlingen terwijl 501 leerlingen de leerlinglijst invulden. Voor meer informatie over de aard en omvang van de tweede proefafname verwijzen wij naar paragraaf 3.1.2. Uiteraard is vergelijking van de gegevens van het normeringsonderzoek en de tweede proefafname alleen zinvol als de uitspraken volledig

identiek zijn wat bij 23 van de 26 schalen het geval was. Bij de schalen Pestgedrag, Storend gedrag en Cognitief zelfbeeld geval zijn naar aanleiding van de tweede proefafname enkele uitspraken verwijderd, geherformuleerd of toegevoegd, waardoor de vergelijking met de gegevens van het normeringsonderzoek niet meer gewaarborgd is.

Tabel 4.4 Gemiddelden en standaarddeviaties per schaal, per afnamemodus, per type respondent voor het normeringsonderzoek en de tweede proefafname

	Normeringsonderzoek								Tweede proefafname			
	LK papier		LL papier		LK digitaal		LL digitaal		LK papier		LL papier	
	Gem	SD	Gem	SD	Gem	SD	Gem	SD	Gem	SD	Gem	SD
LK basis												
pro-sociaal gedrag	24,01	4,20			24,13	4,59			23,30	3,97		
pestgedrag	7,45	2,25			7,35	2,36			--	--		
storend gedrag	9,84	3,41			9,95	3,67			--	--		
teruggetrokken gedrag	13,43	4,03			12,86	4,08			13,47	3,71		
taakgericht gedrag	20,09	4,31			19,93	4,55			19,82	4,20		
motivatie voor leren	21,35	4,04			21,78	4,24			21,40	3,89		
zelfstandigheid	19,15	4,00			19,52	4,32			18,33	4,04		
hyperactief/impulsief	##	##			12,05	4,74			12,43	4,28		
LL basis												
sociaal zelfbeeld			20,29	2,78			20,19	2,78	20,48	3,46	20,62	3,37
cognitief zelfbeeld			23,05	3,03			22,93	3,14	--	--	--	--
schoolbeleving			24,67	3,18			23,78	3,75	24,15	3,93	24,09	3,99
gepest worden			10,12	3,13			10,61	3,17	10,05	3,20	10,37	3,62
motivatie voor leren			21,99	3,41			21,83	3,22	21,96	3,85	21,69	3,66
sociale leeromgeving			22,16	3,30			21,81	3,12	22,74	3,66	22,69	3,27
onveilige leeromgeving			12,39	3,80			12,62	3,54	12,62	3,54	11,67	3,84
onordelijke leeromgeving			13,62	3,72			14,24	3,54	13,46	4,20	13,97	4,07
LL aanvullend												
pro-sociaal gedrag							21,42	3,21	21,18	3,03	20,94	2,92
pestgedrag							9,15	2,37	8,75	2,05	8,93	2,10
storend gedrag							10,86	3,42	11,13	3,38	11,35	3,37
taakgericht gedrag							21,93	3,31	21,97	3,29	21,54	3,59
motivatie voor lezen							14,24	4,14	15,73	3,77	14,87	4,05
motivatie voor rekenen							13,18	4,37	13,72	4,41	13,95	4,45
zelfbeeld lezen							15,22	3,24	15,50	3,19	15,05	3,17
zelfbeeld rekenen							14,40	3,67	14,17	3,90	14,22	3,93
spreken in de klas							15,63	5,36	14,87	5,19	15,10	5,11
toetsangst							14,48	6,24	15,54	5,69	16,66	6,35

##: alleen digitaal afgenomen;

--: geen vergelijking papier versus digitaal mogelijk omdat de uitspraken niet identiek waren.

De vele gemiddelden en standaarddeviaties in tabel 4.4 zijn lastig rechtstreeks met elkaar te vergelijken. Daarom rapporteren we hierna ook de effectgroottes van de verschillen tussen de afnamemodi. Daarnaast presenteren we enkele samenvattende uitkomsten van een serie univariate variantieanalyses naar de invloed van de afnamemodus op de hoogte van de scores: de p-waarde en *Eta squared*. De p-waarde geeft de kans dat het verschil tussen de beide gemiddelden op toeval berust. *Eta squared* is een maat voor de proportie variantie in het sociaal-emotioneel functioneren dat door de afnamemodus verklaard wordt. Eerst bespreken we de verschillen bij LK basis, dan bij LL basis en tot slot bij LL aanvullend.

Niveaoverschillen tussen afnamemodi bij LK basis

Tabel 4.5 toont de effectgroottes van de verschillen tussen afnamemodi bij LK basis. In de kolomtitels verwijst het woord Papier naar de papieren afname, Digitaal naar de digitale afname, Proef naar de proefafname en Norm naar het normeringsonderzoek.

Tabel 4.5 Effectgroottes en resultaten variantieanalyse (p en $Eta squared$) van het verschil tussen de digitale en papieren afnames voor LK basis

	Effectgrootte			Variantie-analyse verschillen tussen Papier norm en Digitaal norm	
	Papier norm minus Digitaal norm	Papier proef minus Digitaal norm	Papier norm minus Papier proef	p	$Eta squared$
pro-sociaal gedrag	-0,03	-0,19	0,17	0,702	0,000
pestgedrag	0,04	--	--	0,510	0,000
storend gedrag	-0,03	--	--	0,666	0,000
teruggetrokken gedrag	0,14	0,16	-0,01	0,036	0,004
taakgericht gedrag	0,04	-0,03	0,06	0,605	0,000
motivatie voor leren	-0,10	-0,09	-0,01	0,115	0,002
zelfstandigheid	-0,09	-0,28	0,20	0,185	0,002
hyperactief/impulsief	--	0,08	--	--	--

De verschillen tussen de gemiddelden van de papieren en digitale afname van de leerkrachtlijst zijn zeer klein. Volgens de vuistregel van Cohen (1977) is er slechts bij één van de schalen sprake van een klein verschil (d.w.z. groter dan +0,20 of kleiner dan -0,20). Het betreft het verschil tussen de digitale afname in het normeringsonderzoek en de papieren afname in de proefafname bij de schaal Zelfstandigheid (Cohen's $d = -0,28$). Inspectie van de p -waarden van de variantieanalyse van de verschillen tussen de papieren en digitale afname in het normeringsonderzoek laat zien dat één van de verschillen significant is op het 5%-niveau ($p = 0,036$). Bij geen van de acht schalen verklaart de afnamemodus echter meer dan viertiende procent van de variantie in sociaal-emotioneel functioneren.

Niveaoverschillen tussen afnamemodi bij LL basis

Tabel 4.6 toont de effectgroottes van de verschillen tussen afnamemodi voor LL basis. In de kolomtitels verwijst de term Papier naar de papieren afname, Digitaal naar de digitale afname, Proef naar de proefafname en Norm naar het normeringsonderzoek.

Tabel 4.6 *Effectgroottes en resultaten variantieanalyse (p en Eta squared) van het verschil tussen de digitale en papieren afnames voor LL basis*

	Effectgrootte			Variantieanalyse verschillen tussen Papier norm en Digitaal norm	
	Papier norm minus Digitaal norm	Papier proef minus Digitaal norm	Papier norm minus Papier proef	<i>p</i>	<i>Eta squared</i>
sociaal zelfbeeld	0,03	0,11	-0,08	0,496	0,000
cognitief zelfbeeld	0,04	-.-	-.-	0,448	0,000
schoolbeleving	0,26	0,09	0,15	< 0,001	0,015
gepest worden	-0,16	-0,13	-0,02	0,002	0,005
motivatie voor leren	0,05	0,01	0,04	0,317	0,001
sociale leeromgeving	0,11	0,28	-0,16	0,027	0,003
onveilige leeromgeving	-0,06	-0,26	0,19	0,208	0,001
onordelijke leeromgeving	-0,17	-0,15	-0,01	0,001	0,006

Net als bij LK basis zijn de verschillen tussen de gemiddelden van de papieren en digitale afname – in termen van Cohen – hooguit ‘klein’. Slechts bij één van de acht schalen van LL basis is er in het normeringsonderzoek sprake van een klein verschil tussen beide afnamemodi. Het betreft de schaal Schoolbeleving ($d = 0,26$) en het verschil is significant ($p < 0,001$), evenals bij Gepest worden ($p = 0,002$), Sociale leeromgeving ($p = 0,027$) en Onordelijke leeromgeving ($p = 0,001$). Dat het hier wederom minieme verschillen betreft, blijkt uit de proportie verklaarde variantie *Eta squared*. Bij geen van de acht schalen verklaart de afnamemodus meer dan anderhalf procent van de variantie in het sociaal-emotioneel functioneren.

Niveaueverschillen tussen afnamemodi bij LL aanvullend

De overeenkomstige resultaten voor LL aanvullend zijn samengevat in tabel 4.7. In de kolomtitels verwijst de term Papier naar de papieren afname, Digitaal naar de digitale afname, Proef naar de proefafname en Norm naar het normeringsonderzoek.

Tabel 4.7 *Effectgroottes en resultaten variantieanalyse (p en Eta squared) van het verschil tussen de digitale en papieren afnames voor LK aanvullend*

Schaal	Effectgrootte Papier proef minus Digitaal norm	Variantieanalyse verschillen tussen Papier proef en Digitaal norm	
		<i>p</i>	<i>Eta squared</i>
pro-sociaal gedrag	-0,11	0,078	0,003
pestgedrag	-0,15	0,020	0,005
storend gedrag	0,11	0,091	0,003
taakgericht gedrag	-0,04	0,488	0,000
motivatie voor lezen	0,28	< 0,001	0,019
motivatie voor rekenen	0,15	0,017	0,005
zelfbeeld lezen	0,03	0,679	0,000
zelfbeeld rekenen	-0,06	0,361	0,001
spreken in de klas	-0,13	0,042	0,004
toetsangst	0,25	< 0,001	0,015

Volgens de vuistregel van Cohen (1977) is er bij twee van de tien schalen van LL aanvullend sprake van een klein verschil: Motivatie voor lezen ($d = 0,28$) en Toetsangst ($d = 0,25$). In beide gevallen geeft de variantieanalyse significantie te zien ($p < 0,001$), evenals bij Pestgedrag ($p = 0,020$), Motivatie voor rekenen

($p = 0,017$) en Spreekangst ($p = 0,042$). De proporties verklaarde variantie laten zien dat het hier marginale verschillen betreft. Bij geen van de tien schalen verklaart de afnamemodus namelijk meer dan 2% van de variantie in het sociaal-emotioneel functioneren.

Betrouwbaarheidsverschillen tussen afnamemodi

De betrouwbaarheden (Cronbachs alfa) van de digitale en papieren afname in het normeringsonderzoek en de papieren afname in de proefafname zijn weergegeven in tabel 4.8. In de kolomtitels verwijst het woord Papier naar de papieren afname, Digitaal naar de digitale afname, Proef naar de proefafname en Norm naar het normeringsonderzoek.

Tabel 4.8 *Betrouwbaarheidsverschillen tussen afnamemodi*

	LK papier norm	LK digitaal norm	LK papier proef	LL papier norm	LL digitaal norm	LL papier proef
LK basis	0,91	0,93	0,91			
pro-sociaal gedrag	0,86	0,89	--			
pestgedrag	0,87	0,87	--			
storend gedrag	0,88	0,90	0,88			
teruggetrokken gedrag	0,92	0,92	0,92			
taakgericht gedrag	0,90	0,92	0,91			
motivatie voor leren	0,91	0,92	0,91			
zelfstandigheid	0,91	0,93	0,91			
hyperactief/impulsief		0,95	0,93			
LL basis						
sociaal zelfbeeld				0,83	0,78	0,90
cognitief zelfbeeld				0,82	0,84	--
schoolbeleving				0,83	0,84	0,85
gepest worden				0,82	0,83	0,87
motivatie voor leren				0,84	0,82	0,85
sociale leeromgeving				0,85	0,81	0,85
onveilige leeromgeving				0,85	0,84	0,89
onordelijke leeromgeving				0,82	0,78	0,83
LL aanvullend						
pro-sociaal gedrag					0,82	0,79
pestgedrag					0,81	0,80
storend gedrag					0,81	0,77
taakgericht gedrag					0,89	0,90
motivatie voor lezen					0,93	0,93
motivatie voor rekenen					0,81	0,78
zelfbeeld lezen					0,88	0,88
zelfbeeld rekenen					0,87	0,86
spreken in de klas					0,92	0,90
toetsangst					0,88	0,91

Bij LK basis zijn de betrouwbaarheidscoëfficiënten van de digitale afname en de beide papieren afnames nagenoeg identiek. Bij geen van de acht schalen is het verschil tussen de betrouwbaarheden groter dan tweehonderdste.

Bij LL basis geeft slechts één van de vergelijkingen tussen de digitale en papieren afname een betrouwbaarheidsverschil te zien dat groter is dan vijfhonderdste. Het betreft de vergelijking van de digitale en papieren afname in het normeringsonderzoek waar de betrouwbaarheid van de papieren meting van Onordelijke leeromgeving achthonderdste hoger is dan die van de digitale meting (alfa respectievelijk 0,80 en 0,72). De betrouwbaarheidsverschillen tussen de digitale en papieren afname zijn ongeveer even klein als die tussen de beide papieren afnames.

Bij LL aanvullend varieert het betrouwbaarheidsverschil tussen de digitale afname in het normeringsonderzoek en de papieren afname in de proefafname van 0,00 voor Zelfbeeld lezen tot 0,037 voor Storend gedrag. Ook bij LL aanvullend is het effect van de afnamemodus op de betrouwbaarheid van de schalen dus verwaarloosbaar klein.

Differential Item Functioning

Behalve in de hoogte van de scores en de betrouwbaarheid kunnen verschillen tussen afnamemodi ook tot uiting komen in de betekenis van de uitspraken voor leerkrachten en leerlingen. Voor de schalen waarvoor in het normeringsonderzoek zowel papieren als digitale gegevens beschikbaar zijn, is nagegaan in hoeverre de uitspraken *Differential Item Functioning* (DIF) vertonen. Er is sprake van DIF als leerlingen met een gelijk niveau van sociaal-emotioneel functioneren afhankelijk van de afnamemodus op een uitspraak hoger of lager scoren dan op grond van hun totaalscore verwacht zou worden. Als een uitspraak DIF vertoont, wordt deze uitspraak op papier anders geïnterpreteerd dan digitaal. De analyse is per schaal uitgevoerd met de Mantel-Haenzel procedure (Heuvelmans, 2013). Van de uitkomsten wordt uitgebreid verslag gedaan in paragraaf 6.2 in het hoofdstuk over de validiteit. Hier volstaan we met de constatering dat er voor geen enkele uitspraak van de schalen uit het normeringsonderzoek DIF geconstateerd is. De uitspraken functioneren in beide afnamemodi op dezelfde manier. Er is dus geen reden om te veronderstellen dat de afnamemodus invloed heeft gehad op de interpretatie van de vragen (en daarmee op de normering).

Normeringsverschillen tussen afnamemodi

Zoals uiteengezet in de handleiding (Cito, 2015) worden de scores in de rapportage van VISEON 2.0 voor de gebruikers gerapporteerd als *signalen*. Er zijn drie signalen: Niet afwijkend (N), Afwijkend (A) en Zeer afwijkend (Z). Een signaal geeft aan of en in welke mate de score van een leerling afwijkt van leerlingen in de landelijke normgroep. Hoe de omzetting van scores naar signalen technisch in zijn werk gaat, is het onderwerp van paragraaf 4.3.3 over de normering. In het navolgende beperken we ons tot de vraag in hoeverre de op alle beschikbare gegevens bepaalde normen in de subgroepen die deelnamen aan de papieren en digitale afname tot dezelfde uitkomsten leiden (waarbij we aantekenen dat deze analyses voor LL aanvullend niet mogelijk waren omdat dit onderdeel alleen digitaal aan leerlingen is voorgelegd). Idealiter ontvangen in beide groepen de 80% laagst scorende leerlingen signaal N, de 5% hoogst scorende leerlingen signaal Z en de 15% tussen N en Z het signaal A. De afkappunten op de verdeling van de ruwe somscores tussen N en A en tussen A en Z zijn zo gekozen dat de ideale verdeling 80% in N, 15% in A en 5% in Z zo dicht mogelijk benaderd wordt (zie voor meer uitleg, paragraaf 4.3.3). De afkappunten zijn berekend op alle beschikbare gegevens, dus zowel van de digitale als de papieren afname. Onder de aanname van vergelijkbaarheid zouden beide afnamemodi tot vergelijkbare percentages leerlingen in de categorieën N, A en Z moeten leiden. Tabel 4.9 toont de verdeling van de leerlingen over de drie signalen afzonderlijk voor de papieren en de digitale afname.

Tabel 4.9 Verdeling van de leerlingen over de drie signalen voor de papieren en digitale afname in het normeringsonderzoek (percentages)

	Papieren afname normeringsonderzoek			Digitale afname normeringsonderzoek			p
	Niet afwijkend	Afwijkend	Ze er afwijkend	Niet afwijkend	Afwijkend	Ze er afwijkend	
LK basis							
pro-sociaal gedrag	84	11	5	80	16	4	0,147
pestgedrag	78	16	6	79	15	6	0,689
storend gedrag	80	15	4	80	15	6	0,714
teruggetrokken gedrag	79	14	7	82	13	5	0,154
taakgericht gedrag	80	14	6	77	15	8	0,254
motivatie voor leren	78	16	5	82	13	5	0,118
zelfstandigheid	82	12	6	81	14	5	0,912
hyperactief/impulsief	--	--	--	80	14	6	--
LL basis							
sociaal zelfbeeld	77	17	6	76	19	5	0,933
cognitief zelfbeeld	83	12	4	85	10	5	0,605
schoolbeleving	85	11	4	78	14	8	< 0,001
gepest worden	82	13	5	78	18	4	0,067
motivatie voor leren	77	17	6	79	15	6	0,365
sociale leeromgeving	82	14	4	82	14	5	0,822
onveilige leer- omgeving	77	18	5	77	18	5	0,828
onordelijke leer- omgeving	81	14	5	77	17	6	0,084

De verdelingen blijken vrijwel zonder uitzondering sterk op elkaar te lijken. Met behulp van ordinale logistische regressieanalyse (McCullagh, 1980) is getoetst in hoeverre de uniforme normen in beide afnamegroepen tot dezelfde uitkomsten leiden. De rechter kolom van tabel 4.9 toont de p-waarde. Slechts één van de 26 uitgevoerde significantietoetsingen is significant op 5%-niveau. Het betreft de schaal Schoolbeleving waar de verdeling over de signalen N, A en Z voor de papieren afname 85%, 11% en 4% bedraagt tegen 78%, 14% en 8% voor de digitale afname. Bij de overige 25 schalen geven de geconstateerde verschillen geen aanleiding om de vergelijkbaarheid van de normen op basis van de papieren en digitale gegevens in twijfel te trekken.

4.2.2 Verschillen tussen jaargroepen

In het normeringsonderzoek en de proefafnames zijn gegevens verzameld over het sociaal-emotioneel functioneren van leerlingen in groep 3 tot en met 8 (leerkrachtvragenlijst) of 5 tot en met 8 (leerlingvragenlijst). In deze paragraaf gaan we in op de vraag of er afzonderlijke normen per jaargroep noodzakelijk zijn dan wel dat volstaan kan worden met één normgroep voor allen. Als de verschillen tussen jaargroepen verwaarloosbaar of zeer klein zijn, is er geen aanleiding om afzonderlijke normgroepen per jaargroep te vormen. Ter kruisvalidatie analyseren we de verschillen tussen jaargroepen zowel op basis van de gegevens van het normeringsonderzoek als de daaraan voorafgaande proefafname.

Niveaueverschillen tussen jaargroepen

Tabel 4.10 toont de effectgroottes van de verschillen tussen jaargroepen in het gemiddeld niveau van sociaal-emotioneel functioneren, afzonderlijk voor het normeringsonderzoek en de daaraan voorafgaande tweede proefafname. Achtereenvolgens zijn weergegeven de effectgroottes voor het verschil tussen jaargroep 3 minus 4 (alleen voor de leerkrachtvragenlijst), 4 minus 5, 5 minus 6, 6 minus 7 en tot slot 7 minus 8. Daarin is het verschil tussen de beide gemiddelden telkens uitgedrukt als een proportie van de standaarddeviatie (waarbij we opmerken dat de tabel met de gemiddelden en standaarddeviaties vanwege

de enorme omvang niet is weergegeven). Bij de evaluatie van de grootte van de verschillen baseren we ons op de vuistregel van Cohen (1977) waarbij 0,20, 0,50 en 0,80 respectievelijk een klein, middelgroot en groot verschil vertegenwoordigt. Om van een duidelijke dalende of stijgende trend te kunnen spreken moeten de opeenvolgende effectgroottes groter zijn dan +0,20 of -0,20.

Tabel 4.10 *Effectgroottes van de verschillen tussen jaargroepen in het normeringsonderzoek en de proefafname*

	Normeringsonderzoek					Proefafname				
	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8
LK basis										
pro-sociaal gedrag	0,12	-0,01	-0,14	-0,13	-0,09	-0,12	0,43	0,00	-0,52	0,15
pestgedrag	0,16	-0,56	0,06	-0,05	0,40	--	--	--	--	--
storend gedrag	-0,11	0,27	-0,12	-0,22	0,48	--	--	--	--	--
teruggetrokken gedrag	0,00	0,05	-0,07	0,34	-0,18	-0,21	-0,27	0,19	0,30	-0,07
taakgericht gedrag	0,41	-0,04	-0,40	0,17	-0,20	0,29	-0,05	-0,05	0,02	-0,19
motivatie voor leren	0,28	0,11	-0,04	0,08	-0,18	0,34	0,29	-0,16	-0,10	0,15
zelfstandigheid	0,11	-0,11	-0,07	-0,03	0,07	0,03	0,24	0,00	-0,11	0,09
hyperactief/impulsief	0,01	0,08	0,02	0,00	0,24	-0,08	-0,12	0,15	0,51	-0,24
LL basis										
sociaal zelfbeeld			0,05	-0,19	0,05			0,12	-0,09	-0,17
cognitief zelfbeeld			0,29	0,00	-0,04			--	--	--
schoolbeleving			-0,06	0,07	-0,07			-0,01	-0,26	-0,10
gepest worden			-0,02	0,16	-0,04			0,07	0,26	0,22
motivatie voor leren			0,32	0,12	0,18			0,23	0,05	0,10
sociale leeromgeving			0,04	-0,13	0,01			0,22	0,03	-0,02
onveilige leeromgeving			0,14	-0,06	-0,07			0,15	-0,01	0,13
onordelijke leeromgeving			-0,07	-0,08	0,12			0,18	-0,11	0,14
LL aanvullend										
pro-sociaal gedrag			0,12	-0,13	0,13			0,05	-0,29	-0,12
pestgedrag			-0,16	-0,09	0,02			-0,68	0,26	-0,04
storend gedrag			-0,38	-0,09	0,09			-0,41	-0,10	-0,08
taakgericht gedrag			0,29	0,24	0,18			0,37	0,09	0,05
motivatie voor lezen			-0,01	0,24	0,18			0,20	-0,03	0,48
motivatie voor rekenen			0,35	0,02	0,27			0,18	-0,17	0,40
zelfbeeld lezen			0,12	0,09	-0,14			0,15	0,09	-0,13
zelfbeeld rekenen			0,22	-0,04	0,05			0,23	-0,11	0,22
spreken in de klas			0,20	-0,08	0,14			-0,01	-0,26	0,27
toetsangst			0,18	0,04	0,26			0,06	-0,01	0,11

Overeenkomstig de vuistregel van Cohen (1977) zijn de effectgroottes in tabel 4.10 overwegend verwaarloosbaar klein. Bovendien vertonen ze vrijwel zonder uitzondering een grillig patroon aangezien er geen sprake is van een duidelijke en gelijkmatige toe- of afname van de lagere naar de hogere jaargroepen. Bij het normeringsonderzoek doen zich echter drie uitzonderingen voor. Bij de schaal Motivatie voor leren uit LL basis lijkt er een vrij regelmatige afname te zijn van groep 5 naar groep 8 (waarbij we opmerken dat alleen het verschil tussen groep 5 en 6 volgens de vuistregel van Cohen aan de eis van een klein effect voldoet). Ook bij de schalen Taakgericht gedrag en Motivatie voor rekenen uit LL aanvullend zien we een vrij duidelijke en regelmatige afname (waarbij er volgens Cohen bij twee van de drie contrasten sprake is van een klein verschil). Kijken we echter naar de overeenkomstige effectgroottes voor de proefafname, dan is deze ogenschijnlijke afname veel minder duidelijk aangezien de afname bij Motivatie voor leren en Taakgericht gedrag zich alleen van groep 5 naar groep 6 voordoet en bij Motivatie voor rekenen alleen van groep 7 naar groep 8.

Om meer zicht te krijgen op de grootte van de geconstateerde verschillen tussen jaargroepen is een serie univariate variantieanalyses uitgevoerd. De resultaten van de significantietoetsing van de verschillen tussen jaargroepen (p -waarde) en de proportie verklaarde variantie in het sociaal-emotioneel functioneren (*Eta squared*) zijn weergegeven in tabel 4.11 (ter kruisvalidatie afzonderlijk voor het normeringsonderzoek en de proefafname). Wellicht ten overvloede merken we hierbij op dat de p -waarden in het normeringsonderzoek en de proefafname vanwege het verschillende aantal waarnemingen minder goed vergelijkbaar zijn.

Tabel 4.11 Resultaten variantieanalyse verschillen tussen jaargroepen in het normeringsonderzoek en de proefafname

	Normeringsonderzoek		Proefafname	
	p	<i>Eta squared</i>	p	<i>Eta squared</i>
LK basis				
pro-sociaal gedrag	0,015	0,021	0,003	0,041
pestgedrag	< 0,001	0,037	--	--
storend gedrag	0,005	0,026	--	--
teruggetrokken gedrag	0,040	0,018	0,018	0,031
taakgericht gedrag	0,001	0,031	0,325	0,014
motivatie voor leren	0,070	0,016	0,006	0,037
zelfstandigheid	0,634	0,005	0,339	0,013
hyperactief/impulsief	0,253	0,010	0,002	0,043
LL basis				
sociaal zelfbeeld	0,339	0,006	0,325	0,007
cognitief zelfbeeld	0,048	0,014	--	--
schoolbeleving	0,906	0,001	0,008	0,024
gepest worden	0,556	0,004	0,001	0,036
motivatie voor leren	< 0,001	0,049	0,017	0,021
sociale leeromgeving	0,670	0,003	0,111	0,012
onveilige leeromgeving	0,563	0,004	0,228	0,009
onordelijke leeromgeving	0,563	0,004	0,314	0,007
LL aanvullend				
pro-sociaal gedrag	0,515	0,004	0,004	0,027
pestgedrag	0,149	0,009	< 0,001	0,060
storend gedrag	< 0,001	0,031	< 0,001	0,049
taakgericht gedrag	< 0,001	0,065	< 0,001	0,043
motivatie voor lezen	0,001	0,026	< 0,001	0,053
motivatie voor rekenen	< 0,001	0,048	0,010	0,023
zelfbeeld lezen	0,303	0,006	0,254	0,008
zelfbeeld rekenen	0,147	0,009	0,045	0,016
spreken in de klas	0,135	0,01	0,090	0,013
toetsangst	0,003	0,024	0,670	0,003

Tabel 4.11 bevestigt onze eerdere conclusie dat de verschillen tussen jaargroepen doorgaans verwaarloosbaar klein zijn. Zoals we eerder op basis van de effectgroottes constateerden, zijn de verschillen tussen jaargroepen bij Motivatie voor leren (LL basis), Taakgericht gedrag (LL aanvullend) en Motivatie voor rekenen (LL aanvullend) wat minder klein dan bij de overige schalen. De toetsing van de verschillen tussen jaargroepen geeft bij deze drie schalen significantie te zien ($p < 0,05$) wat op basis van het relatief grote aantal waarnemingen weinig verbazing wekt. Kijken we naar *Eta squared*, dan blijken de verschillen tussen jaargroepen bij Motivatie voor leren, Taakgericht gedrag en Motivatie voor rekenen in het normeringsonderzoek respectievelijk 5%, 7% en 5% van de variantie in het sociaal-emotioneel functioneren te verklaren. In de proefafname verklaren Motivatie voor leren, Taakgericht gedrag en Motivatie voor rekenen daarentegen veel minder variantie, respectievelijk 2%, 4% en 2%. Al met al bevestigen de uitkomsten van de variantieanalyses de eerdere constatering dat de verschillen tussen jaargroepen overwegend verwaarloosbaar klein zijn. We maken dus geen grote fout als we de gegevens van de

onderscheiden jaargroepen tot één normgroep voor allen zouden samenvoegen. In het tekstdeel hierna presenteren we aanvullende evidentie voor deze beslissing.

Differential Item Functioning

Behalve in de hoogte van de scores en de betrouwbaarheid kunnen verschillen tussen jaargroepen ook tot uiting komen in de betekenis van de uitspraken voor leerkrachten en leerlingen. Nagegaan is in hoeverre de uitspraken *Differential Item Functioning* (DIF) ten gevolge van de jaargroep vertonen. Er is sprake van DIF als leerlingen met een gelijk niveau van sociaal-emotioneel functioneren afhankelijk van de jaargroep op een uitspraak hoger of lager scoren dan op grond van hun totaalscore verwacht zou worden.

Bij afwezigheid van DIF mogen we constateren dat de uitspraken voor leerlingen van verschillende jaargroepen dezelfde betekenis hebben gehad. De analyse is per schaal uitgevoerd met de Mantel-Haenzel procedure (Heuvelmans, 2013). Van de uitkomsten wordt uitgebreid verslag gedaan in paragraaf 6.2 in het hoofdstuk over de validiteit. Hier volstaan we met de constatering dat er voor geen enkele uitspraak van de schalen uit het normeringsonderzoek DIF in relatie tot de jaargroep geconstateerd is. De uitspraken functioneren voor jongere en oudere leerlingen op dezelfde manier. Er is dus geen reden om te veronderstellen dat de jaargroep invloed heeft gehad op de beantwoording van de vragen (en daarmee op de normering).

4.2.3 Conclusies

Zoals op grond van het literatuuronderzoek van Kuhlemeier, Straat en Van der Molen (2015) verwacht werd, bracht de vergelijking van de papieren en digitale afname geen noemenswaardige verschillen tussen de papieren en digitale afname aan het licht. Dit geldt zowel voor de verschillen in het niveau van het sociaal-emotioneel functioneren, de betrouwbaarheid van de meting, de validiteit in termen van *Differential Item Functioning* als de normering.

Behoudens een enkele uitzondering zijn de effectgroottes voor de verschillen tussen het niveau van het sociaal-emotioneel functioneren in de opeenvolgende jaargroepen verwaarloosbaar klein. De effectgroottes voor de verschillen tussen opeenvolgende jaargroepen vertonen een grillig patroon, zonder dat er sprake is van een min of meer gelijkmatige toe- of afname van het sociaal-emotioneel functioneren van de lagere naar de hogere jaargroepen.

De algehele conclusie is dat er geen aanleiding is voor afzonderlijke normering per afnamemodus of per jaargroep. De onder verschillende afnamecondities bij verschillende jaargroepen verzamelde gegevens mogen tot één normgroep worden samengevoegd.

4.3 De representativiteit van de normgroepen

Hiervoor is vastgesteld dat de onder verschillende afnamecondities bij verschillende jaargroepen verzamelde gegevens tot één normgroep mogen worden samengevoegd. Hierna bespreken we de representativiteit van de normgroep van scholen (paragraaf 4.3.1) en van leerlingen (paragraaf 4.3.2). Daarna bespreken we de normering: de omzetting van de ruwe somscores naar percentielen en niveauscores ofwel signalen (paragraaf 4.3.3).

4.3.1 De representativiteit van de normgroepen op schoolniveau

De representativiteit op schoolniveau is bepaald op basis van de volgende schoolkenmerken:

- regio (Noord, Oost, West en Zuid);
- verstedelijking;
- leerlinggewicht: proportie leerlingen op school met een gewicht van 0,00, 0,30 en 1,20;
- schoolgrootte: aantal leerlingen per school;
- denominatie.

De analyse is gebaseerd op de gegevens van alle 6696 scholen voor regulier basisonderwijs in Nederland. De uitkomsten van de analyse van de representativiteit van de responsgroep van scholen naar regio, verstedelijking, leerlinggewicht, schoolgrootte en denominatie zijn weergegeven in tabel 4.12. Zoals hiervoor uiteengezet, is de representativiteit voor elk van de drie onderdelen van VISEON 2.0 afzonderlijk bepaald. De tweede, derde, vierde en vijfde kolom bevatten achtereenvolgens de procentuele verdelingen in de normgroep van scholen voor LK basis, LL basis, LL aanvullend en de populatie van scholen voor regulier basisonderwijs in Nederland.

Tabel 4.12 De representativiteit van de normgroep LK basis, LL basis en LL aanvullend op schoolniveau (percentages)

	LK basis (N = 75)	LL basis (N = 76)	LL aanvullend (N = 16)	Populatie (N = 6696)
Regio				
Noord	11	9	13	15
Oost	20	18	13	25
West	43	43	25	42
Zuid	27	29	50	18
Totaal	100	100	100	100
Verstedelijking				
Zeer sterk stedelijk	15	16	0	12
Sterk stedelijk	23	21	44	22
Matig stedelijk	23	22	19	20
Weinig stedelijk	23	22	13	26
Niet stedelijk	17	18	25	20
Totaal	100	100	100	100
Verstedelijking: stad versus platteland				
Sterk tot zeer sterk verstedelijkt (stad)	37	37	44	35
Niet tot matig verstedelijkt (platteland)	63	63	56	66
Totaal	100	100	100	100
Leerlinggewicht				
tot 10%	63	61	75	64
10%-25%	19	20	13	25
25%-40%	8	9	6	7
> 40%	11	11	6	4
Totaal	100	100	100	100
Schoolgrootte				
tot 100	20	21	13	20
101-200	37	37	25	33
201-300	21	22	38	26
301-400	13	12	13	12
> 400	8	8	13	10
Totaal	100	100	100	100
Denominatie				
Algemeen bijzonder	7	7	13	5
Openbaar	28	28	13	33
Rooms-katholiek	36	36	56	30
PC, Ger, Ref en Evan	23	24	13	30
Overig	7	7	6	3
Totaal	100	100	100	100

Met behulp van Chi-kwadraattoetsing is nagegaan in hoeverre de afwijkingen van representativiteit significant zijn. De uitkomsten van de toetsing staan in tabel 4.13.

Tabel 4.13 *Uitkomsten van de Chi-kwadraattoetsing van de representativiteit op schoolniveau (p-waarden)*

Achtergrondkenmerk	<i>p</i>	<i>p</i>	<i>p</i>
	LK basis	LL basis	LL aanvullend
Regio	0,218	0,055	0,013
Verstedelijking	0,861	0,794	0,157
Verstedelijking: stad versus platteland	0,345	0,377	0,298
Leerlinggewicht	0,040	0,042	0,693
Schoolgrootte	0,848	0,898	0,791
Denominatie	0,187	0,226	0,045

Regio

In de normgroepen voor LK basis en LL basis zijn scholen in regio Zuid enigszins oververtegenwoordigd ten koste van regio Oost en West. De Chi-kwadraattoetsing gaf echter geen significantie te zien (p respectievelijk 0,218 en 0,055).

De verschillen tussen de percentages in de responsgroep en de populatie zijn voor LL aanvullend groter dan voor LK basis en LL basis. In vergelijking met de landelijke verdeling bevat de normgroep voor LL aanvullend zeer veel scholen uit de regio's Oost en West terwijl scholen uit regio Zuid zwaar ondervertegenwoordigd zijn. Dat het hier om serieus te nemen afwijkingen gaat, blijkt ook uit de Chi-kwadraattoetsing die ondanks het kleine aantal scholen in de responsgroep significantie te zien geeft ($p = 0,013$).

Verstedelijking

In de normgroepen voor LK basis en LL basis lijkt de verdeling van de scholen naar verstedelijking sterk op die in de populatie van scholen voor basisonderwijs in Nederland. De verschillen tussen responsgroep en populatie variëren hier van 0% tot 4%. In de normgroep LL aanvullend lijken scholen uit sterk stedelijke gebieden oververtegenwoordigd, terwijl scholen uit zeer sterk stedelijke en niet stedelijke gebieden ondervertegenwoordigd lijken, maar ook hier geeft de Chi-kwadraattoetsing geen significantie te zien.

Verstedelijking: stad versus platteland

In de normgroepen voor LK basis en LL aanvullend wijkt de verdeling van de scholen naar stad versus platteland hooguit 3% af van de populatieverdeling. De groep LL basis bevat naar verhouding wat te veel stadsscholen en wat te weinig plattelandsscholen. Geen van de drie Chi-kwadraattoetsingen leverde significantie op.

Leerlinggewicht

Bij LK basis en LL basis zijn scholen met meer dan 40% gewichtsl leerlingen enigszins oververtegenwoordigd, terwijl scholen met hooguit 10% en 25 à 40% gewichtsl leerlingen iets te zwak vertegenwoordigd zijn.

De verschillen tussen de verdeling in responsgroep en de populatie variëren hier van 1 tot 6% en zijn significant (p respectievelijk 0,040 en 0,042).

In LL aanvullend zijn de afwijkingen wat groter, maar niet significant ($p = 0,693$). Een plausibele verklaring is het gebrek aan onderscheidingsvermogen vanwege het uitzonderlijk kleine aantal waarnemingen ($N = 16$).

Schoolgrootte

In LK basis en LL leerling lijkt de verdeling van de responsgroep naar schoolgrootte zeer sterk op die in de populatie. In LL aanvullend zijn scholen met tussen de 300 en 400 leerlingen lichtelijk oververtegenwoordigd ten koste van zeer kleine scholen (tot 100 leerlingen) en kleine scholen (tussen 100 en 200 leerlingen). De Chi-kwadraatanalyse toont aan dat deze kleine verschillen niet significant zijn (p respectievelijk 0,848, 0,898 en 0,791).

Denominatie

In LK basis en LL basis lijken scholen van openbare en protestants-christelijke signatuur in de responsgroep licht ondervertegenwoordigd, terwijl rooms-katholieke scholen wat oververtegenwoordigd lijken. Gezien de uitkomsten van de Chi-kwadraattoetsing berusten deze verschillen echter op toeval (p is respectievelijk 0,187 en 0,226).

De responsgroep LL aanvullend vertoont een vergelijkbaar beeld, met dien verstande dat de afwijkingen van representativiteit hier groter zijn dan bij het basisinstrumentarium. Volgens de Chi-kwadraattoetsing gaat het hier om een significant verschil ($p = 0,045$).

Conclusie

Al met al hebben we niet kunnen aantonen dat de responsgroepen die gebruikt zijn voor de normering van LK basis en LL basis niet representatief zijn voor de populatie van basisscholen in Nederland. Slechts bij één steekproefkenmerk zien we een significant verschil: beide responsgroepen bevatten naar verhouding wat meer scholen met meer dan 40% gewichtsleerlingen, maar ook hier is de oververtegenwoordiging niet groot.

In de responsgroep LL aanvullend is regio Zuid wat te sterk vertegenwoordigd en regio Oost en West wat te zwak. Ook bevat deze groep relatief veel rooms-katholieke scholen en naar verhouding weinig openbare en protestants-christelijke scholen.

4.3.2 De representativiteit van de normgroepen op leerlingniveau

De gegevens over de representativiteit van de responsgroep van leerlingen naar regio, verstedelijking, leerlinggewicht, schoolgrootte, denominatie en geslacht zijn weergegeven in tabel 4.14. De tweede, derde en vierde kolom bevatten achtereenvolgens de procentuele verdeling van deze kenmerken in de normgroep voor LK basis, LL basis en LL aanvullend. De vijfde kolom geeft de procentuele verdeling voor de bijna anderhalf miljoen basisschoolleerlingen in Nederland (schooljaar 2014-2015).

Het aantal voor de significantietoetsing beschikbare waarnemingen is zo groot dat zelfs zeer kleine afwijkingen van representativiteit significant zijn ($p < 0,001$). Omdat alle toetsingen significantie te zien gaven, ($p < 0,001$) zien we af van verdere rapportage van de uitkomsten van de Chi-kwadraattoetsing en baseren we ons uitsluitend op de verschillen tussen de procentuele verdelingen in de responsgroep en de populatie van leerlingen in Nederland.

Tabel 4.14 De representativiteit van de normgroepen op leerlingniveau (percentages)

	LK basis (N = 1007)	LL basis (N = 1970)	LL aanvullend (N = 624)	Populatie (N = 1456268)
Regio				
Noord	11	6	9	10
Oost	15	19	14	22
West	35	41	24	48
Zuid	39	35	53	20
Totaal	100	100	100	1
Verstedelijking				
Zeer sterk stedelijk	7	12	0	17
Sterk stedelijk	39	24	43	28
Matig stedelijk	20	32	29	21
Weinig stedelijk	15	18	8	22
Niet stedelijk	19	13	20	12
Totaal	100	100	100	1
Verstedelijking: stad versus platteland				
Sterk tot zeer sterk verstedelijkt (stad)	46	36	43	45
Niet tot matig verstedelijkt (platteland)	54	64	57	55
Totaal	100	100	100	1
Leerlinggewicht		0	0	0
tot 10%	65	66	71	67
10%-25%	21	19	18	22
25%-40%	4	5	3	7
> 40%	9	9	8	4
Totaal	100	100	100	0
Schoolgrootte				
tot 100	14	10	9	6
101-200	33	30	22	23
201-300	30	31	45	29
301-400	12	16	16	20
> 400	11	13	9	23
Totaal	100	100	100	100
Denominatie				
Algemeen bijzonder	9	7	13	5
Openbaar	24	26	6	30
Rooms-katholiek	45	41	60	33
PC, Ger, Ref en Evan	16	21	14	28
Overig	7	6	8	3
Totaal	100	100	100	100
Geslacht				
Jongen	50	53	50	50
Meisje	50	47	50	50
Totaal	100	100	100	100

Regio

Leerlingen uit regio Zuid zijn in de normgroepen voor LK basis en LL basis oververtegenwoordigd ten koste van regio West en Oost. In LL aanvullend zijn de afwijkingen aanzienlijk groter: deze responsgroep bevat relatief veel leerlingen uit regio Zuid en relatief weinig leerlingen uit regio West.

Verstedelijking

In LK basis zijn leerlingen uit sterk stedelijke gebieden enigszins oververtegenwoordigd ten koste van leerlingen uit zeer sterk stedelijke, weinig stedelijke en niet stedelijke gebieden. De groep LL basis bevat naar verhouding wat meer leerlingen uit matig stedelijke gebieden.

De groep LL aanvullend bevat relatief veel leerlingen uit zeer sterk en weinig stedelijke gebieden, terwijl de sterk stedelijke en weinig stedelijke gebieden daar minder sterk vertegenwoordigd zijn.

Verstedelijking: stad versus platteland

Kijken we naar de verhouding stads- versus plattelandsscholen, dan blijken de responsgroepen LK basis en LL aanvullend nagenoeg landelijk representatief te zijn. De afwijkingen van de landelijke verdeling zijn hier 1% (LK basis) en 2% (LL aanvullend). Bij de groep LL basis is de afwijking van representativiteit groter: 9%.

Leerlinggewicht

Ten aanzien van leerlinggewicht blijkt de verdeling van de respons in de groepen LK basis, LL basis en LL aanvullend de landelijke verdeling dicht te benaderen. De afwijkingen variëren van 1% tot 6%.

Schoolgrootte

In LK basis blijken leerlingen van grote scholen (301-400 leerlingen) en zeer grote scholen (meer dan 400 leerlingen) enigszins ondervertegenwoordigd, terwijl leerlingen van zeer kleine scholen (0 tot 100 leerlingen) en kleine scholen (101 tot 200 leerlingen) lichtelijk oververtegenwoordigd zijn.

De groep LL basis bevat naar verhouding wat te weinig leerlingen van zeer grote scholen (> 400 leerlingen) ten faveure van kleine scholen (101-200 leerlingen).

In LL aanvullend zitten verhoudingsgewijs iets minder leerlingen van grote scholen (> 400) terwijl deze groep relatief gezien wat meer leerlingen van middelgrote scholen (201-300) bevat.

Denominatie

In vergelijking met de landelijke verdeling zijn leerlingen van openbare en protestants-christelijke scholen in de normgroep voor LK basis enigszins ondervertegenwoordigd, terwijl leerlingen van rooms-katholieke scholen daar lichtelijk oververtegenwoordigd zijn.

In de normgroep voor LL basis zijn de afwijkingen nog iets kleiner. Bij LL aanvullend zien we daarentegen forse afwijkingen: relatief weinig leerlingen van openbare en protestants-christelijke scholen en veel leerlingen van rooms-katholieke scholen.

Conclusie

Ten aanzien van verstedelijking (stad versus platteland) en leerlinggewicht mogen de normgroepen voor LK basis en LL basis als landelijk representatief worden beschouwd, terwijl de verdelingen naar regio, schoolgrootte en denominatie voor sommige categorieën overwegend lichte afwijkingen van het landelijke beeld te zien geven.

De normgroep voor LL aanvullend is landelijk representatief voor wat betreft verstedelijking (stad versus platteland) en leerlinggewicht, maar bij regio, schoolgrootte en denominatie zien we forse afwijkingen van het landelijke beeld.

Weging naar een proportionele afspiegeling van de populatie

Hiervoor concludeerden we dat de verdeling van de leerlingen in de normgroepen bij sommige achtergrondkenmerken afwijkt van de landelijke populatie van basisschoolleerlingen. Een geëigend middel om een niet volledig representatieve responsgroep meer representatief te maken is weging naar een proportionele afspiegeling van de populatie. Weging heeft echter alleen zin als de achtergrondkenmerken aan de hand waarvan de representativiteit bepaald is ertoe doen (dit wil zeggen: variantie in het sociaal-emotioneel functioneren verklaren). Met behulp van univariate variantieanalyse is de grootte van het effect van de achtergrondkenmerken op de gemeten aspecten van het sociaal-emotioneel functioneren bepaald. Als maat voor de effectgrootte hanteren we *Eta squared* die te interpreteren is als een proportie verklaarde variantie.

Tabel 4.15 toont per schaal de laagst en hoogst waargenomen waarde van *Eta squared* en de gemiddelde waarde ervan. De gemiddelde proportie verklaarde variantie is met 0,01 zeer klein. Uit de eveneens zeer kleine standaarddeviaties maken we op dat de eta's maar weinig variëren tussen de 26 schalen. Anders dan in het cognitieve domein blijken de kenmerken op basis waarvan de representativiteit bepaald is niet of

nauwelijks van belang voor het sociaal-emotionele functioneren van de leerlingen. Dit betekent tegelijkertijd dat weging naar een proportionele afspiegeling geen of hooguit een zeer klein effect zal hebben op de schattingen van het sociaal-emotioneel functioneren van de leerlingen (en de normering). Om deze reden hebben wij weging dan ook achterwege gelaten.

Tabel 4.15 *Laagst (min) en hoogst (max) waargenomen proportie verklaarde variantie (Eta squared) voor het effect van de achtergrondkenmerken op het sociaal-emotioneel functioneren.*

Achtergrondkenmerken	Aantal schalen	Eta squared			
		Min	Max	Gem	Stddev
Regio	26	0,00	0,04	0,01	0,01
Verstedelijking	26	0,00	0,05	0,02	0,02
Verstedelijking: stad versus platteland	26	0,00	0,04	0,01	0,01
Leerlinggewicht	26	0,00	0,06	0,01	0,02
Schoolgrootte	26	0,00	0,06	0,02	0,01
Denominatie	26	0,00	0,07	0,01	0,02

4.4 De normering

De gegevens voor de *normering* zijn verzameld via papieren en digitale afnames bij leerlingen in verschillende jaargroepen. Zoals uiteengezet in paragraaf 4.2 mogen de via verschillende afnamemodi en bij verschillende jaargroepen verzamelde gegevens als gelijkwaardig worden beschouwd. In het normeringsonderzoek zijn bij sommige aspecten middelgrote verschillen tussen de scores van jongens en meisjes geconstateerd. Deze sekseverschillen deden zich met name voor bij de schalen Pro-sociaal gedrag, Pestgedrag, Gepest worden en Storend gedrag uit het domein Sociaal gedrag (zie paragraaf 6.4). Het betreft hier reële verschillen in sociaal-emotioneel functioneren die ontwikkelingspsychologisch goed interpreteerbaar zijn en die ook in eerder onderzoek bevestigd zijn (o.a. Kuhlemeier, Van Boxel & Van Til, 2012). Desalniettemin is besloten geen afzonderlijke normen voor jongens en meisjes te presenteren. Een van de redenen is dat afzonderlijke normering voor jongens en meisjes een vertekend beeld van de reële bestaande verschillen tussen leerlingen zou geven. Daarnaast sluiten uniforme normen voor jongens en meisjes aan bij het beleid ten aanzien van andere vragenlijsten en toetsen uit het Cito Volgstelsel waar VISEON deel van uitmaakt. Ook is besloten is om geen statistische weging toe te passen. Zoals we eerder constateerden, hebben de steekproefkenmerken op basis waarvan de representativiteit is vastgesteld geen noemenswaardig effect op de scores gehad. Daardoor is het niet zinvol om de responsgroep op één of meer steekproefkenmerken terug te wegen naar een volledig proportionele afspiegeling van de landelijke verdeling. Om al deze redenen is besloten tot het hanteren van één ongewogen normgroep voor allen, ongeacht de afnamemodus, de jaargroep of het geslacht van de leerling. Voor de gebruiker heeft dit als voordeel dat de scoring en interpretatie betrekkelijk eenvoudig kunnen plaatsvinden. In het vervolg van deze paragraaf beschrijven we hoe de normering precies heeft plaatsgevonden.

In de rapportage voor de gebruikers van VISEON 2.0 zijn de ruwe somscores van de leerlingen per schaal omgezet naar percentielscores. Bij de zogenoemde 'positieve' schalen – de schalen waarbij een hogere score een gunstigere uitslag indiceert – zijn de percentielscores afgetrokken van 100. In de rapportage wijst een hogere percentielscore daardoor bij alle schalen op een minder gunstig sociaal-emotioneel functioneren.

Op basis van hun percentielscores zijn de leerlingen van de normgroep vervolgens per schaal verdeeld over drie zogeheten signalen: niet afwijkend (N), afwijkend (A) en zeer afwijkend (Z). Een signaal geeft aan of en in welke mate de score van een leerling afwijkt van leerlingen in de landelijke normgroep.

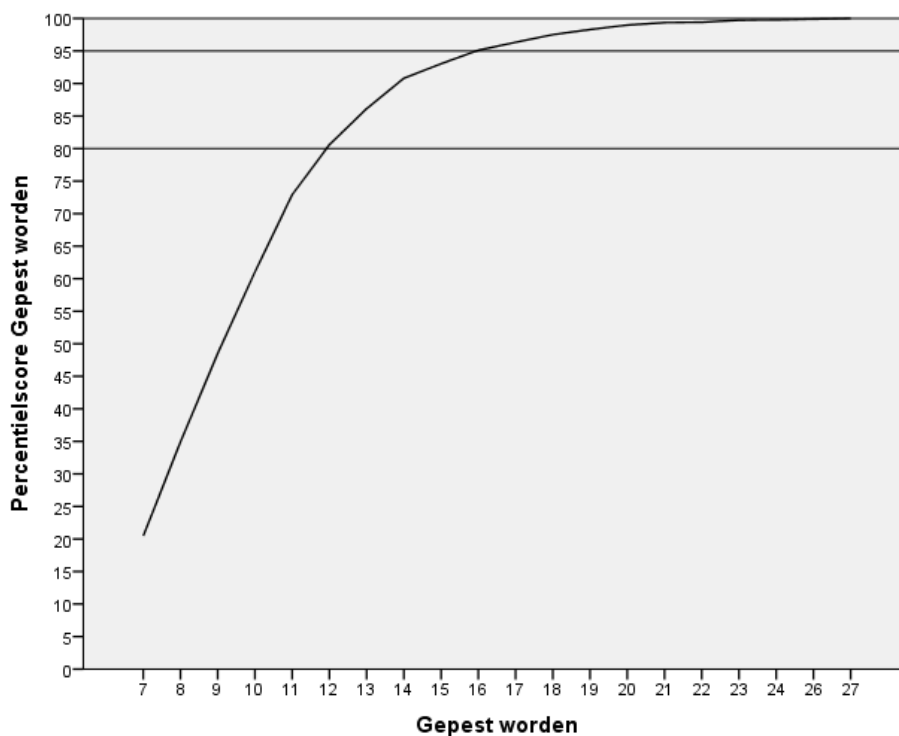
In de gebruikershandleiding (Cito, 2015) zijn deze signalen als volgt omschreven:

Signaal	Interpretatie
N	De score van de leerling wijkt niet af van de meeste andere leerlingen in de normgroep. 80% van de leerlingen behaalt de score N.
A	De score van de leerling wijkt af van de score van de meeste andere leerlingen in de normgroep. 15% van de leerlingen behaalt de score A.
Z	De score van de leerling wijkt zeer af van de score van de meeste andere leerlingen in de normgroep. 5% van de leerlingen behaalt de score Z.

Idealiter wordt er zo genormeerd dat precies 80% van de leerlingen signaal N ontvangt (dat wil zeggen de 80% leerlingen met de 'gunstigste' scores), precies 5% het signaal A (dat wil zeggen dat 5% met de 'minst gunstige' scores) en precies 15% het signaal A (dat wil zeggen de 15% tussen de signalen N en Z). In de praktijk laat de verdeling van de scores een dergelijke precisie niet altijd toe. De cumulatieve verdeling van de opeenvolgende schaalscores maakt soms 'sprongen', bijvoorbeeld als 78% van de leerlingen een score van 14 of lager heeft en 82% een score van 15 of lager. Daardoor moeten we genoegen nemen met kleine afwijkingen van de ideale verdeling. We hebben geprobeerd de afkappunten op de ruwe somscoreverdeling tussen N en A en tussen A en Z zo te kiezen dat de ideale verdeling van 80% in categorie N, 15% in categorie A en 5% in categorie Z zo dicht mogelijk benaderd wordt. Als de gegevens meerdere ongeveer gelijkwaardige verdelingen toelieten, is prioriteit toegekend aan het zo dicht mogelijk benaderen van de 5% in categorie Z boven de 80% in N en de 15% in A. Een nadere toelichting van de gevolgde normeringsprocedure aan de hand van een voorbeeld staat in het volgende kader.

Toelichting op de normeringsprocedure

Hoe de afkappunten precies zijn vastgesteld, illustreren we aan de hand van figuur 4.1 waarin de ruwe scores van de schaal Gepest worden zijn afgezet tegen de percentielscores. De schaal Gepest worden is een voorbeeld van een zogenoemde 'negatieve' schaal waarbij een hogere score een minder gunstig niveau van sociaal-emotioneel functioneren indiceert. De schaal kent een theoretisch bereik van 7 (geen pestgedrag) tot 28 (relatief veel pestgedrag). Figuur 4.1 laat onder meer zien dat 20% van de leerlingen de laagst haalbare score 7 heeft behaald en derhalve naar eigen zeggen op geen enkele manier heeft met pestgedrag door anderen. De percentielpunten P80 en P95 zijn in figuur 4.1 weergegeven met de beide horizontale lijnen. De percentielscores P80 en P95 komen op de ruwe scoreschaal overeen met de score 12 voor de grens tussen N en A en de score 16 voor de grens tussen A en Z. Overeenkomstig deze afkappunten blijkt na afronding 81% van de leerlingen het signaal N te krijgen, 15% het signaal A en 5% het signaal Zeer afwijkend. Deze verdeling komt zeer dicht in de buurt van de wenselijke verdeling 80%, 15% en 5%. Bij sommige andere schalen resulteren de percentielscores P80 en P95 in verdelingen die enigszins afwijken van de nagestreefde verdeling. Daar zijn de afkappunten zo gekozen dat de ideale verdeling van 80%, 15% en 5% zo dicht mogelijk benaderd wordt. Als er meerdere mogelijkheden waren, is daarbij prioriteit gegeven aan de 5% met signaal Z boven de 80% en 15% met respectievelijk signaal N en A.



Figuur 4.1 Cumulatieve verdeling van de ruwe scores in procenten voor de schaal Gepest worden

Nagestreefde versus gerealiseerde verdeling over de signalen N, A en Z

Tabel 4.16 toont de in werkelijkheid gerealiseerde percentages leerlingen in categorie N, A en Z voor alle 26 schalen van VISEON 2.0. Te zien is dat de gerealiseerde verdeling meestal nauwelijks afwijkt van de nagestreefde verdeling. Dat de ideale verdeling niet in alle gevallen volledig gerealiseerd kon worden, blijkt bijvoorbeeld uit de schaal Sociaal zelfbeeld waar 77% van de leerlingen het signaal Niet afwijkend kreeg, 17% het signaal Afwijkend en 6% het signaal Zeer afwijkend.

Tabel 4.16 Verdeling van de leerlingen in de normgroepen over de signalen N, A en Z (percentages)

	Signaal		
	Niet afwijkend	Afwijkend	Zeer afwijkend
LK basis			
pro-sociaal gedrag	81	14	5
pestgedrag	79	15	6
storend gedrag	80	15	5
teruggetrokken gedrag	81	13	6
taakgericht gedrag	78	15	7
motivatie voor leren	81	14	5
zelfstandigheid	82	13	5
hyperactief/impulsief	80	14	6
LL basis			
sociaal zelfbeeld	77	18	6
cognitief zelfbeeld	84	12	5
schoolbeleving	83	12	5
gepest worden	81	15	5
motivatie voor leren	78	16	6
sociale leeromgeving	82	14	4
onveilige leeromgeving	77	18	5
onordelijke leeromgeving	80	15	6
LL aanvullend			
pro-sociaal gedrag	76	18	6
pestgedrag	76	17	7
storend gedrag	81	14	5
taakgericht gedrag	80	14	6
motivatie voor lezen	79	14	7
motivatie voor rekenen	79	14	7
zelfbeeld lezen	81	14	5
zelfbeeld rekenen	78	14	7
spreekangst	78	18	5
toetsangst	79	16	5

Uit tabel 4.17 valt af te lezen met welke ruwe somscores de signalen N, A en Z in de rapportage voor de gebruiker overeenkomen.

De geldigheid van de normen

Ten aanzien van de geldigheid van de in 2015 verzamelde normen worden de criteria aangehouden die de COTAN vermeldt in zijn beoordelingssysteem (Evers, Lucassen, Meijer & Sijtsma, 2010, p. 21). Dit betekent dat de normen ten minste tot 2030 geldig zijn. Gezien de meetpretentie van het instrument is immers niet aan te nemen dat er sprake zal zijn van abnormale slijtage. Niettemin zal Cito op basis van dataretour periodiek nagaan of er sprake is van verschuivingen in de normen. Mocht daarvan onverhoopt eerder sprake zijn, dan zal de normering tijdig worden aangepast.

Tabel 4.17 De omzetting van ruwe scores naar signalen voor 'positieve' en 'negatieve' schalen

Schaal	Richting	Signaal		
		Niet afwijkend	Afwijkend	Zeer afwijkend
LK basis				
pro-sociaal gedrag	positief	32 - 21	20 - 17	16 - 8
pestgedrag	negatief	6 - 8	9 - 11	12 - 24
storend gedrag	negatief	6 - 12	13 - 16	17 - 24
teruggetrokken gedrag	negatief	7 - 16	17 - 19	20 - 28
taakgericht gedrag	positief	28 - 17	16 - 14	13 - 7
motivatie voor leren	positief	28 - 19	18 - 15	14 - 7
zelfstandigheid	positief	28 - 16	15 - 14	13 - 7
hyperactief/impulsief	negatief	7 - 15	16 - 19	20 - 28
LL basis				
sociaal zelfbeeld	positief	24 - 19	18 - 16	15 - 6
cognitief zelfbeeld	positief	28 - 21	20 - 18	17 - 7
schoolbeleving	positief	28 - 22	21 - 19	18 - 7
gepest worden	negatief	7 - 12	13 - 16	17 - 28
motivatie voor leren	positief	28 - 20	19 - 17	16 - 7
sociale leeromgeving	positief	28 - 20	19 - 17	16 - 7
onveilige leeromgeving	negatief	7 - 14	15 - 19	20 - 28
onordelijke leeromgeving	negatief	7 - 16	17 - 20	21 - 28
LL aanvullend				
pro-sociaal gedrag	positief	28 - 20	19 - 17	16 - 7
pestgedrag	negatief	7 - 10	11 - 13	14 - 28
storend gedrag	negatief	7 - 13	14 - 16	17 - 28
taakgericht gedrag	positief	28 - 20	19 - 17	16 - 7
motivatie voor lezen	positief	20 - 11	10 - 8	7 - 5
motivatie voor rekenen	positief	20 - 10	9 - 6	5
zelfbeeld lezen	positief	20 - 13	12 - 10	9 - 5
zelfbeeld rekenen	positief	20 - 12	11 - 9	8 - 5
spreekangst	negatief	8 - 19	20 - 25	26 - 32
toetsangst	negatief	8 - 18	19 - 28	29 - 32

5 Betrouwbaarheid en meetnauwkeurigheid

WISEON 2.0 is niet bedoeld om er belangrijke beslissingen mee te onderbouwen. Het doel reikt niet verder dan het in kaart brengen en volgen van gedrag en functioneren. In dit hoofdstuk gaan we na in hoeverre de vastgestelde betrouwbaarheid voldoet aan de eisen die het COTAN Beoordelingssysteem hieraan stelt (Evers, Lucassen, Meijer & Sijtsma, 2010). Daarbij bespreken we drie aspecten van de betrouwbaarheid en meetnauwkeurigheid:

- interne consistentie en meetnauwkeurigheid;
- interbeoordelaarsbetrouwbaarheid;
- test-hertestbetrouwbaarheid.

WISEON 2.0 bestaat uit 26 schalen die ieder bestaan uit vijf à acht uitspraken over het sociaal-emotioneel functioneren van de leerlingen. In paragraaf 5.1 bespreken we de item-restcorrelaties (Rir) en de homogeniteitscoëfficiënten die de mate van interne consistentie van de schalen tot uitdrukking brengen. Daarbij bespreken ook twee aspecten van de meetnauwkeurigheid, te weten de 95%-betrouwbaarheidsintervallen rond de schatting van de interne consistentie en de standaardmeetfout.

De vragenlijst voor leerkrachten is een observatielijst en geen objectief scorebare (leervorderingen)toets. Het is niet onbelangrijk te weten in hoeverre subjectiviteit in de observatie en beoordeling van kinderen een rol speelt. Daarom schenken we in paragraaf 5.2 aandacht aan de interbeoordelaarsbetrouwbaarheid. We vragen ons daarbij af in hoeverre beoordelaars onafhankelijk van elkaar tot vergelijkbare oordelen komen.

WISEON 2.0 is niet bedoeld voor voorspellingen over ontwikkeling door de tijd heen en we verwachten evenmin dat het construct zoals gemeten in WISEON 2.0 gerelateerd is aan leeftijd (zie hoofdstuk 2). Wel kunnen er zich bij individuele leerlingen veranderingen in het sociaal-emotionele functioneren voordoen die mogelijk extra aandacht rechtvaardigen. De COTAN adviseert het beoogd testgebruik mee te laten spelen bij de keuze van de lengte van het test-hertestinterval (Evers, Lucassen, Meijer & Sijtsma, 2010). In de handleiding (Cito, 2015) wordt gebruikers geadviseerd WISEON 2.0 in een schooljaar twee keer af te nemen: in november en in april. In paragraaf 5.3 besteden we daarom aandacht aan de test-her-test-betrouwbaarheid over de periode november-april. Vanwege de lengte van het interval geven de berekende waarden niet alleen een indicatie van de meetfouten, maar ook van veranderingen van de ware scores (stabiliteit), dit wil zeggen 'echte' veranderingen in het sociaal-emotioneel functioneren van leerlingen ten gevolge van gebeurtenissen in de genoemde periode. Helaas laat ons onderzoeksontwerp met slechts twee metingen het niet toe om de invloed van meetfouten en werkelijke veranderingen van elkaar te scheiden. Daarvoor zijn, zoals Coleman (1968) aantoonde, ten minste drie metingen nodig. De hier gepresenteerde waarden van de test-hertestbetrouwbaarheid zijn daarmee waarschijnlijk eerder te interpreteren in termen van stabiliteit dan van meet(on)nauwkeurigheid (Heise, 1969).

5.1 Interne consistentie, betrouwbaarheidsintervallen en standaardmeetfout

De betrouwbaarheid per schaal is bepaald aan de hand van Cronbachs alfa. Tabel 5.1 toont per schaal de laagst, hoogst en gemiddeld waargenomen item-restcorrelatie (Rir), Cronbachs alfa met het bijbehorende 95%-betrouwbaarheidsinterval en de standaardmeetfout. Deze gegevens zijn berekend op basis van de volledige normeringsgroep (zie hoofdstuk 4).

Tabel 5.1 *Item-restcorrelaties (Rir), interne consistentie (Cronbachs alfa) met bijbehorend 95%-betrouwbaarheidsinterval en standaardmeetfout*

	Item-restcorrelaties			Cronbachs alfa met 95%-betrouwbaarheidsinterval			Standaardmeetfout met 90%- en 95%-betrouwbaarheidsinterval		
	Min	Max	Gem	Alfa	Onder	Boven	Sem	90%	95%
LK basis									
pro-sociaal gedrag	0,69	0,80	0,75	0,93	0,92	0,93	1,18	±1,94	±2,31
pestgedrag	0,61	0,78	0,71	0,88	0,87	0,89	0,80	±1,32	±1,58
storend gedrag	0,59	0,76	0,66	0,87	0,85	0,88	1,29	±2,12	±2,53
teruggetrokken gedrag	0,57	0,77	0,69	0,89	0,88	0,90	1,35	±2,22	±2,65
taakgericht gedrag	0,68	0,81	0,76	0,92	0,91	0,93	1,26	±2,08	±2,48
motivatie voor leren	0,60	0,86	0,75	0,92	0,91	0,92	1,18	±1,94	±2,31
zelfstandigheid	0,64	0,80	0,74	0,92	0,91	0,92	1,19	±1,96	±2,33
hyperactief/impulsief	0,76	0,87	0,82	0,95	0,94	0,95	1,18	±1,94	±2,31
LL basis									
sociaal zelfbeeld	0,43	0,67	0,58	0,81	0,80	0,83	1,21	±1,99	±2,38
cognitief zelfbeeld	0,50	0,65	0,57	0,82	0,81	0,84	1,30	±2,14	±2,54
schoolbeleving	0,49	0,68	0,59	0,84	0,82	0,85	1,36	±2,23	±2,66
gepest worden	0,49	0,63	0,57	0,82	0,81	0,83	1,34	±2,20	±2,62
motivatie voor leren	0,51	0,66	0,59	0,84	0,82	0,85	1,34	±2,20	±2,63
sociale leeromgeving	0,47	0,70	0,59	0,84	0,82	0,85	1,30	±2,14	±2,55
onveilige leeromgeving	0,49	0,69	0,60	0,84	0,83	0,85	1,49	±2,45	±2,92
onordelijke leeromgeving	0,47	0,61	0,54	0,81	0,80	0,82	1,60	±2,64	±3,14
LL aanvullend									
pro-sociaal gedrag	0,41	0,61	0,53	0,80	0,77	0,82	1,44	±2,36	±2,81
pestgedrag	0,40	0,68	0,58	0,82	0,80	0,85	1,01	±1,65	±1,97
storend gedrag	0,24	0,72	0,58	0,81	0,79	0,83	1,49	±2,45	±2,92
taakgericht gedrag	0,43	0,60	0,55	0,81	0,78	0,83	1,44	±2,37	±2,83
motivatie voor lezen	0,59	0,80	0,74	0,89	0,88	0,91	1,37	±2,26	±2,69
motivatie voor rekenen	0,66	0,89	0,81	0,93	0,92	0,94	1,16	±1,90	±2,27
zelfbeeld lezen	0,48	0,69	0,61	0,81	0,79	0,84	1,41	±2,32	±2,77
zelfbeeld rekenen	0,64	0,79	0,71	0,88	0,86	0,89	1,27	±2,09	±2,49
spreekangst	0,56	0,72	0,63	0,87	0,86	0,89	1,93	±3,18	±3,79
toetsangst	0,66	0,77	0,73	0,92	0,91	0,93	1,76	±2,90	±3,46

De gemiddelde Rir over alle 174 uitspraken van de 26 schalen bedraagt 0,65 wat als hoog aangemerkt kan worden. De standaarddeviatie van de Rir's is met 0,11 klein hetgeen betekent dat de Rir's weinig van uitspraak tot uitspraak verschillen.

Geen van de gerapporteerde alfa's is lager dan 0,80, terwijl de 95%-betrouwbaarheidsintervallen rond de schatting van alfa zeer smal zijn.

De gepresenteerde standaardmeetfouten (Sem) zijn zonder uitzondering zeer klein. Het bereik van plusminus (bijna) twee standaardmeetfouten rond de somscore geeft aan waar de ware score zich met 95% zekerheid zal bevinden. Zo ligt de ware somscore voor pro-sociaal gedrag uit LK basis met 95% waarschijnlijkheid tussen 22,52 en 25,66.

De conclusie is dat de betrouwbaarheid van alle 26 schalen als goed beoordeeld kan worden.

5.2 Interbeoordelaarsbetrouwbaarheid

Leerkrachten met een duobaan zijn beiden goed bekend met het sociaal-emotioneel functioneren van de leerlingen in hun groep en daardoor bij uitstek geschikt als informant in onderzoek naar de interbeoordelaarsbetrouwbaarheid. Ter bepaling van de interbeoordelaarsbetrouwbaarheid zijn in totaal 290

leerlingen door twee duo-leerkrachten beoordeeld. Dit onderzoek is niet in het kader van het normeringsonderzoek uitgevoerd, maar als een afzonderlijke deelstudie. Er waren in totaal zestien beoordelingsparen van zeven scholen in het onderzoek betrokken. Van de 290 leerlingen kwam 12% uit groep 3, 35% uit groep 4, 13% uit groep 6, 18% uit groep 7 en 22% uit groep 8. Het onderzoek is uitgevoerd in de maand oktober van 2015, dus na de zomervakantie na het normeringsonderzoek. De leerkrachten is uitdrukkelijk verzocht de vragen over het sociaal-emotioneel functioneren van de leerling geheel onafhankelijk van elkaar te beantwoorden.

De vergelijking tussen de oordelen van de duo-leerkrachten baseren we op drie gegevens:

- de correlatie tussen de somscores van beide leerkrachten;
- de percentages overeenstemming per vraag per antwoordmogelijkheid;
- *Cohens Kappa*.

De correlatie geeft aan in hoeverre de rangordering van de leerlingen op basis van het oordeel van de ene leerkracht afwijkt van de rangordering van de andere leerkracht. Een vergelijkbare rangordering is nog geen garantie voor een grote overeenkomst tussen de scores. De ene leerkracht kan bijvoorbeeld systematisch hogere scores toekennen dan de ander. Daarom rapporteren we ook het percentage overeenstemming per uitspraak. Dat is een eenvoudige maat voor de mate waarin de per uitspraak toegekende scores van de beide leerkrachten overeenkomen. De leerkracht kan bij elke vraag uit telkens vier antwoordmogelijkheden kiezen. We rapporteren niet alleen het percentage exacte overeenstemming, maar ook de percentages scores die 1, 2 of 3 punten afwijken van de score van de andere duo-leerkracht. Het percentage exacte overeenstemming houdt er geen rekening mee dat de scores van beide leerkrachten alleen op basis van toeval al tot op zekere hoogte met elkaar zullen overeenstemmen. Denk bijvoorbeeld aan een vier-keuzeopgave waarbij de kans dat een examenkandidaat het juiste alternatief kiest louter op basis van toeval al een kwart is. Een bekende maat die corrigeert voor toevallige overeenstemming is *Kappa* (Cohen, 1960). Deze maat is 1 als de scores perfect overeenstemmen en 0 als de overeenstemming niet groter is dan men op basis van toeval mag verwachten. Landis en Koch (1997) geven de volgende vuistregel voor het interpreteren van de hoogte van *Kappa*: 0,00 tot 0,20 is *slight*, 0,21 tot 0,40 is *fair*, 0,41 tot 0,60 is *moderate*, 0,61 tot 0,80 is *substantial* en 0,81 tot 1,00 is *almost perfect*.

Tabel 5.2 toont de correlaties tussen de somscores van de beide leerkrachten voor de acht leerkrachtschalen van VISEON 2.0, de percentages overeenstemming per antwoordmogelijkheid en *Cohens Kappa*.

Tabel 5.2 De samenhang tussen de oordelen van duo-leerkrachten over het sociaal-emotioneel functioneren van hun leerlingen

Schaal	Percentage overeenstemming					Kappa
	<i>r</i>	Exact	±1	±2	±3	
pro-sociaal gedrag	0,71	71%	27%	1%	0%	0,48
pestgedrag	0,81	88%	11%	1%	0%	0,60
storend gedrag	0,68	76%	22%	2%	0%	0,67
teruggetrokken gedrag	0,65	67%	30%	3%	0%	0,43
taakgericht gedrag	0,76	68%	30%	2%	0%	0,47
motivatie voor leren	0,53	69%	29%	2%	0%	0,55
zelfstandigheid	0,66	66%	31%	3%	0%	0,43
hyperactief/impulsief	0,81	75%	23%	2%	0%	0,58

De correlaties tussen de somscores van de beide leerkrachten variëren van 0,53 voor Motivatie voor leren tot 0,81 voor Pestgedrag en Hyperactief/impulsief. Over het geheel genomen zijn de correlaties tussen de beoordelingsparen beduidend lager dan bij de vorige versie van VISEON (Kamphuis e.a., 2004; $0,84 < r <$

0,92) en de versie voor kleuters (Van Boxtel, Hop & Hemker, 2014; r bedraagt 0,70, 0,75 en 0,75) maar ongeveer even hoog of hoger dan bij andere instrumenten met een vergelijkbare meetpretentie. Zo variëren de gerapporteerde correlaties tussen de beoordelingsparen bij Zien! (Broer, Haverhals & De Bruin, 2012) van 0,40 tot 0,80 met een gemiddelde van gemiddeld 0,61 en bij de SDQ (Diepenmaat, Van Eijnsden, Janssens, Loomans & Stone, 2014) van 0,20 tot 0,77 met een gemiddelde van 0,50.

Extreem grote verschillen tussen de somscores van twee leerkrachten kunnen wijzen op fouten in de data. Daarvan kan bijvoorbeeld sprake zijn als de scores van twee duo-leerkrachten per ongeluk niet op dezelfde leerling betrekking zouden hebben. Gezien de duidelijke aanwijzingen voor de toewijzing van leerkrachtlijsten aan leerlingen en het gegeven dat de school, de jaargroep, de voornaam, de achternaam en de geboortedatum van de leerling in het bestand voor beide duo-leerkrachten volledig identiek waren, achten wij deze alternatieve verklaring zeer onwaarschijnlijk. Desalniettemin hebben we geprobeerd het effect van extreme, potentieel onjuiste scores op de correlaties te kwantificeren. Daarbij hebben wij twee criteria voor een extreem groot verschil gehanteerd, te weten maximaal 30% en 25% van de maximumscore. Omwille van de vergelijkbaarheid over schalen zijn de ruwe somscores omgezet naar een percentage van de maximumscore. Daardoor betekent het 25%-criterium bijvoorbeeld dat de antwoorden van de beide leerkrachten gemiddeld één punt per uitspraak van elkaar verschillen. Overeenkomstig het 30%- en 25%-criterium zijn achtereenvolgens respectievelijk 1% en 2% van de in totaal 2182 naar percentages omgezette somscores uit de dataset verwijderd. Tabel 5.3 toont de correlaties voor de volledige dataset en na verwijdering van potentieel onjuiste scores volgens onze twee criteria.

Tabel 5.3 Samenhang tussen duo-leerkrachten op basis van alle gegevens en na verwijdering van de 1% en 2% meest extreme scores volgens het 30%- en 25%-criterium (correlaties)

	Alle scores	Verschil in percentage van de maximumscore	
		30%	25%
pro-sociaal gedrag	0,71	0,73	0,75
pestgedrag	0,81	0,83	0,83
storend gedrag	0,68	0,68	0,70
teruggetrokken gedrag	0,65	0,70	0,73
taakgericht gedrag	0,76	0,78	0,81
motivatie voor leren	0,53	0,53	0,53
zelfstandigheid	0,66	0,71	0,75
hyperactief/impulsief	0,81	0,83	0,84

Verwijdering van de 1% en 2% meest extreme afwijkingen (met als criteria een verschil van respectievelijk 30% en 25% van de maximumscore) leidt over het algemeen tot een minieme verhoging van de samenhang tussen de somscores van de twee leerkrachten. De lage correlatie bij de schaal Motivatie voor leren blijft zelfs onveranderd 0,53. Deze bevindingen vormen een extra ondersteuning voor de integriteit van de verzamelde gegevens.

Een grote overeenstemming in rangorde is geen garantie voor een grote overeenkomst in toegekende scores. De ene duo-leerkracht kan systematisch hogere of lagere scores toekennen dan de ander zonder dat dit in de correlatie tot uitdrukking komt. Tabel 5.2 bevat daarom ook informatie over de grootte van de verschillen tussen de beide beoordelende leerkrachten in de vorm van het percentage overeenstemming per uitspraak. De percentages exacte overeenstemming liggen in het bereik van 66% voor Zelfstandigheid tot 88% voor Pestgedrag. De antwoorden van de ene leerkracht blijken zelden meer dan één punt af te wijken van die van de andere leerkracht. *Cohens Kappa* – een overeenstemmingsmaat die corrigeert voor toevallige overeenstemming – varieert van 0,43 tot 0,67 wat volgens de vuistregel van Landis en Koch (1997) als *moderate* tot *substantial* geïnterpreteerd mag worden.

Conclusie

De bevinding dat de antwoorden van de duo-leerkrachten zelden meer dan één scorepunt van elkaar verschillen geeft aanleiding tot vertrouwen in de objectiviteit van de beoordeling. De overeenstemming over de mate van pestgedrag is zeer hoog. Dit is opmerkelijk omdat pesten zich vaak buiten het gezichtsveld van de leerkracht afspeelt en men juist hier grotere verschillen tussen beoordelaars zou verwachten (Diepenmaat e.a., 2014). Ook over Hyperactiviteit/impulsiviteit zijn duo-leerkrachten het in hoge mate met elkaar eens. De wat minder hoge overeenstemming bij Teruggetrokken gedrag en Zelfstandigheid kan wellicht verklaard worden doordat internaliserende problematiek voor leerkrachten vaak minder zichtbaar is dan externaliserend gedrag.

De interbeoordelaarsbetrouwbaarheid van de meeste schalen is voldoende hoog, maar wel lager dan wij oorspronkelijk verwachtten. Een relatief lage betrouwbaarheid doet zich voor bij Motivatie voor leren, Teruggetrokken gedrag, Zelfstandigheid en Storend gedrag. Scoreverschillen tussen duo-leerkrachten zijn overigens niet altijd toe te schrijven aan een gebrek aan objectiviteit bij één of beiden (Kamphuis e.a., 2014). Het kan ook werkelijk zo zijn dat een leerling zich bij de ene leerkracht anders gedraagt dan bij de andere leerkracht. Zo zou de leerling zich afhankelijk van de leerkracht meer of minder gemotiveerd kunnen gedragen, terwijl hyperactiviteit/impulsiviteit meer een eigenschap zou kunnen zijn die zich ongeacht de leerkracht in alle lessen in ongeveer even sterke mate voordoet. In de gebruikershandleiding (Cito, 2015) adviseren wij duo-leerkrachten dan ook om de leerkrachtlijst voor een leerling ofwel gezamenlijk in te vullen, ofwel dat volkomen onafhankelijk van elkaar te doen. De ingevulde lijsten kunnen de leerkrachten daarna per leerling onderling vergelijken en bespreken. Daarbij is het belangrijk om te bespreken waarom zij bij bepaalde schalen of uitspraken verschillend oordelen. Bij duidelijke verschillen van mening kan men besluiten om het gedrag van de leerling (nogmaals) te observeren om op grond daarvan tot een definitief oordeel te komen.

5.3 Test-hertestbetrouwbaarheid

De samenhang tussen scores die op verschillende momenten bij dezelfde onderzoeksgroep verzameld worden, geeft inzicht in de test-hertestbetrouwbaarheid. Het interval tussen beide metingen varieert in de onderzoekspraktijk van enkele weken tot een half jaar (Goodman, 2001; Stone e.a., 2010; Diepenmaat e.a., 2014). De COTAN adviseert het beoogde testgebruik mee te laten spelen bij de keuze van de lengte van het test-hertestinterval (Evers, Lucassen, Meijer & Sijtsma, 2010). VISEON 2.0 is niet bedoeld voor voorspellingen door de tijd heen en we verwachten evenmin dat het gemeten construct gerelateerd is aan jaargroep of leeftijd. Wel pretendeert het instrumentarium leerkrachten in staat te stellen het sociaal-emotioneel functioneren van individuele leerlingen systematisch te volgen. In de handleiding (Cito, 2015) wordt gebruikers geadviseerd VISEON 2.0 in november en in april van een schooljaar af te nemen. De aanname is dat er zich tussen beide geadviseerde afnamemomenten systematische veranderingen in het sociaal-emotioneel functioneren van een leerling kunnen voordoen die mogelijk extra aandacht vereisen. In overeenstemming met dit beoogd gebruik hebben wij de test-hertestbetrouwbaarheid bepaald over de periode van november tot april. De gegevens van november 2014 zijn afkomstig van de tweede proefafname en die van april 2015 van het normeringsonderzoek. Van de 437 leerlingen die de leerkracht bij de tweede proefafname beoordeelde, werden er 255 ook in het normeringsonderzoek beoordeeld. Van de 504 leerlingen die aan de tweede proefafname deelnamen, vulden er 249 ook de leerlinglijst van het normeringsonderzoek in. Dit biedt de mogelijkheid om het sociaal-emotioneel functioneren van de leerlingen over de in de handleiding geadviseerde periode te vergelijken. De vergelijking tussen het sociaal-emotioneel functioneren in november en april baseren we op drie gegevens:

- de correlatie tussen de somscores per schaal in november en april;
- de percentages overeenstemming per vraag per antwoordmogelijkheid;
- *Cohens Kappa*.

De correlatie geeft aan in hoeverre de rangordering van de leerlingen op basis van het niveau van sociaal-emotioneel functioneren in november afwijkt van de rangordering in april. Het percentage overeenstemming per uitspraak is een eenvoudige maat voor de mate waarin de per uitspraak toegekende scores op beide

momenten overeenkomen. *Cohens Kappa* is een voor toevallige overeenstemming gecorrigeerde overeenstemmingsmaat (Cohen, 1960). Zie voor meer uitleg en de interpretatie van de hoogte van *Cohens Kappa* paragraaf 5.2.

Tabel 5.4 toont de correlaties tussen de somscores voor de 26 schalen van VISEON 2.0, de percentages overeenstemming per antwoordmogelijkheid en *Cohens Kappa*. Het ontbreken van sommige gegevens in tabel 5.4 vereist een nadere toelichting. Uit de leerkrachtschalen Pestgedrag en Storend gedrag zijn op basis van de tweede proefafname enkele uitspraken verwijderd. Daardoor konden de somscores en hiermee ook de correlatie tussen de somscores in de proefafname en het normeringsonderzoek niet voor de volledige schalen berekend worden. De percentages overeenstemming en *Cohens Kappa* zijn gebaseerd op de gemeenschappelijke uitspraken van Pestgedrag en Storend gedrag in beide onderzoeken.

Aan de leerlingschaal Cognitief zelfbeeld zijn naar aanleiding van de tweede proefafname twee uitspraken toegevoegd. Ook hier zijn de percentages overeenstemming en *Cohens Kappa* berekend op basis van de wel voor vergelijking beschikbare uitspraken.

Tabel 5.4 Samenhang tussen de in november en april verzamelde gegevens over het sociaal-emotioneel functioneren (correlaties)

	r	Exact	Percentage overeenstemming			Kappa
			±1	±2	±3	
LK basis						
pro-sociaal gedrag	0,62	62%	36%	1%	0%	0,29
pestgedrag	--	83%	17%	0%	0%	0,41
storend gedrag	--	59%	36%	4%	1%	0,34
teruggetrokken gedrag	0,59	56%	41%	3%	0%	0,28
taakgericht gedrag	0,67	57%	41%	2%	0%	0,30
motivatie voor leren	0,72	61%	36%	2%	0%	0,36
zelfstandigheid	0,64	57%	41%	2%	0%	0,30
hyperactief/impulsief	0,72	65%	34%	1%	0%	0,42
LL basis						
sociaal zelfbeeld	0,39	57%	36%	5%	2%	0,25
cognitief zelfbeeld	--	55%	39%	5%	1%	0,22
schoolbeleving	0,51	60%	30%	6%	4%	0,31
gepest worden	0,48	62%	33%	4%	1%	0,30
motivatie voor leren	0,50	51%	42%	6%	1%	0,21
sociale leeromgeving	0,36	53%	40%	5%	2%	0,22
onveilige leeromgeving	0,62	55%	36%	7%	2%	0,24
onordelijke leeromgeving	0,44	50%	38%	10%	2%	0,22
LL aanvullend						
pro-sociaal gedrag	0,59	56%	38%	5%	1%	0,26
pestgedrag	0,43	74%	24%	2%	0%	0,36
storend gedrag	0,53	61%	33%	4%	2%	0,30
taakgericht gedrag	0,60	57%	37%	5%	1%	0,29
motivatie voor lezen	0,57	53%	34%	9%	4%	0,33
motivatie voor rekenen	0,67	54%	36%	8%	3%	0,37
zelfbeeld lezen	0,70	56%	36%	6%	2%	0,35
zelfbeeld rekenen	0,69	53%	39%	6%	2%	0,33
spreekangst	0,57	49%	36%	11%	4%	0,27
toetsangst	0,61	45%	40%	11%	4%	0,22

--: geen correlatie beschikbaar

LK basis

De oordelen van de leerkrachten in november vertonen een sterke samenhang met de oordelen over dezelfde leerlingen vijf maanden later in mei. De correlaties variëren van 0,59 voor Teruggetrokken gedrag tot 0,72 voor Motivatie voor leren en Hyperactief/impulsief. De conclusie is dat leerkrachten over het algemeen zeer consistent zijn in hun oordeel over het sociaal-emotioneel functioneren. Dit blijkt ook uit de grote mate van overeenstemming tussen de antwoorden op de afzonderlijke vragen waarbij het antwoord in november zelden meer dan 1 punt verschilt van het antwoord in november. De voor toevallige overeenstemming gecorrigeerde overeenstemmingscoëfficiënt Kappa ligt tussen 0,28 en 0,42 wat volgens de vuistregel van Landis en Koch (1997) als *fair* of *moderate* geëvalueerd kan worden.

LL basis en LL aanvullend

Bij de leerlingvragenlijst is de samenhang tussen de scores in november en april door de bank genomen iets lager dan bij LK basis het geval is. De correlatie varieert hier van 0,36 voor Sociale leeromgeving tot 0,70 voor Zelfbeeld lezen. De overeenstemming op het niveau van de antwoorden op de individuele uitspraken is over het algemeen ook wat minder gunstig dan bij de leerkrachtlijst, al variëren de percentages ook hier sterk van schaal tot schaal. Zo lijkt Pestgedrag een stabiel aspect van het sociaal-emotioneel functioneren dat ook na vijf maanden nog helder in het geheugen van de leerling gegrift staat (74% volledig identieke antwoorden). Dit in tegenstelling tot bijvoorbeeld Spreekangst en Toetsangst die sterker van moment tot moment lijken te fluctueren (respectievelijk 49% en 47% volledig identieke antwoorden).

Discussie en conclusie

Ter bepaling van de invloed van meetfouten worden in onderzoek naar de test-hertestbetrouwbaarheid van cognitieve vaardigheidstoetsen over het algemeen intervallen van vier tot zes weken aanbevolen. Een zeer kort interval – tot enkele weken – wordt vanwege herinneringseffecten minder zinvol geacht (Evers, Lucassen, Meijer & Sijtsma, 2010). In ons onderzoek was het interval echter aanmerkelijk langer, namelijk ongeveer vijf maanden. Vanwege de lengte van het interval geven de berekende waarden niet alleen een indicatie van de meetfouten (betrouwbaarheid), maar ook van veranderingen van de ware scores (stabiliteit), dit wil zeggen 'echte' veranderingen in het sociaal-emotioneel functioneren van leerlingen ten gevolge van relevante gebeurtenissen die gedurende de vijf maanden kunnen plaatsvinden. Hierbij merken we op dat de temporele variabiliteit van sociaal-emotioneel functioneren waarschijnlijk groter is dan van cognitieve vaardigheden als rekenen en lezen (die normaliter bij kinderen een relatief gestage toename te zien zullen geven). Helaas laat ons onderzoeksontwerp met slechts twee metingen het niet toe om de invloed van meetfouten en werkelijke veranderingen van elkaar te scheiden (of de test-hertestcorrelatie te corrigeren voor 'reële' veranderingen in het sociaal-emotioneel functioneren). Daarvoor zijn, zoals Coleman (1968) aantoonde, ten minste drie metingen nodig. De hiervoor gepresenteerde waarden van de test-hertestbetrouwbaarheid zijn daarmee waarschijnlijk eerder op te vatten als indicatoren van stabiliteit dan van meetnauwkeurigheid (Heise, 1969).

De in november verzamelde scores vertonen een middelmatig tot sterke samenhang met de scores voor dezelfde leerlingen in april. De geconstateerde correlaties over een periode van vijf maanden zijn vergelijkbaar met de gemiddelde correlatie van 0,62 over de beoordelingen van leerkrachten, leerlingen en ouders met de SDQ met een interval van vier tot zes maanden (Goodman, 2001). Eveneens met de SDQ rapporteren Diepenhorst e.a. (2014) correlaties tussen metingen in het najaar en het voorjaar (variërend van 0,59 tot 0,77) met een gemiddelde van 0,68. Vliek, Riet en Weide (2012) onderzochten de stabiliteit van de Kanjervragenlijst over een periode van twee tot vier maanden en vonden correlaties in het bereik van 0,68 tot 0,78 (met een gemiddelde van 0,72). De gevonden stabiliteit is dus in lijn met uitkomsten van ander onderzoek in het sociaal-emotionele domein over een vergelijkbaar lange periode.

De geconstateerde stabiliteit is niet zo hoog dat we moeten uitgaan van sociaal-emotioneel functioneren als een stabiel en onveranderbaar persoonskenmerk. De geconstateerde middelmatige tot sterke samenhang geeft aanleiding tot de veronderstelling dat er zich gedurende vijf maanden veranderingen in het sociaal-emotionele functioneren van individuele leerlingen kunnen voordoen die mogelijk extra aandacht vereisen.

Tegelijkertijd is de samenhang niet zo laag dat er twijfel opkomt of er beide momenten wel dezelfde eigenschappen gemeten worden.

In test-hertestonderzoek met een interval van vier tot zes weken zijn de gevonden correlaties vrijwel nooit meer dan 0,10 lager dan de betrouwbaarheid in termen van interne consistentie. In ons onderzoek was de interne consistentie (Cronbachs alfa) van alle 26 schalen hoger dan 0,80 (zie paragraaf 5.1). De correlaties tussen beide metingen bleken echter regelmatig meer dan 0,10 lager dan de interne consistentie.

De gevonden test-hertestcorrelaties voldoen hiermee maar zelden aan de minimumeis voor een voldoende betrouwbare meting op groepsniveau of voor het nemen van minder belangrijke beslissingen op individueel niveau (Evers, Lucassen, Meijer & Sijtsma, 2010). Een plausibele verklaring is de uitzonderlijk lange periode tussen de eerste en tweede meting. Anders dan de interne consistentie, is de gevonden test-hertestcorrelatie geattenuëerd door 'echte' veranderingen in het sociaal-emotioneel functioneren die tijdens het lange interval zijn opgetreden. Onze resultaten geven hiermee een ondergrensschatting van de test-hertestbetrouwbaarheid in termen van meetnauwkeurigheid (Diepenmaat e.a., 2014; Goodman, 2001; Stone e.a., 2010;). Hadden wij met andere woorden voor een veel korter tijdsinterval gekozen, dan waren de correlaties zonder twijfel navenant hoger geweest (maar waren de uitkomsten voor de gebruiker minder relevant geweest). Een andere verklaring voor het als substantieel ervaren verschil tussen de interne consistenties en de test-hertestcorrelaties veronderstelt dat de temporele variabiliteit van sociaal-emotioneel functioneren groter is dan van cognitieve vaardigheden als rekenen en lezen.

6 Validiteit

In dit hoofdstuk doen we verslag van de uitkomsten van analyses die zijn uitgevoerd met het oog op de begripsvaliditeit van VISEON 2.0. Het gebruiksdoel van het instrument reikt, net als bij andere instrumenten in het Cito Volgsysteem primair en speciaal onderwijs, niet verder dan het beschrijven en volgen van leerlingen, in dit geval van hun sociaal-emotionele functioneren. Omdat VISEON 2.0 niet bedoeld is voor 'voorspellend gebruik' blijven voorspellende kwaliteiten en criteriumvaliditeit in dit hoofdstuk verder onbesproken.

Er is op verschillende manieren onderzoek gedaan naar de begripsvaliditeit. In paragraaf 6.1 doen we verslag van een analyse van de interne structuur van de meting van het sociaal-emotioneel functioneren. Daarbij gaan we eerst in op de vraag in hoeverre het gelukt is om voor de onderscheiden aspecten van het sociaal-emotioneel functioneren uni-dimensionele schalen te construeren (paragraaf 6.1.1). Vervolgens kijken we met behulp van exploratieve factoranalyse naar de factorzuiverheid ofwel de empirische onderscheidbaarheid van de veronderstelde aspecten van het sociaal-emotioneel functioneren (paragraaf 6.1.2). Daarna bespreken we de correlaties tussen de schalen zoals deze door de factoranalyses bevestigd werden (paragraaf 6.1.3). Tot slot gaan we na in hoeverre de gevonden interne structuur geldig is voor relevante groepen leerlingen (paragraaf 6.1.4).

In paragraaf 6.2 gaan we in op de vraag of de uitspraken op dezelfde manier functioneren voor verschillende groepen, te weten respondenten die de lijst schriftelijk dan wel digitaal invulden, leerlingen van verschillende jaargroepen en jongens en meisjes. Dit doen we door middel van DIF-analyses.

In paragraaf 6.3 rapporteren we de uitkomsten met betrekking tot de externe structuur. Die bestuderen we aan de hand van correlaties met (onderdelen van) andere instrumenten voor sociaal-emotioneel functioneren. Het betreft de 'oude' versie van VISEON (eerste generatie), de *Strength and Difficulties Questionnaire* (SDQ), de Kanjervragenlijst en de Schoolvragenlijst (SVL). Er is sprake van convergente validiteit als een aspect relatief hoog samenhangt met verwante aspecten en van discriminante validiteit als er sprake is van een ontbrekende of een relatief lage samenhang met niet verwante of minder verwante aspecten (Campbell & Fiske, 1959). Gegevens over de convergente en discriminante validiteit zijn ook van belang om te kunnen uitsluiten dat niet 'per ongeluk' een ander dan het bedoelde begrip gemeten wordt (Evers, Lucassen, Meijer & Sijtsma, 2010).

Tot slot gaan we in paragraaf 6.4 in op verschillen tussen groepen leerlingen. Eerst rapporteren we de uitkomsten van een analyse van de verschillen qua afnamemodus (schriftelijk versus digitaal) en jaargroep (oudere versus jongere leerlingen). Het onderzoek naar afnamemodi en jaargroep is van belang omdat we hier geen verschillen tussen de onderscheiden groepen verwachten. Vervolgens gaan we in op verschillen qua verstedelijking (stad versus platteland), leerlinggewicht en geslacht. Over de invloed van verstedelijking en leerlinggewicht op het sociaal-emotioneel functioneren is nog weinig bekend. Deze analyses zijn echter van belang omdat de gebruiker deze informatie kan gebruiken bij de interpretatie van de afnameresultaten. De analyse van verschillen tussen jongens en meisjes is voor de onderbouwing van de begripsvaliditeit relevant omdat we hier op basis van eerder onderzoek (o.a. Kuhlemeier, Van Boxtel & Van Til, 2012) hypothesen konden formuleren. Als we tegen de verwachting in geen verschillen tussen jongens en meisjes zouden vinden, is het minder aannemelijk dat VISEON 2.0 zoals bedoeld sociaal-emotioneel functioneren meet (vergelijk Evers, Lucassen, Meijer & Sijtsma, 2010).

6.1 Interne structuur

6.1.1 Unidimensionaliteit

Met het oog op een zo eenduidig mogelijke interpretatie van de scores door de gebruiker is tijdens de constructiefase veel aandacht besteed aan het ontwikkelen van betrouwbare en unidimensionele schalen. In hoofdstuk 3 en 5 hebben we laten zien dat de betrouwbaarheid van alle 26 schalen op basis van de interne consistentie (Cronbachs alfa) en de item-restcorrelaties goed te noemen is. Als extra check op de nagestreefde uni-dimensionaliteit zijn wij met behulp van exploratieve factoranalyse – principale componentenanalyse met varimaxrotatie – nagegaan in hoeverre de uitspraken inderdaad slechts één factor meten dan wel een beroep doen op meerdere factoren. Tabel 6.1 toont voor elke schaal het aantal uitspraken en het aantal gevonden factoren met het bijbehorende percentage verklaarde variantie.

Tabel 6.1 Resultaten van de exploratieve factoranalyses per schaal

	Aantal uitspraken	Aantal factoren	Percentage verklaarde variantie	
			Factor 1	Factor 2
LK basis				
pro-sociaal gedrag	8	1	66	
pestgedrag	6	1	64	
storend gedrag	6	2	60	18
teruggetrokken gedrag	7	1	61	
taakgericht gedrag	7	1	68	
motivatie voor leren	7	1	67	
zelfstandigheid	7	1	67	
hyperactief/impulsief	7	1	75	
LL basis				
sociaal zelfbeeld	6	1	52	
cognitief zelfbeeld	7	1	49	
schoolbeleving	7	1	49	
gepest worden	7	1	49	
motivatie voor leren	7	1	51	
sociale leeromgeving	7	1	52	
onveilige leeromgeving	7	1	52	
onordelijke leeromgeving	7	1	47	
LL aanvullend				
pro-sociaal gedrag	7	1	46	
pestgedrag	7	1	51	
storend gedrag	7	1	53	
taakgericht gedrag	7	2	47	15
motivatie voor lezen	5	1	71	
motivatie voor rekenen	5	1	78	
zelfbeeld lezen	5	1	58	
zelfbeeld rekenen	5	1	67	
spreekangst	8	1	54	
toetsangst	8	1	64	

De uitspraken blijken bij 24 van de 26 schalen slechts één factor te meten (zie tabel 6.1). De twee uitzonderingen zijn Storend gedrag uit LK basis en Taakgericht gedrag uit LL aanvullend die een beroep doen op twee factoren (waarbij de tweede factor duidelijk minder variantie naar zich toetrekt dan de eerste factor).

De ladingen van de uitspraken op de beide factoren van Storend gedrag zijn weergegeven in tabel 6.2. Het onderscheid tussen de beide factoren lijkt inhoudelijk moeilijk te interpreteren. Wellicht heeft de

gevonden twee-factorstructuur ermee te maken dat de beide op de tweede factor ladende uitspraken expliciet aan regels en regelhandhaving gerelateerd zijn en de andere drie wat minder uitdrukkelijk.

Tabel 6.2 Factorladingen van de uitspraken over Storend gedrag uit LK basis

	Factor	
	1	2
stoort de les	0,73	0,44
houdt zich aan de gedragsregels		0,90
doet wat de leerkracht zegt		0,91
houdt andere kinderen van hun werk	0,83	
bemoeit zich ongevraagd met het werk van andere kinderen	0,82	
kan tijdens de les zijn/haar mond niet houden	0,84	

De ladingen van de uitspraken van Taakgericht gedrag zijn weergegeven in tabel 6.3. Hier heeft het onderscheid te maken met een verschil tussen enerzijds aandachtig en zelf-gereguleerd werken en anderzijds nauwkeurig werken. Dit onderscheid werd overigens ook reeds gevonden tijdens de proefafname die aan het normeringsonderzoek vooraf ging (Kuhlemeier, 2015). Omdat nauwkeurig werken tot de kern van het begrip Taakgericht gedrag behoort, is destijds besloten om de twee-factorstructuur voor lief te nemen en de beide uitspraken over nauwkeurig werken te handhaven

Tabel 6.3 Factorladingen van de uitspraken over Taakgericht gedrag uit LL aanvullend

	Factor	
	1	2
Ik let goed op bij de lessen	0,72	
Als de juf of meester een opdracht geeft, begin ik meteen	0,73	
Ik werk slordig		0,92
Als we in groepjes werken, let ik goed op	0,70	
Ik werk netjes		0,81
Als andere kinderen kletsen, werk ik rustig door	0,71	
Als de leerkracht de klas uitloopt, werk ik rustig door	0,76	

6.1.2 Factorzuiverheid

Hiervoor hebben we vastgesteld dat 24 van de 26 schalen uni-dimensioneel zijn. Naast homogeniteit is tijdens de constructie ook veel aandacht besteed aan het verkrijgen van factor-zuivere schalen, dit wil zeggen schalen waarvan de inhoud zo weinig mogelijk overlapt met die van andere schalen. De meting van het sociaal-emotioneel functioneren geldt als factor-zuiver als de veronderstelde aspecten (schalen) in de gegevens terug te vinden zijn. In deze paragraaf onderzoeken we de factor-zuiverheid ofwel de empirische onderscheidbaarheid van de geconstrueerde schalen. Dat doen we met behulp van exploratieve factoranalyse: principale componentenanalyse met varimax rotatie, uitgevoerd op de gegevens van het normeringsonderzoek. We beschouwen een schaal als factor-zuiver als de ladingen op de factor die de schaal pretendeert te meten hoog zijn (d.w.z. $\geq 0,40$) en de ladingen op de 'eigen' factor hoger zijn dan die op andere factoren die de schaal niet beoogt te meten. Omdat het aantal uitspraken relatief groot is in verhouding tot het aantal respondenten, hebben we de factoranalyses afzonderlijk uitgevoerd voor LK basis, LL basis en LL aanvullend.

Factorzuiverheid van LK basis

De uitkomsten van een exploratieve factoranalyse op de uitspraken van zeven van de acht schalen van de leerkrachtlijst zijn weergegeven in tabel 6.4. Daarbij zijn factorladingen tussen -0,40 en 0,40 niet weergegeven. De uitspraken van de schaal Hyperactiviteit/impulsiviteit zijn in deze analyse niet meegenomen. Deze schaal was oorspronkelijk bedoeld voor het aanvullende instrumentarium en is daarom alleen in de digitale afname aan leerkrachten voorgelegd. Daardoor zijn er voor deze schaal veel minder waarnemingen beschikbaar dan voor de overige zeven schalen van LK basis (waarvoor zowel digitale als papieren gegevens beschikbaar zijn).

Tabel 6.4 Uitkomsten van een exploratieve factoranalyse van de uitspraken van LK basis

	Factor						
	1	2	3	4	5	6	7
Pro-sociaal gedrag							
kan goed dingen delen	0,67						
bemiddelt bij een ruzie tussen andere kinderen	0,71						
kan goed met andere kinderen samenwerken in groepjes	0,67						
kan andere kinderen goed aanvoelen	0,76						
troost andere kinderen als zij pijn of verdriet hebben	0,78						
kan ruzies met andere kinderen zelf oplossen	0,67						
kan zich aanpassen aan andere kinderen	0,75						
helpt andere kinderen	0,76						
Pestgedrag							
is onredelijk boos of driftig							0,63
maakt ruzie zonder aanleiding							0,74
treitert of pest andere kinderen							0,70
slaat of schopt andere kinderen							0,75
scheldt andere kinderen uit							0,78
bedreigt een ander kind							0,75
Storend gedrag							
stoort de les							0,75
houdt zich aan de gedragsregels							0,66
doet wat de leerkracht zegt							0,64
houdt andere kinderen van hun werk							0,74
bemoeit zich ongevraagd met het werk van andere kinderen							0,66
kan tijdens de les zijn/haar mond niet houden							0,70
Teruggetrokken gedrag							
heeft veel emotionele steun nodig van de leerkracht		0,77					
gedraagt zich te afhankelijk van de leerkracht		0,74					
heeft uitzonderlijk weinig zelfvertrouwen		0,76					
is snel in tranen		0,74					
maakt een bijzonder onzekere indruk		0,76					
is bang om contacten met andere kinderen aan te gaan		0,51					
maakt zich snel zorgen		0,78					
Taakgericht gedrag							
checkt bij een taak of hij/zij alles goed gedaan heeft					0,80		
controleert zelf of de taak volledig is uitgevoerd					0,80		
herstelt uit zichzelf fouten tijdens het uitvoeren van een taak					0,77		
werkt zorgvuldig					0,78		
kan de aandacht goed bij spel of werk houden			0,46		0,51	-0,41	
let goed op in de groep			0,49		0,44	-0,43	
werkt netjes					0,66		
Motivatie voor leren							
leest met plezier			0,64				
toont weinig interesse voor school			0,70				
rekent met plezier			0,66				
moet tot leren worden aangezet			0,68				
heeft meestal zin om taken uit te voeren			0,77				
is gemotiveerd om een taak af te ronden			0,70				
leert met plezier			0,80				
Zelfstandigheid							
weet precies wat hij/zij wil				0,56			
kan goed voor zichzelf opkomen		-0,47		0,57			
neemt het initiatief				0,79			
komt zelfverzekerd over		-0,52		0,68			
toont veel zelfvertrouwen		-0,53		0,66			
neemt bij groepswork anderen op sleeptouw				0,80			
voert de regie bij groepswork				0,83			

Op basis van de theoretische overwegingen die aan de constructie ten grondslag lagen (zie hoofdstuk 2) en de uitkomsten van de beide proefafnames is een gedwongen zeven-factorstructuur aan de gegevens opgelegd. De zeven factoren blijken in totaal 68% van de variantie in het sociaal-emotioneel functioneren te verklaren. De verklaarde variantie voor factor 1 tot en met 7 is achtereenvolgens 32%, 14%, 7%, 5%, 3%, 3% en 3%.

Het patroon van factorladingen blijkt vrijwel volledig overeen te komen met de beoogde structuur. Alle uitspraken laden hoger dan 0,40 op hun 'eigen' factor en vrijwel alle uitspraken laden hoger op hun 'eigen' factor dan op alle overige factoren. Een kleine smet op dit verder zo tevredenstellende beeld vormen enkele, overwegend kleine dubbelladingen. Twee daarvan betreffen uitspraken uit de schaal Taakgericht gedrag op de factor voor de Motivatie voor leren: 'Kan de aandacht goed bij spel of werk houden' en 'Let goed op in de groep'. Niet geheel duidelijk is waarom juist deze twee uitspraken het verder zo consistente beeld verstoren. Omdat de dubbelladingen niet hoog zijn en mogelijk aan toevallige steekproeffluctuaties te wijten zijn, is besloten de verdeling van de uitspraken over de schalen intact te laten.

Factorzuiverheid van LL basis

De resultaten van een voorbereidende factoranalyse van de uitspraken van de acht schalen van het basisgedeelte van de leerlinglijst (LK basis) gaven aanleiding om één uitspraak te verwijderen. Het betrof de uitspraak 'Ik voel me buitengesloten op deze school' uit de schaal Schoolbeleving. De reden voor verwijdering was dat de uitspraak even hoog op de 'eigen' factor laadde als op de factor voor Sociaal zelfbeeld (in beide gevallen 0,43).

De uitkomsten van een exploratieve factoranalyse op de resterende uitspraken van de acht schalen van LK basis zijn weergegeven in tabel 6.5. De acht factoren blijken in totaal 51% van de variantie in het sociaal-emotioneel functioneren te verklaren. De verklaarde variantie voor factor 1 tot en met 8 is achtereenvolgens 22%, 9%, 5%, 4%, 3%, 3%, 3% en 2%.

Het patroon van factorladingen blijkt vrijwel volledig overeen te komen met de structuur zoals beoogd tijdens de constructie. Alle factorladingen zijn hoger dan 0,40 en alle uitspraken laden hoger op hun 'eigen' factor dan op de overige factoren.

Tabel 6.5 *Uitkomsten van een exploratieve factoranalyse van de uitspraken van LL basis*

	Factor							
	1	2	3	4	5	6	7	8
Sociaal zelfbeeld								
Ik ben tevreden over het aantal vrienden/vriendinnen dat ik heb						0,64		
Ik ben tevreden over hoe ik met andere kinderen kan opschieten						0,56		
Ik ben tevreden over hoe aardig andere kinderen mij vinden						0,71		
Ik ben tevreden over hoe vaak ik met andere kinderen afspreek						0,51		
Ik ben tevreden over hoe leuk leeftijdsgenoten mij vinden						0,68		
Ik ben er tevreden over hoe graag kinderen mij mogen						0,71		
Cognitief zelfbeeld								
Ik ben meestal tevreden over mijn werk op school				0,69				
Ik ben blij met hoe ik het doe op school				0,73				
Dingen kunnen altijd beter, maar ik vind dat ik de dingen op school goed doe				0,67				
Ik ben gelukkig met mijn manier van werken op school				0,61				
Als ik mijn best wil doen op school, dan lukt me dat ook				0,52				
Ik vind mezelf een goede leerling				0,65				
Ik ben tevreden met hoe ik het doe tijdens de les				0,61				
Schoolbeleving								
Ik vind het leuk op deze school								0,73
Ik voel me ongelukkig op deze school								0,59
De sfeer bij ons op school is goed								0,58
Ik voel me thuis op deze school								0,69
Ik heb een hekel aan deze school								0,75
Ik hoor er op deze school niet echt bij								0,50
Ik voel me veilig op deze school								0,62
Gepest worden								
Andere kinderen lachen mij uit					0,46			
Een ander kind heeft mij uitgescholden					0,62			
Een ander kind maakt mij belachelijk					0,54			
Een ander kind heeft mij gepest					0,57			
Een ander kind heeft mij geschopt of geslagen (met opzet)					0,78			
Een ander kind heeft met mij gevochten (de ander is begonnen)					0,61			
Een ander kind heeft mij met opzet verwond (blauwe plek of erger)					0,70			
Motivatie voor leren								
Als ik leer, werk ik zo hard mogelijk	0,63							
Als ik aan iets begin, maak ik het af	0,61							
Als ik leer, doe ik mijn uiterste best	0,62							
Als ik leer, werk ik hard door, zelfs als de leerstof moeilijk is	0,73							
Ik ben een echte doorzetter, ook als het tegenzit	0,70							
Ik geef de moed niet op, ook al is de opdracht nog zo moeilijk	0,71							
Ik werk net zo lang aan een probleem tot ik het heb opgelost	0,62							
Sociale leeromgeving								
De kinderen in de groep zijn vrienden of vriendinnen van elkaar		0,62						
In mijn groep wordt niemand buitengesloten		0,44						
De kinderen in de groep gaan leuk met elkaar om		0,70						
De leerlingen in de groep zijn aardig voor elkaar		0,70						
De kinderen in de groep helpen elkaar		0,62						
De kinderen in de groep kunnen goed met elkaar opschieten		0,73						
De kinderen kunnen goed met elkaar spelen en samenwerken		0,69						
Onveilige leeromgeving								
Een kind roddelt over een ander kind				0,67				
Een kind lacht een ander kind uit				0,72				
Een kind scheldt een ander kind uit				0,68				
Een kind krabt of knijpt een ander kind				0,46				
Een kind schopt of slaat een ander kind				0,58				
Een kind beschadigt of vernielt iets van een ander kind				0,45				
Een kind treitert of pest een ander kind				0,67				
Onordelijke leeromgeving								
Er is veel onrust en lawaai in de klas								0,67
De leerlingen doen niet wat de juf of meester zegt								0,65
De juf of meester krijgt de klas maar niet stil								0,71
De leerlingen kunnen niet goed werken								0,57
De leerlingen gaan pas aan het werk als de les al lang begonnen is								0,55
Leerlingen kletsen en klieren als ze in groepjes werken								0,60
De leerlingen verstoren de les								0,68

Factorzuiverheid van LL aanvullend

In een voorbereidende analyse zijn alle 64 uitspraken van het aanvullend gedeelte van de leerlinglijst (LL aanvullend) aan een exploratieve factoranalyse onderworpen. Op grond van theoretische overwegingen en de resultaten van de tijdens de proefafnames uitgevoerde factoranalyses (Kuhlemeier, 2015) is een gedwongen tien-factorstructuur aan de gegevens opgelegd. De tien factoren blijken in totaal 59% van de variantie in sociaal-emotioneel functioneren te verklaren. De percentages verklaarde variantie per factor bedragen achtereenvolgens 17%, 12%, 8%, 6%, 4%, 4%, 3%, 3%, 2% en 2%. De factorladingen zijn weergegeven in tabel 6.6.

Tabel 6.6 Uitkomsten van een exploratieve factoranalyse van de uitspraken van LL aanvullend

	Factor									
	1	2	3	4	5	6	7	8	9	10
Pro-sociaal gedrag										
Ik kan goed dingen delen met anderen					0,48					
Ik help andere kinderen om hun ruzie op te lossen					0,63					
Ik help andere kinderen als ze iets niet leuk vinden					0,67					
Ik troost andere kinderen als zij pijn of verdriet hebben					0,63					
Ik kan ruzies met andere kinderen zelf oplossen					0,56					
Ik help andere kinderen					0,73					
Ik kom op voor andere kinderen					0,73					
Pestgedrag										
Ik maak ruzie met een ander kind										
Ik lach andere kinderen uit						0,56				
Ik heb een ander kind met opzet verwond (blauwe plek of erger)						0,67				
Ik treiter of pest een ander kind						0,73				
Ik sla of schop een ander kind						0,72				
Ik vecht met een ander kind (ik ben zelf begonnen)						0,74				
Ik scheld een ander kind uit						0,68				
Storend gedrag										
Ik stoort de les		0,70								
Ik houd me aan de gedragsregels in de klas										
Ik houd andere kinderen van hun werk			0,64							
Ik kan tijdens de les mijn mond niet houden			0,63							
Ik stoort andere kinderen bij hun werk			0,72							
Ik praat er doorheen als de juf of meester aan het woord is			0,68							
Ik val andere kinderen lastig tijdens de les			0,65							
Taakgericht gedrag										
Ik let goed op bij de lessen		-0,55								
Als de juf of meester een opdracht geeft, begin ik meteen		-0,47								
Ik werk slordig										0,81
Als we in groepjes werken, let ik goed op		-0,48								
Ik werk netjes										0,70
Als andere kinderen kletsen, werk ik rustig door		-0,61								
Als de leerkracht de klas uitloopt, werk ik rustig door		-0,62								
Motivatie voor lezen										
Ik vind lezen op school leuk							0,84			
Ik lees graag thuis							0,64			
Ik vind lezen op school saai							0,84			
Als we gaan lezen op school, heb ik daar zin in							0,83			
Ik lees omdat ik dat leuk vind							0,79			
Motivatie voor rekenen										
Ik vind rekenen op school leuk			0,82							
Ik vind het leuk om thuis te rekenen			0,73							
Ik reken graag			0,85							
Ik los graag sommen op			0,79							
Als we gaan rekenen op school, heb ik daar zin in			0,85							

Vervolg tabel 6.6

	Factor									
	1	2	3	4	5	6	7	8	9	10
Zelfbeeld lezen										
Ik begrijp de moeilijkste teksten										0,53
Ik lees goed										0,73
Ik lees langzamer dan de andere leerlingen in mijn klas										0,78
Ik lees niet zo goed als de meeste andere kinderen in mijn klas										0,73
Ik vind lezen makkelijk										0,78
Zelfbeeld rekenen										
Ik ben goed in rekenen			0,41						0,70	
Als ik iets uitreken, heb ik daar meer tijd voor nodig dan andere kinderen in mijn klas									0,71	
Als ik iets moet uitrekenen, heb ik daar meestal geen enkele moeite mee									0,66	
Ik reken niet zo goed als de meeste andere kinderen in mijn klas									0,73	
Ik vind rekenen makkelijk			0,44						0,73	
Spreekangst										
Als ik voor de hele klas iets moet vertellen, raak ik vaak in paniek				0,71						
Ik vind het eng als ik in een groepje iets moet vertellen				0,70						
Als ik in de klas iets hardop moet voorlezen, voel ik me erg gespannen				0,58						
Ik ben vaak bang dat andere kinderen mij uitlachen als ik in een groepje iets moet zeggen				0,68						
Als de juf of meester onverwachts mijn naam noemt, krijg ik het meestal heel erg warm				0,69						
Als de juf of meester mij opeens een vraag stelt, ben ik bang dat ik niet uit mijn woorden kan komen				0,66						
Ik word helemaal zenuwachtig als ik eraan denk dat de hele klas naar mij kijkt				0,75						
Als ik hoor dat we iets in groepjes moeten bespreken, raak ik vaak helemaal in paniek				0,63						
Toetsangst										
Als ik een toets maak, ben ik bang om fouten te maken	0,74									
Voordat ik een toets maak, heb ik vaak last van buikpijn	0,67									
Als ik met een toets bezig ben, ben ik zo gespannen dat ik niet goed kan nadenken	0,76									
Als ik een toets maak, voel ik me zenuwachtig	0,80									
Als ik weet dat ik morgen een toets heb, maak ik me heel erg druk	0,78									
Als ik een toets maak, denk ik vaak dat ik het niet goed doe	0,79									
Het lijkt dat mijn hart extra hard klopt wanneer ik een toets maak	0,79									
Bij het maken van een toets denk ik vaak dat het me niet zal lukken	0,79									

Behoudens twee uitzonderingen zijn de absolute waarden van de factorladingen groter dan 0,40.

De uitzonderingen betreffen de uitspraak 'Ik maak ruzie met een ander kind' uit de schaal Pestgedrag met een factorlading van 0,35 en de uitspraak 'Ik houd me aan de gedragsregels in de klas' uit de schaal Storend gedrag met een factorlading van 0,28. Na rijp beraad is besloten om deze twee uitspraken ondanks de lage factorlading toch te handhaven, enerzijds omdat ze tot de kern van het te meten begrip behoren en anderzijds omdat de afwijkingen mogelijk aan toevallige steekproeffluctuaties te wijten zijn (het aantal leerlingen in de analyses is bij LL aanvullend immers minder groot dan bij LK basis en LL basis).

Voor acht van de tien schalen van LL aanvullend geldt dat vrijwel alle uitspraken uniek op hun 'eigen' factor laden. Twee uitzonderingen doen zich voor bij de schaal Zelfbeeld rekenen waar twee uitspraken een zwakke bijlading (0,41 en 0,44) op de factor voor Motivatie voor rekenen te zien geven.

De schaal die het fraaie beeld enigszins lijkt te verstoren is Taakgericht gedrag. Zoals we eerder zagen bij de analyse van de uni-dimensionaliteit (zie paragraaf 6.1.1) bestaat deze schaal uit vijf uitspraken over aandachtig/zelf-gereguleerd werken en twee uitspraken over nauwkeurig werken. Zoals valt op te maken uit

tabel 6.6 hebben de vijf uitspraken over aandachtig/zelf-gereguleerd werken negatieve ladingen op de factor die mede bepaald wordt door de uitspraken over Storend gedrag. De tweede factor is hiermee te interpreteren als een bipolaire factor die vooral een dimensie storend gedrag versus aandachtig werken lijkt te weerspiegelen. Dat taakgerichte leerlingen minder geneigd zijn tot storend gedrag, lijkt begrijpelijk. Gezien de verschillen tussen de uitspraken lijkt het echter weinig aannemelijk dat beide schalen één factor zouden vormen. Daarnaast moet er rekening mee gehouden worden dat de afwijkingen van het veronderstelde patroon vanwege het relatief kleine aantal waarnemingen aan toevallige steekproef-fluctuaties te wijten zijn. Vandaar dat besloten is de schalen voor Storend gedrag en Taakgericht gedrag vooralsnog te handhaven.

De beide resterende uitspraken van de schaal Taakgericht gedrag bepalen samen de tiende factor. Gezien de formulering van deze uitspraken – ‘Ik werk slordig’ en ‘Ik werk netjes’ – kunnen we deze tiende factor interpreteren als nauwkeurig werken. De gevonden twee-factorstructuur van Taakgericht gedrag komt niet onverwachts. Zoals uiteengezet in paragraaf 6.1.1, vonden we dit onderscheid reeds in de proefafname die aan het normeringsonderzoek vooraf ging (Kuhlemeier, 2015). Omdat nauwkeurig werken tot de kern van het begrip Taakgericht gedrag behoort, is destijds besloten om de twee-factorstructuur voor lief te nemen. Een relativerende kanttekening is echter dat het aantal leerlingen bij LL aanvullend minder groot is dan bij LK basis en LL basis. De kans dat afwijkingen van de veronderstelde structuur aan toevallige fluctuaties te wijten zijn, is daardoor eveneens groter. Om deze reden hebben we de factoranalyse nog eens afzonderlijk uitgevoerd op uitsluitend de uitspraken van het domein Werkhouding (dat bestaat uit de uitspraken over Taakgericht gedrag, Motivatie voor lezen en Motivatie voor rekenen). De ladingen van de uitspraken uit het domein Werkhouding zijn weergegeven in tabel 6.7.

Tabel 6.7 Uitkomsten van een exploratieve factoranalyse van de uitspraken uit het domein Werkhouding (LL aanvullend)

	1	2	3
Taakgericht gedrag			
Ik let goed op bij de lessen			0,72
Als de juf of meester een opdracht geeft, begin ik meteen			0,65
Ik werk slordig			0,55
Als we in groepjes werken, let ik goed op			0,71
Ik werk netjes			0,71
Als andere kinderen kletsen, werk ik rustig door			0,64
Als de leerkracht de klas uitloopt, werk ik rustig door			0,65
Motivatie voor lezen			
Ik vind lezen op school leuk		0,86	
Ik lees graag thuis		0,70	
Ik vind lezen op school saai		0,85	
Als we gaan lezen op school, heb ik daar zin in		0,86	
Ik lees omdat ik dat leuk vind		0,85	
Motivatie voor rekenen			
Ik vind rekenen op school leuk	0,88		
Ik vind het leuk om thuis te rekenen	0,73		
Ik reken graag	0,92		
Ik los graag sommen op	0,85		
Als we gaan rekenen op school, heb ik daar zin in	0,90		

Anders dan in de hiervoor gerapporteerde factoranalyse over alle 64 uitspraken van LL aanvullend, blijken de zeven uitspraken van Taakgericht gedrag nu één factor te vormen (die we kunnen interpreteren als aandachtig, zelf-gereguleerd en nauwkeurig werken). Kennelijk is het antwoord op de vraag of de uitspraken over Taakgericht gedrag één of twee factoren vormen mede afhankelijk van de min of meer

toevallige samenstelling van de set van uitspraken die tegelijkertijd aan een factoranalyse onderworpen wordt. Deze gunstige resultaten waren voor ons extra aanleiding om de schaal voor Taakgericht gedrag te handhaven.

De conclusie is dat acht van de tien schalen van LL aanvullend empirisch goed onderscheidbaar zijn. Tegelijkertijd zijn er voldoende argumenten om de beide minder goed onderscheidbare schalen voor Taakgericht gedrag en Storend gedrag vooralsnog te handhaven. Zodra er voldoende gegevens uit dataretour beschikbaar zijn, zullen wij de kwestie taakgericht versus storend gedrag opnieuw onder de loep nemen.

6.1.3 Correlaties tussen de schalen

De begripsvaliditeit van de meting van sociaal-emotioneel functioneren moet onder meer blijken uit de interne structuur van de meting. Als de veronderstelde of beoogde structuur hout snijdt, zou die ook in de gegevens terug te vinden moeten zijn. In de toetsingsfase van het valideringsproces (Cronbach, 1970) zijn er tenminste drie situaties die aanleiding geven om de begripsvaliditeit van de meting van sociaal-emotioneel functioneren in twijfel te trekken:

- Ten eerste kunnen de schalen die de onderscheiden aspecten van sociaal-emotioneel functioneren pretenderen te meten perfect samenhangen. Het onderscheid tussen de aspecten is dan empirisch betekenisloos. Dit kan aanleiding zijn om het instrument (en de eraan ten grondslag liggende theorie) te vereenvoudigen door het onderscheid tussen de desbetreffende aspecten te laten vallen.
- Ten tweede kan er sprake zijn van een niet perfecte samenhang van de aspecten, terwijl de richting van de samenhang anders is dan op grond van theorie te verwachten valt. Ook dan is er aanleiding om de validiteit van de meting en/of de onderliggende theorie in twijfel te trekken.
- Ten derde kunnen aspecten geen enkele of slechts een geringe samenhang vertonen. Ze kunnen dan niet meer worden beschouwd als behorend tot hetzelfde overkoepelende begrip. Het ware dan beter om te spreken van afzonderlijke, op zichzelf staande begrippen.

Hiervóór hebben we beargumenteerd dat er weinig reden is om aan de uni-dimensionaliteit en factor-zuiverheid van de meting van de onderscheiden aspecten van het sociaal-emotioneel functioneren te twifelen. De resultaten van de factoranalyses bleken in grote mate overeen te komen met de op basis van theoretische argumenten gebaseerde indeling van de 174 uitspraken over de schalen van LK basis, LL basis en LL aanvullend. In deze paragraaf bespreken we de correlaties tussen de schalen zoals deze door factoranalyses bevestigd werden. Vanwege het zeer grote aantal relaties bespreken we in deze paragraaf alleen de correlaties tussen de schalen van elk van de onderdelen LK basis, LL basis en LL aanvullend afzonderlijk. De correlaties tussen de schalen van verschillende onderdelen worden gepresenteerd en besproken in bijlage 1.

Interne structuur van LK basis

De correlaties tussen de acht schalen van de leerkrachtlijst zijn weergegeven in tabel 6.8.

Tabel 6.8 *Correlaties tussen de schalen van LK basis*

	1	2	3	4	5	6	7	8
1 pro-sociaal gedrag	1							
2 pestgedrag	-0,53	1						
3 storend gedrag	-0,43	0,58	1					
4 teruggetrokken gedrag	-0,39	0,16	0,06	1				
5 taakgericht gedrag	0,50	-0,40	-0,54	-0,22	1			
6 motivatie voor leren	0,53	-0,41	-0,49	-0,32	0,72	1		
7 zelfstandigheid	0,40	-0,02	0,05	-0,62	0,28	0,39	1	
8 hyperactief/impulsief	-0,49	0,55	0,75	0,23	-0,61	-0,52	-0,11	1

Over het geheel genomen zijn de richting en de sterkte van de correlaties tussen de schalen van de leerkrachtlijst overeenkomstig theoretische verwachtingen. Zo vertoont Pro-sociaal gedrag een middelmatig negatieve samenhang met schalen die externaliserend probleemgedrag beogen te meten: Pestgedrag ($r = -0,53$), Storend gedrag ($r = -0,43$) en Hyperactiviteit/impulsiviteit ($r = -0,49$). Daarnaast hangt Pro-sociaal positief samen met andere 'positieve' indicatoren van sociaal-emotioneel functioneren, zoals Taakgericht gedrag ($r = 0,50$), Motivatie voor leren ($r = 0,53$) en Zelfstandigheid ($r = 0,40$). Pestgedrag hangt zoals verwacht positief samen met Storend gedrag ($r = 0,58$), maar negatief met Taakgericht gedrag ($r = -0,40$) en Motivatie voor leren ($r = -0,41$). Evenmin onverwacht is dat Taakgericht gedrag en Motivatie voor leren sterk positief samenhangen ($r = 0,72$).

Evenzeer in lijn met theoretische noties is de sterk positieve samenhang tussen Hyperactief/impulsief en Storend gedrag ($r = 0,75$), de sterk negatieve samenhang tussen Hyperactief/impulsief en Taakgericht gedrag ($r = -0,61$) en de sterk negatieve samenhang tussen Zelfstandigheid en Teruggetrokken gedrag ($r = -0,62$). Deze laatste drie correlaties zijn weliswaar hoog, maar niet zo hoog dat aan de onderscheidbaarheid van deze aspecten van het sociaal-emotioneel functioneren getwijfeld moet worden. Het patroon van negatieve correlaties met Teruggetrokken gedrag en Zelfstandigheid doet vermoeden dat beide begrippen min of meer op zichzelf staande vormen van minder sociaal, niet-externaliserend gedrag vertegenwoordigen. Daarbij merken we op dat minder positieve scores voor Teruggetrokken gedrag en Zelfstandigheid niet noodzakelijkerwijs als problematisch beschouwd moeten worden.

Interne structuur van LL basis

De correlaties tussen de acht schalen van LL basis zijn weergegeven in tabel 6.9.

Tabel 6.9 Correlaties tussen de schalen van LL basis

	1	2	3	4	5	6	7	8
1 sociaal zelfbeeld	1							
2 cognitief zelfbeeld	0,39	1						
3 schoolbeleving	0,52	0,35	1					
4 gepest worden	-0,38	-0,17	-0,40	1				
5 motivatie voor leren	0,36	0,58	0,33	-0,13	1			
6 sociale leeromgeving	0,48	0,27	0,49	-0,33	0,30	1		
7 onveilige leeromgeving	-0,27	-0,16	-0,35	0,62	-0,12	-0,46	1	
8 onordelijke leeromgeving	-0,29	-0,22	-0,36	0,40	-0,22	-0,38	0,52	1

De sterkte en richting van de samenhang is in grote lijnen overeenkomstig de verwachtingen. Zo hangt Gepest worden sterk samen met Onveilige leeromgeving ($r = 0,62$). Schoolbeleving vertoont een middelmatig positieve samenhang met Sociaal zelfbeeld ($r = 0,52$) en Sociale leeromgeving ($r = 0,48$) maar correleert middelmatig negatief met Gepest worden ($r = -0,40$).

Evenmin onverwacht is de middelmatig positieve relatie tussen Motivatie voor leren en Cognitief zelfbeeld ($r = 0,58$).

De correlatie tussen sociaal en cognitief zelfbeeld ($r = 0,39$) is relatief laag als men bedenkt dat beide schalen op hetzelfde hiërarchische niveau van zelfbeeld betrekking hebben, maar wel in overeenstemming met relevante literatuur (Harter, 1999; Marsh, 1990; Marsh & Shavelson, 1985; Rosenberg, Schooler, Schoenbach & Rosenberg, 1995; Shavelson, Hubner & Stanton, 1976). Een vergelijkbaar lage samenhang tussen sociaal en cognitief zelfbeeld werd overigens ook gevonden in de Peiling Sociale Competentie (Kuhlemeier, Van Boxtel & Van Til, 2012).

Interne structuur van LL aanvullend

De correlaties tussen de tien schalen van de aanvullende leerlingenlijst zijn weergegeven in tabel 6.10.

Tabel 6.10 Correlaties tussen de schalen van LL aanvullend

		1	2	3	4	5	6	7	8	9	10
1	pro-sociaal gedrag	1									
2	pestgedrag	-0,26	1								
3	storend gedrag	-0,26	0,58	1							
4	taakgericht gedrag	0,44	-0,35	-0,60	1						
5	motivatie voor lezen	0,31	-0,24	-0,28	0,37	1					
6	motivatie voor rekenen	0,26	-0,13	-0,22	0,37	0,27	1				
7	zelfbeeld lezen	0,20	-0,15	-0,18	0,28	0,37	-0,02	1			
8	zelfbeeld rekenen	0,08	0,00	-0,16	0,21	-0,04	0,52	0,15	1		
9	spreekangst	-0,17	0,05	0,10	-0,13	0,00	0,01	-0,21	-0,23	1	
10	toetsangst	0,03	0,10	0,09	0,05	0,02	0,05	-0,09	-0,24	0,55	1

De correlaties tussen de schalen van LL aanvullend blijken overwegend lager dan bij LK basis en LL basis het geval was. Maar ook nu is het patroon in overeenstemming met theoretische verwachtingen.

Eén correlatie kan als hoog worden aangemerkt: Storend gedrag hangt zoals verwacht sterk negatief samen met Taakgericht gedrag ($r = -0,60$).

Vier correlaties vallen in het bereik middelmatig. Het betreft de samenhang tussen Pro-sociaal gedrag en Taakgericht gedrag ($r = 0,44$), tussen Pestgedrag en Storend gedrag ($r = 0,58$), tussen Zelfbeeld rekenen en Motivatie voor rekenen ($r = 0,52$) en tussen Spreekangst en Toetsangst ($r = 0,55$).

Binnen het domein Sociaal gedrag hangen Pro-sociaal gedrag en Storend gedrag zoals gezegd sterk samen, maar Pro-sociaal gedrag vertoont een zwak negatieve samenhang met Pestgedrag en Storend gedrag (r respectievelijk $-0,26$ en $-0,26$).

De correlatie tussen Zelfbeeld lezen en Zelfbeeld rekenen is verwaarloosbaar laag ($r = 0,15$), terwijl Motivatie voor lezen slechts zwak samenhangt met Motivatie voor rekenen ($r = 0,27$). Beide motivatieschalen hangen zwak samen met Taakgericht gedrag (r respectievelijk $0,37$ en $0,37$). Vergelijkbaar lage correlaties tussen het zelfbeeld van lezen en rekenen en tussen de motivatie voor lezen en rekenen zijn ook gevonden in de Peiling Sociale Competentie (Kuhlemeier, Van Boxtel & Van Til, 2012) en wat betreft het zelfbeeld ook in het onderzoek van Marsh (1986). Leerlingen die vinden dat zij goed kunnen lezen of gemotiveerd zijn voor lezen zijn slechts in zeer beperkte mate ook goed in rekenen of gemotiveerd voor rekenen.

Spreek- en toetsangst hangen zoals verwacht positief met elkaar samen ($r = 0,55$), terwijl de correlaties met niet-verwante schalen van LL aanvullend verwaarloosbaar klein zijn (wellicht met uitzondering van de zwak negatieve correlatie van Spreekangst met Zelfbeeld lezen en Zelfbeeld rekenen en tussen Toetsangst en Zelfbeeld rekenen).

Overeenkomstig de eis van discriminante validiteit hangen Zelfbeeld lezen, Zelfbeeld rekenen, Spreekangst en Toetsangst niet of nauwelijks samen met Pro-sociaal gedrag, Pestgedrag en Storend gedrag ($-0,18 < r < 0,20$).

Conclusie

Al met al is de relatieve sterkte en de richting van de correlaties tussen de schalen van LK-basis, LL basis en LL-aanvullend in overeenstemming met theoretische noties.

Correlaties tussen schalen van verschillende onderdelen

Hiervoor is alleen ingegaan op de correlaties tussen de schalen binnen elk van de drie onderdelen van VISEON 2.0 afzonderlijk. De correlaties tussen schalen van verschillende onderdelen zijn vanwege de omvang niet in de hoofdttekst opgenomen. Daarvoor verwijzen wij de geïnteresseerde lezer naar bijlage 1.

6.1.4 Invariantie over relevante groepen

Eerder in dit hoofdstuk bespraken we de interne structuur van de meting van het sociaal-emotioneel functioneren aan de hand van de correlaties tussen de schalen (die goed interpreteerbaar bleken). In deze paragraaf gaan we na in hoeverre deze interne structuur vergelijkbaar is voor verschillende groepen leerlingen, te weten jongens en meisjes, deelnemers aan de papieren en digitale afname en leerlingen van verschillende jaargroepen.

Geslacht

Uit eerder onderzoek is bekend dat er bij sommige aspecten van het sociaal-emotioneel functioneren aanzienlijke verschillen bestaan tussen de gemiddelde scores van jongens en meisjes (o.a. Kuhlemeier, Van Boxtel & Van Til, 2012). Zoals we zullen zien in paragraaf 6.3.2 vinden we deze verschillen ook terug in de dataset van het normeringsonderzoek. In deze paragraaf kijken we niet naar sekseverschillen in het niveau van sociaal-emotioneel functioneren maar richten we ons op de vraag in hoeverre de hiervoor gerapporteerde interne structuur geldig is voor jongens en meisjes. Vanwege het grote aantal schalen bespreken we de uitkomsten afzonderlijk voor LK basis, LL basis en LL aanvullend.

Geslacht: LK basis

De correlaties tussen de acht schalen van LK basis zijn weergegeven in tabel 6.11, waarbij de correlaties voor meisjes onder de diagonaal staan en die voor jongens erboven. Ondanks de soms forse verschillen qua gemiddelden (zie paragraaf 6.3.2) blijkt het correlatiepatroon voor meisjes sterk te lijken op dat voor jongens. Het absolute verschil tussen overeenkomstige correlaties voor jongens en meisjes bedraagt maximaal 0,11 met een gemiddelde van ,04. De sterkste afwijking doet zich voor bij de samenhang tussen Hyperactief/impulsief en Motivatie voor leren waarbij de correlatie voor meisjes sterker is dan voor jongens (r respectievelijk -0,56 en -0,45).

Tabel 6.11 Correlaties tussen de acht schalen van LK basis voor meisjes (onder de diagonaal) en jongens (boven de diagonaal)

	1	2	3	4	5	6	7	8
1 pro-sociaal gedrag	1,00	-0,53	-0,43	-0,40	0,46	0,50	0,41	-0,43
2 pestgedrag	-0,49	1,00	0,59	0,19	-0,38	-0,39	-0,01	0,53
3 storend gedrag	-0,38	0,54	1,00	0,06	-0,53	-0,48	0,04	0,73
4 teruggetrokken gedrag	-0,40	0,14	0,04	1,00	-0,21	-0,28	-0,62	0,24
5 taakgericht gedrag	0,48	-0,35	-0,50	-0,25	1,00	0,70	0,31	-0,55
6 motivatie voor leren	0,53	-0,42	-0,46	-0,37	0,72	1,00	0,40	-0,45
7 zelfstandigheid	0,38	0,00	0,11	-0,63	0,24	0,36	1,00	-0,07
8 hyperactief/impulsief	-0,51	0,54	0,74	0,24	-0,62	-0,56	-0,11	1,00

Geslacht: LL basis

De overeenkomstige correlaties voor LL basis zijn weergegeven in tabel 6.12. Net als bij LK basis het geval was, lijken de correlaties voor jongens sterk op die voor meisjes. Het absolute verschil tussen overeenkomstige correlaties voor jongens en meisjes bedraagt maximaal 0,08 met een gemiddelde van 0,04. De sterkste afwijking doet zich voor bij de samenhang tussen Sociale leeromgeving en Onordelijke leeromgeving die voor meisjes sterker is dan voor jongens (r respectievelijk -0,42 en -0,34).

Tabel 6.12 *Correlaties tussen de acht schalen van LL basis voor meisjes (onder de diagonaal) en jongens (boven de diagonaal)*

	1	2	3	4	5	6	7	8
1 sociaal zelfbeeld	1,00	0,37	0,49	-0,36	0,38	0,47	-0,25	-0,26
2 cognitief zelfbeeld	0,40	1,00	0,38	-0,17	0,59	0,28	-0,15	-0,23
3 schoolbeleving	0,56	0,32	1,00	-0,37	0,35	0,49	-0,33	-0,38
4 gepest worden	-0,41	-0,14	-0,42	1,00	-0,12	-0,32	0,62	0,39
5 motivatie voor leren	0,33	0,56	0,29	-0,13	1,00	0,31	-0,10	-0,22
6 sociale leeromgeving	0,49	0,25	0,50	-0,35	0,29	1,00	-0,43	-0,34
7 onveilige leeromgeving	-0,30	-0,14	-0,35	0,60	-0,13	-0,49	1,00	0,51
8 onordelijke leeromgeving	-0,31	-0,20	-0,34	0,42	-0,22	-0,42	0,54	1,00

Geslacht: LL aanvullend

De overeenkomstige correlaties voor LL aanvullend zijn weergegeven in tabel 6.13. Net als bij LK basis en LL basis het geval was, vertonen de correlaties voor jongens en meisjes een sterke gelijkenis. Het absolute verschil tussen overeenkomstige correlaties voor jongens en meisjes bedraagt maximaal 0,22 met een gemiddelde van 0,07. De sterkste afwijkingen doen zich voor bij de relatie van Toetsangst met Storend gedrag en Taakgericht gedrag, van Spreekangst met Storend gedrag en Taakgericht gedrag (voor meisjes wat zwakker dan voor jongens) en bij de relatie tussen Pro-sociaal gedrag en Pestgedrag (voor meisjes wat zwakker negatief dan voor jongens). Een geruststellende constatering is dat de afwijkingen bij de hogere (en meestal significante) correlaties over het algemeen veel kleiner zijn dan bij de lagere (en doorgaans niet-significante) correlaties. Wel zijn de afwijkingen gemiddeld groter dan bij LK basis en LL basis het geval was, maar dat is zonder twijfel ook toe te schrijven aan toevallige fluctuaties ten gevolge van het veel kleinere aantal waarnemingen.

Tabel 6.13 *Correlaties tussen de tien schalen van LL aanvullend voor meisjes (onder de diagonaal) en jongens (boven de diagonaal)*

	1	2	3	4	5	6	7	8	9	10
1 pro-sociaal gedrag	1,00	-0,31	-0,23	0,42	0,27	0,33	0,09	0,10	-0,15	0,01
2 pestgedrag	-0,12	1,00	0,61	-0,38	-0,17	-0,16	-0,15	-0,07	0,11	0,19
3 storend gedrag	-0,21	0,44	1,00	-0,57	-0,24	-0,23	-0,17	-0,24	0,23	0,24
4 taakgericht gedrag	0,41	-0,24	-0,55	1,00	0,32	0,39	0,27	0,37	-0,24	-0,09
5 motivatie voor lezen	0,28	-0,16	-0,25	0,33	1,00	0,29	0,31	0,04	-0,04	-0,02
6 motivatie voor rekenen	0,25	-0,15	-0,21	0,42	0,30	1,00	-0,03	0,57	-0,04	0,00
7 zelfbeeld lezen	0,27	-0,14	-0,26	0,31	0,40	0,03	1,00	0,15	-0,26	-0,13
8 zelfbeeld rekenen	0,20	-0,14	-0,22	0,26	0,01	0,51	0,21	1,00	-0,24	-0,21
9 spreekangst	-0,25	0,08	0,07	-0,07	-0,01	0,03	-0,22	-0,18	1,00	0,53
10 toetsangst	-0,06	0,16	0,02	0,11	0,01	0,10	-0,12	-0,18	0,59	1,00

Afnamemodus

Eerder in deze verantwoording constateerden wij dat de afnamemodus – afname op papier versus digitaal – nauwelijks van invloed is op de hoogte en spreiding van de scores (zie paragraaf 4.2.1). In deze paragraaf richten we ons op de vraag in hoeverre de eerder gerapporteerde interne structuur geldig is voor de gegevens van de papieren en digitale afname. Vanwege het grote aantal schalen bespreken we de uitkomsten afzonderlijk voor LK basis en LL basis. LL aanvullend ontbreekt omdat dit onderdeel alleen digitaal aan leerlingen is voorgelegd.

Afnamemodus: LK basis

De correlaties tussen de acht schalen van LK basis zijn weergegeven in tabel 6.14, waarbij die voor de digitale afname onder de diagonaal staan en die voor papieren afname erboven. De schaal Hyperactief/impulsief ontbreekt omdat deze uitspraken alleen digitaal aan leerkrachten zijn voorgelegd. Inspectie van de beide correlatiepatronen brengt vrijwel alleen kleine verschillen aan het licht. Het absolute verschil tussen overeenkomstige correlaties voor de papieren en digitale afname is maximaal 0,16 met een gemiddelde van 0,05. De afwijking van 0,16 betreft de relatie tussen Zelfstandigheid en Teruggetrokken gedrag: in de digitale afname is de correlatie -0,67 en in de papieren afname -0,51.

Tabel 6.14 Correlaties tussen de zeven schalen van LK basis voor de digitale afname (onder de diagonaal) en de papieren afname (boven de diagonaal)

		1	2	3	4	5	6	7
1	pro-sociaal gedrag	1,00	-0,54	-0,40	-0,34	0,49	0,52	0,38
2	pestgedrag	-0,53	1,00	0,55	0,21	-0,44	-0,44	-0,05
3	storend gedrag	-0,45	0,60	1,00	0,10	-0,54	-0,48	0,12
4	teruggetrokken gedrag	-0,41	0,13	0,04	1,00	-0,17	-0,31	-0,51
5	taakgericht gedrag	0,51	-0,38	-0,54	-0,24	1,00	0,72	0,22
6	motivatie voor leren	0,54	-0,40	-0,50	-0,32	0,73	1,00	0,32
7	zelfstandigheid	0,41	-0,01	0,02	-0,67	0,31	0,42	1,00

Afnamemodus: LL basis

De overeenkomstige correlaties voor LL basis zijn weergegeven in tabel 6.15. De gemiddelde absolute verschillen tussen overeenkomstige correlaties voor de papieren en digitale afname zijn wederom klein. Het verschil is maximaal 0,11 met een gemiddelde van 0,03. De afwijking van 0,11 doet zich voor bij de relatie tussen Onveilige leeromgeving en Schoolbeleving (-0,27 digitaal versus -0,38 papier) en tussen Onordelijke leeromgeving en Gepest worden (0,48 digitaal en 0,37 papier).

Tabel 6.15 Correlaties tussen de acht schalen van LL basis voor de digitale afname (onder de diagonaal) en de papieren afname (boven de diagonaal)

		1	2	3	4	5	6	7	8
1	sociaal zelfbeeld	1,00	0,39	0,53	-0,38	0,36	0,49	-0,27	-0,27
2	cognitief zelfbeeld	0,39	1,00	0,34	-0,17	0,59	0,28	-0,15	-0,23
3	schoolbeleving	0,50	0,37	1,00	-0,39	0,30	0,50	-0,38	-0,36
4	gepest worden	-0,38	-0,16	-0,43	1,00	-0,12	-0,32	0,61	0,37
5	motivatie voor leren	0,34	0,54	0,36	-0,14	1,00	0,30	-0,10	-0,22
6	sociale leeromgeving	0,46	0,24	0,46	-0,36	0,29	1,00	-0,46	-0,38
7	onveilige leeromgeving	-0,28	-0,18	-0,27	0,63	-0,16	-0,44	1,00	0,51
8	onordelijke leeromgeving	-0,33	-0,19	-0,37	0,48	-0,22	-0,37	0,56	1,00

Jaargroep

Eerder in deze verantwoording constateerden wij dat de jaargroep waarin de leerling zich bevindt nauwelijks invloed heeft op de hoogte en spreiding van de scores (zie paragraaf 4.2.2). In deze paragraaf stellen we de vraag in hoeverre de eerder gerapporteerde interne structuur geldig is voor leerlingen in verschillende leerjaren. De leerkrachtlijst is ingevuld voor leerlingen in groep 3, 4, 5, 6, 7 en 8. Om het aantal analyses en tabellen te beperken, hebben in de analyse twee groepen onderscheiden: leerlingen in groep 3, 4 en 5 versus leerlingen in groep 6, 7 en 8. Bij de leerlinglijst die is ingevuld in jaargroep 5, 6, 7 en 8 onderscheiden we eveneens twee groepen: groep 5 en 6 versus groep 7 en 8. Vanwege het grote aantal schalen bespreken we de uitkomsten afzonderlijk voor LK basis, LL basis en LL aanvullend.

Jaargroep: LK basis

De correlaties tussen de acht schalen van LK basis zijn weergegeven in tabel 6.16, waarbij de correlaties voor de leerlingen in groep 6, 7 en 8 onder de diagonaal staan en die voor groep 3, 4 en 5 erboven. De verschillen blijken klein te zijn. Het absolute verschil tussen overeenkomstige correlaties voor jongere en oudere leerlingen bedraagt maximaal 0,18 met een gemiddelde van 0,08. De vier sterkste afwijkingen betreffen relaties van Pestgedrag met Hyperactief/impulsief (0,44 versus 0,63), Taakgericht gedrag (-0,29 versus -0,45), Teruggetrokken gedrag (0,06 versus 0,20) en Motivatie voor leren (-0,32 versus -0,45). In alle vier gevallen is de samenhang voor oudere leerlingen wat sterker dan voor jongere leerlingen.

Tabel 6.16 Correlaties tussen de acht schalen van LK basis voor groep 6, 7 en 8 (onder de diagonaal) en groep 3, 4 en 5 (boven de diagonaal)

		1	2	3	4	5	6	7	8
1	pro-sociaal gedrag	1,00	-0,48	-0,44	-0,38	0,46	0,51	0,34	-0,49
2	pestgedrag	-0,56	1,00	0,55	0,06	-0,29	-0,32	0,04	0,44
3	storend gedrag	-0,44	0,60	1,00	-0,01	-0,46	-0,41	0,05	0,67
4	teruggetrokken gedrag	-0,39	0,20	0,08	1,00	-0,22	-0,36	-0,70	0,25
5	taakgericht gedrag	0,51	-0,45	-0,58	-0,21	1,00	0,74	0,35	-0,55
6	motivatie voor leren	0,55	-0,45	-0,52	-0,30	0,73	1,00	0,46	-0,50
7	zelfstandigheid	0,42	-0,05	0,06	-0,58	0,25	0,36	1,00	-0,16
8	hyperactief/impulsief	-0,49	0,63	0,80	0,21	-0,65	-0,54	-0,06	1,00

Jaargroep: LL basis

De overeenkomstige correlaties voor LL basis zijn weergegeven in tabel 6.17. De verschillen tussen overeenkomstige correlaties voor jongere en oudere leerlingen zijn wederom klein. De grootste absolute afwijking is 0,13 en de gemiddelde absolute afwijking bedraagt 0,04. De twee grootste afwijkingen betreffen de relatie tussen Onveilige leeromgeving en Schoolbeleving (-0,24 versus -0,37) en tussen Onordelijke leeromgeving en Gepest worden (0,52 versus 0,38).

Tabel 6.17 Correlaties tussen de acht schalen van LL basis voor groep 7 en 8 (onder de diagonaal) en groep 5 en 6 (boven de diagonaal)

		1	2	3	4	5	6	7	8
1	sociaal zelfbeeld	1,00	0,40	0,48	-0,33	0,40	0,41	-0,27	-0,31
2	cognitief zelfbeeld	0,38	1,00	0,40	-0,17	0,56	0,26	-0,17	-0,19
3	schoolbeleving	0,53	0,35	1,00	-0,41	0,38	0,47	-0,24	-0,38
4	gepest worden	-0,39	-0,17	-0,40	1,00	-0,12	-0,33	0,65	0,52
5	motivatie voor leren	0,35	0,58	0,31	-0,14	1,00	0,28	-0,10	-0,17
6	sociale leeromgeving	0,49	0,27	0,49	-0,33	0,31	1,00	-0,39	-0,38
7	onveilige leeromgeving	-0,27	-0,16	-0,37	0,61	-0,12	-0,47	1,00	0,59
8	onordelijke leeromgeving	-0,28	-0,23	-0,36	0,38	-0,24	-0,37	0,51	1,00

Jaargroep: LL aanvullend

De correlaties voor LL aanvullend zijn weergegeven in tabel 6.18. De absolute verschillen tussen overeenkomstige correlaties voor oudere en jongere leerlingen zijn wat groter dan bij LK basis en LL basis het geval was. Dit is zonder twijfel mede toe te schrijven aan toevallige fluctuaties ten gevolge van het wat minder grote aantal waarnemingen. De grootste afwijking is 0,24 en het gemiddelde bedraagt 0,09. De zes grootste afwijkingen betreffen relaties tussen:

- Spreekangst en Pro-sociaal gedrag (-0,31 versus -0,07);
- Spreekangst en Pestgedrag (0,16 versus -0,01);

- Spreekangst en Taakgericht gedrag (-0,23 versus -0,05);
- Pestgedrag en Motivatie voor lezen (-0,31 versus -0,14);
- Storend gedrag en Zelfbeeld lezen (-0,27 versus -0,12);
- Spreekangst en Toetsangst (0,49 versus 0,62).

Tabel 6.18 Correlaties tussen de tien schalen van LL aanvullend voor groep 7 en 8 (onder de diagonaal) en groep 5 en 6 (boven de diagonaal)

		1	2	3	4	5	6	7	8	9	10
1	pro-sociaal gedrag	1,00	-0,24	-0,21	0,49	0,36	0,29	0,15	0,04	-0,07	0,02
2	pestgedrag	-0,34	1,00	0,54	-0,33	-0,14	-0,07	-0,10	0,01	-0,01	0,15
3	storend gedrag	-0,31	0,61	1,00	-0,54	-0,21	-0,19	-0,12	-0,17	0,08	0,12
4	taakgericht gedrag	0,42	-0,40	-0,63	1,00	0,31	0,40	0,28	0,24	-0,05	0,07
5	motivatie voor lezen	0,24	-0,31	-0,32	0,36	1,00	0,27	0,32	-0,08	0,03	0,06
6	motivatie voor rekenen	0,24	-0,18	-0,21	0,32	0,22	1,00	-0,06	0,50	0,06	0,06
7	zelfbeeld lezen	0,23	-0,21	-0,27	0,29	0,40	0,05	1,00	0,13	-0,22	-0,12
8	zelfbeeld rekenen	0,16	-0,04	-0,16	0,21	0,00	0,58	0,19	1,00	-0,19	-0,23
9	spreekangst	-0,31	0,16	0,18	-0,23	-0,04	-0,13	-0,23	-0,28	1,00	0,62
10	toetsangst	0,02	0,06	0,11	-0,03	-0,06	-0,06	-0,08	-0,27	0,49	1,00

De verschillen tussen de correlaties voor jongere en oudere leerlingen doen zich niet bij alle schalen in gelijke mate voor. Zo zijn vier van de zes grootste afwijkingen gerelateerd aan spreekangst. De samenhang van spreekangst met andere aspecten van het sociaal-emotioneel functioneren lijkt bij oudere leerlingen meestal sterker dan bij jongere leerlingen.

6.2 Kwaliteit van de itemparameters: DIF-analyse

Nagegaan is in hoeverre de uitspraken van VISEON 2.0 *Differential Item Functioning* (DIF) vertonen ten gevolge van de afnamemodus (schriftelijke versus digitale afname), de jaargroep (oudere versus jongere leerlingen) en geslacht. Er is sprake van DIF als leerlingen met een gelijk niveau van sociaal-emotioneel functioneren afhankelijk van het achtergrondkenmerk op een uitspraak hogere of lagere scores behalen dan op grond van hun totaalscore verwacht zou worden. DIF-analyse is onder meer van belang omdat het informatie kan opleveren over mogelijke meer-dimensionaliteit van het gemeten begrip. De analyse is per schaal uitgevoerd met de Mantel-Haenzel procedure (Heuvelmans, 2013). De procedure resulteert in een toetsingsgrootheid – DIFSTAT genoemd – die een waarde heeft van omstreeks 1 als er geen sprake is van DIF; de uitspraken hebben dan voor de onderscheiden subgroepen dezelfde betekenis. Als de waarde groter is dan 1, dan scoort de eerste groep hoger dan de tweede groep. Is de waarde kleiner dan 1, dan scoort de tweede groep hoger dan de eerste. In de meeste gevallen kan het programma ook een z-waarde berekenen die aangeeft of de DIFSTAT significant van 1 verschilt. Er is sprake van een significant verschil op 1%-niveau als de absolute waarde van z gelijk of groter is dan de absolute waarde van 2,58.

6.2.1 Afnamemodus

Voor de schalen waarvoor zowel papieren als digitale gegevens beschikbaar zijn, is nagegaan in hoeverre de uitspraken *Differential Item Functioning* (DIF) ten gevolge van de afnamemodus vertonen. De resultaten van de Mantel-Haenzel procedure zijn weergegeven in het linker gedeelte van tabel 6.19. Weergegeven zijn de laagst en hoogst waargenomen waarde van DIFSTAT en z en hun gemiddelde waarde. Een DIFSTAT-waarde groter dan 1 betekent dat leerlingen die aan de papieren afname deelnamen gegeven een gelijke sociaal-emotionele vaardigheid op de desbetreffende uitspraak relatief hoge scores

behalen. Een DIFSTAT-waarde kleiner dan 1 betekent dat leerlingen die aan de digitale afname deelnamen naar verhouding hoge scores behaalden.

Tabel 6.19 Resultaten van de analyse van Differential Item Functioning ten gevolge van afnamemodus, jaargroep en geslacht

	Statistic	Afnamemodus			Jaargroep			Geslacht		
		Min	Max	Gem	Min	Max	Gem	Min	Max	Gem
pro-sociaal gedrag	DIFSTAT	0,937	1,035	0,972	0,850	1,077	1,010	0,840	1,074	0,988
	z	-0,180	0,097	-0,081	-0,448	0,207	0,021	-0,504	0,215	-0,043
pestgedrag	DIFSTAT	0,730	1,380	0,935	0,544	1,079	0,848	0,905	2,335	1,344
	z	##	##	##	##	##	##	##	##	##
storend gedrag	DIFSTAT	0,706	1,120	0,916	0,710	1,037	0,890	0,754	1,206	1,024
	x	-0,722	0,285	-0,209	-0,797	0,076	-0,297	-0,699	0,442	-0,001
teruggetrokken gedrag	DIFSTAT	0,907	1,267	1,023	0,739	1,217	0,986	0,886	1,183	0,983
	z	-0,253	0,683	0,058	-0,834	0,511	-0,069	-0,357	0,442	-0,071
taakgericht gedrag	DIFSTAT	0,917	1,105	1,011	0,716	1,119	0,947	0,671	1,065	0,943
	z	-0,252	0,290	0,026	-0,964	0,318	-0,201	-1,170	0,189	-0,207
motivatie voor leren	DIFSTAT	0,908	1,066	1,002	0,839	1,125	1,005	0,876	1,611	1,006
	z	-0,279	0,174	0,000	-0,475	0,325	0,006	-0,383	1,362	-0,043
zelfstandigheid	DIFSTAT	0,883	1,071	0,973	0,864	1,139	1,021	0,884	1,213	1,001
	z	-0,360	0,207	-0,085	-0,431	0,379	0,048	-0,381	0,593	-0,017
hyperactief/impulsief	DIFSTAT	--	--	--	0,722	0,977	0,883	0,889	1,356	1,185
	z	--	--	--	##	##	##	-0,246	0,620	0,323
sociaal zelfbeeld	DIFSTAT	0,834	1,080	0,943	0,881	1,198	1,035	0,884	1,055	0,989
	z	-0,635	0,272	-0,218	-0,371	0,506	0,080	-0,467	0,194	-0,046
cognitief zelfbeeld	DIFSTAT	0,860	1,044	0,958	0,854	1,176	1,032	0,897	1,167	0,987
	z	-0,495	0,158	-0,153	-0,480	0,464	0,068	-0,435	0,597	-0,061
schoolbeleving	DIFSTAT	0,924	1,266	1,062	0,740	1,316	0,948	0,750	1,307	0,966
	z	-0,221	0,589	0,143	-0,883	0,657	-0,195	-0,818	0,822	-0,172
gepest worden	DIFSTAT	0,791	1,202	1,014	0,817	1,805	1,143	0,568	3,015	1,218
	z	-0,601	0,550	0,027	-0,521	1,548	0,223	-1,651	2,945	0,126
motivatie voor leren	DIFSTAT	0,921	1,043	0,985	0,910	1,113	1,008	0,903	1,113	1,016
	z	-0,308	0,169	-0,059	-0,294	0,314	0,012	-0,370	0,436	0,059
sociale leeromgeving	DIFSTAT	0,787	1,098	0,968	0,866	1,266	1,010	0,908	1,125	1,010
	z	-0,831	0,361	-0,129	-0,454	0,661	-0,008	-0,371	0,512	0,041
onveilige leeromgeving	DIFSTAT	0,805	1,415	1,021	0,633	1,568	1,104	0,660	1,177	1,036
	z	-0,753	1,406	0,063	-1,460	1,289	0,107	-1,766	0,618	0,044
onordelijke leeromgeving	DIFSTAT	0,727	1,235	0,988	0,833	1,393	1,071	0,931	1,082	0,995
	z	-1,093	0,861	-0,068	-0,603	0,936	0,130	-0,290	0,344	-0,022
pro-sociaal gedrag	DIFSTAT	--	--	--	0,875	1,147	1,013	0,447	0,869	0,717
	z	--	--	--	-0,330	0,307	0,011	##	##	##
pestgedrag	DIFSTAT	--	--	--	0,728	1,487	1,027	0,915	4,321	2,832
	z	--	--	--	##	##	##	##	##	##
storend gedrag	DIFSTAT	--	--	--	0,746	1,768	0,951	1,047	2,209	1,597
	z	--	--	--	-0,553	1,149	-0,172	##	##	##
taakgericht gedrag	DIFSTAT	--	--	--	0,845	1,209	1,047	0,511	0,851	0,712
	z	--	--	--	-0,372	0,463	0,096	##	##	##
motivatie voor lezen	DIFSTAT	--	--	--	0,875	1,177	1,030	0,492	0,656	0,589
	z	--	--	--	-0,286	0,359	0,056	##	##	##
motivatie voor rekenen	DIFSTAT	--	--	--	0,793	1,157	1,012	0,883	1,141	1,073
	z	--	--	--	-0,490	0,322	0,009	##	##	##
zelfbeeld lezen	DIFSTAT	--	--	--	0,833	1,440	1,023	0,796	0,911	0,850
	z	--	--	--	-0,405	0,782	0,009	##	##	##
zelfbeeld rekenen	DIFSTAT	--	--	--	0,879	1,093	0,976	1,380	1,643	1,516
	z	--	--	--	-0,296	0,196	-0,066	##	##	##
spreekangst	DIFSTAT	--	--	--	0,802	1,197	1,019	0,676	0,932	0,791
	z	--	--	--	-0,439	0,418	0,023	##	##	##
toetsangst	DIFSTAT	--	--	--	1,030	1,730	1,203	0,597	0,879	0,732
	z	--	--	--	0,062	1,014	0,326	##	##	##

--: alleen digitaal afgenomen; ##: z-waarde niet beschikbaar

Voor de 48 uitspraken van zeven van de acht schalen van LK basis (exclusief Hyperactiviteit/impulsiviteit waarvoor alleen digitale afnamegegevens beschikbaar waren) varieerde DIFSTAT van 0,71 tot 1,38 met een gemiddelde van 0,98. Geen van de waarden van DIFSTAT was significant op 1%-niveau (waarbij we aantekenen dat het programma de z-waarden van de uitspraken van de schaal Pestgedrag vanwege de scheefheid van de verdeling niet kon berekenen, ook niet nadat de beide minst frequent gekozen antwoordmogelijkheden samengenomen waren).

Voor de 55 uitspraken van LL basis varieerde DIFSTAT van 0,73 tot 1,41 met een gemiddelde van 0,99. Geen van de waarden van DIFSTAT was significant op 1%-niveau.

Van LL aanvullend zijn alleen digitale afnamegegevens beschikbaar waardoor we eventuele DIF ten gevolge de afnamemodus hier niet kunnen vaststellen.

Al met al is er bij geen enkele uitspraak van LK basis en LL basis DIF geconstateerd. De uitspraken functioneren in beide afnamemodi op dezelfde manier. Er is dus geen reden om te veronderstellen dat de afnamemodus invloed heeft gehad op de beantwoording van de vragen (en daarmee op de normering).

6.2.2 Jaargroep

Nagegaan is in hoeverre de uitspraken *Differential Item Functioning* (DIF) ten gevolge van de jaargroep vertonen. Om het aantal analyses en tabellen te beperken, hebben we in de analyse twee groepen onderscheiden. Bij de leerkrachtlijst die is ingevuld voor jaargroep 3 tot en met 8 maken we een onderscheid tussen jongere leerlingen in groep 3, 4 en 5 versus oudere leerlingen in groep 6, 7 en 8. Bij de leerlinglijst die is afgenomen in jaargroep 5, 6, 7 en 8 onderscheiden we eveneens twee groepen: groep 5 en 6 versus groep 7 en 8. Vanwege het grote aantal schalen bespreken we de uitkomsten afzonderlijk voor LK basis, LL basis en LL aanvullend. Om het programma in staat te stellen de DIFSTAT en bijbehorende z-waarden uit te rekenen, zijn waar dat nodig was de beide antwoordmogelijkheden met het kleinste aantal waarnemingen samengenomen (waardoor de vier-puntschaal een drie-puntschaal werd). Een DIFSTAT-waarde groter dan 1 betekent dat jongere leerlingen gegeven een gelijke sociaal-emotionele vaardigheid op de desbetreffende uitspraak relatief hoge scores behalen en een DIFSTAT-waarde kleiner dan 1 dat oudere leerlingen op deze uitspraak relatief hoge scores behalen. De resultaten van de Mantel-Haenzel procedure zijn weergegeven in het middelste gedeelte van tabel 6.19.

Voor de 55 uitspraken van LK basis varieerde DIFSTAT van 0,54 tot 1,22 met een gemiddelde van 0,96. Geen van de waarden van DIFSTAT was significant op 1%-niveau (waarbij we opmerken dat het programma voor Pestgedrag en Hyperactief/impulsief vanwege de scheefheid van de verdelingen geen z-waarde kon berekenen).

Voor de 55 uitspraken van de acht schalen van LK basis varieerde DIFSTAT van 0,63 tot 1,81 met een gemiddelde van 1,04. Geen van de waarden van DIFSTAT was significant op 1%-niveau.

Voor de uitspraken van de tien schalen van LK aanvullend varieerde DIFSTAT van 0,73 tot 1,77 met een gemiddelde van 1,04. Geen van de waarden van DIFSTAT was significant op 1%-niveau (waarbij we opmerken dat het programma de z-waarden van Pestgedrag ondanks het samenvoegen van de beide antwoorden met het kleinste aantal waarnemingen niet kon berekenen).

Al met al is er bij geen enkele uitspraak van VISEON 2.0 bias ten gevolge van de jaargroep geconstateerd. De uitspraken functioneren voor jongere en oudere leerlingen op dezelfde manier.

6.2.3 Geslacht

Nagegaan is in hoeverre de uitspraken *Differential Item Functioning* (DIF) ten gevolge van geslacht vertonen. Een DIFSTAT-waarde groter dan 1 betekent dat jongens gegeven een gelijke sociaal-emotionele vaardigheid op de desbetreffende uitspraak relatief hoge scores behalen en een DIFSTAT-waarde kleiner dan 1 meisjes op deze uitspraak naar verhouding hoge scores behalen. De resultaten van de Mantel-Haenzel procedure zijn weergegeven in het rechter gedeelte van tabel 6.19.

Voor de 55 uitspraken van LK basis varieerde DIFSTAT van 0,67 tot 2,34 met een gemiddelde van 1,05. Geen van de waarden van DIFSTAT was significant op 1%-niveau (waarbij we opmerken dat het programma voor Pestgedrag vanwege de scheefheid van de verdeling geen z-waarden kon berekenen). Voor de 55 uitspraken van de acht schalen van LL basis varieerde DIFSTAT van 0,57 tot 3,01 met een gemiddelde van 1,03. Eén van de waarden van DIFSTAT (3,01) was significant op 1%-niveau ($z = 2,95$). Het betrof de uitspraak 'Een ander kind heeft met mij gevochten (de ander is begonnen)' uit de schaal Gepest worden. Bij deze uitspraak behaalden jongens gegeven een gelijke sociaal-emotionele vaardigheid hogere scores dan meisjes.

Voor de uitspraken van de tien schalen van LL aanvullend varieerde DIFSTAT van 0,45 tot 4,32 met een gemiddelde van 1,15. Helaas kon het programma de z-waarden ondanks het samenvoegen van de beide antwoorden met het kleinste aantal waarnemingen niet berekenen. Van de zeven uitspraken met een DIFSTAT-waarde groter dan 2,00 behoorden er zes tot de schaal Pestgedrag en de overige uitspraak met seksebias was afkomstig uit de schaal Storend gedrag. In alle zeven gevallen behaalden jongens naar verhouding hogere scores dan meisjes.

Al met al is seksebias waarschijnlijk alleen bij de schaal Pestgedrag uit LL aanvullend aanwezig. Voor de overige uitspraken geldt dat deze voor jongens en meisjes dezelfde betekenis hebben.

6.3 Externe structuur

In deze paragraaf bespreken we de externe structuur zoals deze tot uiting komt in de relaties van VISEON 2.0 met de vorige versie van VISEON (eerste generatie), de *Strength and Difficulties Questionnaire* (SDQ), de Kanjervragenlijst en de Schoolvragenlijst (SVL). Er is sprake van convergente validiteit als een aspect relatief hoog samenhangt met verwante aspecten en van discriminante validiteit als er sprake is van een ontbrekende of een relatief lage samenhang met niet verwante of minder verwante aspecten (Campbell & Fiske, 1959).

6.3.1 Relaties met VISEON (eerste generatie)

Leerkrachtlijst VISEON (eerste generatie)

De leerkrachtlijst van de vorige versie van VISEON bevat 44 uitspraken die betrekking hebben op concreet waarneembaar gedrag tijdens de alledaagse gang van zaken binnen de groep (Kamphuis, Rosier, Sluifster, Van der Lubbe, Van Putten & Meijer, 2004). Door het invullen van de lijst krijgt de leerkracht een beeld van vier aspecten van het sociaal-emotioneel functioneren van een leerling:

- zorgvuldige versus onzorgvuldige werkhouding;
- emotionele stabiliteit versus emotionele instabiliteit;
- aangenaam versus storend gedrag;
- sociaal versus teruggetrokken gedrag.

Hierna geven we een inhoudelijke typering van de vier leerkrachtschalen die vrijwel letterlijk ontleend is aan Kamphuis e.a. (2004). Daarbij geven we ook aan welke leerkrachtschalen uit VISEON 2.0 een vergelijkbare meetpretentie hebben. Overeenkomstig de eis van convergente en discriminante validiteit veronderstellen we hogere correlaties tussen inhoudelijk verwante schalen dan tussen inhoudelijk niet of minder verwante schalen.

De schaal *Zorgvuldige werkhouding versus onzorgvuldige werkhouding* heeft betrekking op het beeld dat de leerkracht heeft van het werkgedrag van de leerling; zijn of haar houding ten aanzien van het leren op school. Enkele aspecten zijn: inzet en zorgvuldigheid bij het uitvoeren van taken en het gedurende langere tijd geconcentreerd bezig kunnen zijn. Schalen in VISEON 2.0 met een vergelijkbare meetpretentie zijn Taakgericht gedrag en Motivatie voor leren.

De schaal *Emotionele stabiliteit versus emotionele instabiliteit* richt zich op de emotionele kant van het leerlinggedrag zoals de leerkracht dat waarneemt. In de uitspraken wordt de zekere, emotioneel stabiele

leerling geplaatst tegenover de onzekere leerling die snel van slag is. Enkele aspecten zijn: het al dan niet tegen kritiek kunnen, de mate waarin de leerling zich over iets vervelends heen kan zetten en het al dan niet snel van slag zijn. De schaal uit VISEON 2.0 die hier het meest op lijkt, is Teruggetrokken gedrag (waarbij de gelijkenis niet zozeer betrekking heeft op de meetpretentie als wel op de inhoud van een deel van de uitspraken).

De schaal *Aangenaam gedrag versus storend gedrag* heeft betrekking op het beeld dat de leerkracht heeft van het gedrag van de leerling ten opzichte van zijn of haar klasgenoten en van de leerkracht zelf en de mate waarin de leerling volgens de leerkracht rekening kan houden met de omgeving. Enkele aspecten zijn: het zich al dan niet houden aan regels en afspraken, en het plagen of juist in bescherming nemen van anderen. Schalen in VISEON 2.0 met een soortgelijke meetpretentie zijn Pestgedrag, Storend gedrag en Hyperactief/impulsief.

De schaal *Sociaal gedrag versus teruggetrokken gedrag* heeft betrekking op de mate waarin de leerling volgens de leerkracht gericht is op het participeren in groepsactiviteiten en interactie met de leerkracht en de medeleerlingen. Aan de ene kant gaat het om de aanwezige, sociale leerling die open is en zelfvertrouwen heeft en zelf initiatief neemt tot het leggen van contacten. Aan de andere kant betreft het de gesloten en teruggetrokken leerling die zich schuchter gedraagt, het liefst op de achtergrond blijft en weinig initiatieven neemt. Enkele aspecten zijn: het al dan niet actief contacten leggen in de groep en het al dan niet verlegen zijn van een leerling. Inhoudelijk verwante schalen uit VISEON 2.0 zijn Zelfstandigheid en Teruggetrokken gedrag.

Leerlinglijst VISEON (eerste generatie)

De leerlinglijst richt zich op gedrag dat tot uiting komt in het omgaan met medeleerlingen, met schoolse taken en met de leerkracht. De lijst bevat 42 uitspraken waar een leerling al dan niet mee kan instemmen. In de leerlinglijst komen vijf aspecten van het sociaal-emotioneel functioneren van een leerling aan de orde:

- schoolbeeld: de houding van de leerling ten opzichte van de school;
- relatie met andere leerlingen;
- werkhouding;
- relatie met de leerkracht;
- zelfvertrouwen.

Hierna geven we een inhoudelijke typering van de vijf schalen voor leerlingen waarbij we tevens aangeven welke leerlingsschalen uit VISEON 2.0 een soortgelijke meetpretentie hebben. Overeenkomstig de eis van convergente en discriminante validiteit veronderstellen we hogere correlaties tussen inhoudelijk verwante schalen dan tussen inhoudelijk niet of minder verwante schalen.

De schaal *Schoolbeeld* heeft betrekking op het beeld dat de leerling zich van de school gevormd heeft. Voor het kind moet de school een veilige plaats zijn; een plek waar het zich thuis voelt, zichzelf durft te zijn, zich geaccepteerd weet en vertrouwen in zichzelf en anderen kan hebben. Deze schaal maakt duidelijk of een leerling het leuk vindt op school en zich daar veilig voelt. De schaal in VISEON 2.0 met een vergelijkbare meetpretentie is Schoolbeleving.

De schaal *Relatie met andere leerlingen* heeft betrekking op de houding van de leerling ten opzichte van zijn of haar klasgenoten. Aspecten zijn onder meer: het al dan niet leuk met elkaar omgaan, het contact (durven) maken met de andere leerlingen. Schalen uit VISEON 2.0 die hiermee enige verwantschap vertonen, zijn Sociaal zelfbeeld, Sociale leeromgeving en Gepest worden.

De schaal *Werkhouding* heeft betrekking op de houding van de leerling ten aanzien van het leren op school. Aspecten zijn onder meer: inzet bij het uitvoeren van taken, taakgerichtheid en doorzettingsvermogen. Schalen in VISEON 2.0 met een soortgelijke meetpretentie zijn Motivatie voor leren en Taakgericht gedrag.

De schaal *Relatie met de leerkracht* heeft betrekking op de houding van de leerling ten opzichte van zijn of haar leerkracht. Deze schaal maakt duidelijk hoe de leerling de leerkracht ervaart; of de leerling vindt dat de leerkracht een positieve dan wel negatieve invloed op hem of haar uitoefent. Deze schaal heeft geen duidelijke pendant in VISEON 2.0.

De schaal *Zelfvertrouwen* geeft aan in welke mate het kind vertrouwen heeft in eigen kunnen en opgewassen is tegen de gestelde eisen. Het zelfvertrouwen van een leerling wordt bepaald door zijn of haar zelfbeeld. Inhoudelijk verwante schalen uit VISEON 2.0 zijn Cognitief zelfbeeld, Zelfbeeld van lezen en Zelfbeeld van rekenen.

De matching van VISEON 2.0 met VISEON (eerste generatie)

Ter bepaling van de samenhang met VISEON (eerste generatie) is nagegaan in hoeverre het bestand van het normeringsonderzoek scholen bevatte die gedurende de periode september 2014 tot juni 2015 ook de eerste generatie van VISEON gebruikt hadden. Dit resulteerde in 29 scholen. Als er voor een leerling gegevens van meerdere afnamedata beschikbaar waren, is de afname geselecteerd van de datum die het afnamemoment van het normeringsonderzoek het dichtst benaderde.

Van de 1007 door de leerkracht beoordeelde leerlingen uit het normeringsonderzoek waren er 467 met volledige zekerheid in het gebruikersbestand van de eerste generatie VISEON terug te vinden. Dit wil zeggen dat schoolnummer, jaargroep, geslacht, geboortedatum, voornaam en achternaam in beide bestanden volledig overeenkwamen. Van deze 467 leerlingen hadden er 396 in het normeringsonderzoek deelgenomen aan de digitale afname en 71 aan de papieren afname. VISEON (eerste generatie) is alleen digitaal afgenomen.

Van de 1970 leerlingen die in het normeringsonderzoek een leerlinglijst hadden ingevuld konden er 680 met volledige zekerheid in het gebruikersbestand van de eerste generatie VISEON getraceerd worden (dit wil zeggen dat schoolnummer, jaargroep, geslacht, geboortedatum, voornaam en achternaam volledig overeenkwamen). Van deze 680 leerlingen hadden er 430 in het normeringsonderzoek deelgenomen aan de digitale afname en 250 aan de papieren afname.

De gegevens van de eerste generatie VISEON zijn verzameld bij leerlingen uit jaargroep 3 tot en met 8 (leerkrachtlijst) of jaargroep 5 tot en met 8 (leerlinglijsten). Met behulp van variantieanalyse is nagegaan in hoeverre verschillen tussen jaargroepen in het gemiddeld niveau van sociaal-emotioneel functioneren significant en betekenisvol zijn. De analyse geeft bij acht van de negen schalen significantie te zien ($p < 0,01$). Net als we eerder zagen bij VISEON 2.0 (zie paragraaf 4.2.2) verklaart de jaargroep nauwelijks variantie in het sociaal-emotioneel functioneren. De waarde van *Eta squared* bedraagt voor Zorgvuldige werkhouding 0,005, voor Aangenaam versus Storend gedrag 0,005, voor Emotionele stabiliteit 0,010, voor Sociaal versus Teruggetrokken gedrag 0,004, voor Zelfvertrouwen 0,005, voor Werkhouding 0,006, voor Relatie met leerkracht 0,003, voor Relatie met andere leerlingen 0,024 en tot slot voor Schoolbeeld 0,009. Net als bij VISEON 2.0 het geval was, maken we dus geen grove fout als we jaargroep in de analyse van de correlaties tussen de schalen van VISEON eerste generatie en VISEON 2.0 negeren.

Enkele beschrijvende statistische gegevens van VISEON (eerste generatie)

Enkele beschrijvende gegevens voor de negen schalen van de eerste generatie VISEON zijn weergegeven in tabel 6.20. Te zien is onder meer dat alle schalen voldoende betrouwbaar zijn om onderzoek naar de relatie met VISEON 2.0 te rechtvaardigen.

Tabel 6.20 Enkele beschrijvende statistische gegevens voor VISEON (eerste generatie)

	N	Min	Max	Gem	Stddev	Scheefheid	Kurtosis	Aantal uitspraken	Alfa
Leerkracht									
Zorgvuldige werkhouding	466	13,28	155,32	107,60	31,82	-0,02	-0,69	11	0,92
Emotionele stabiliteit	467	27,24	123,28	82,45	19,66	0,31	-0,24	12	0,85
Aangenaam versus storend gedrag	467	72,84	193,52	154,85	28,71	-0,11	-0,85	11	0,85
Sociaal versus teruggetrokken gedrag	465	8,92	100,52	67,87	18,14	-0,28	0,00	10	0,87
Leerling									
Schoolbeeld	646	24,76	111,00	80,93	16,86	-0,02	0,27	8	0,77
Relatie met andere leerlingen	646	19,56	101,28	78,58	15,88	-0,33	0,08	9	0,74
Werkhouding	646	26,52	134,44	87,74	15,58	0,28	1,05	9	0,75
Relatie met leerkracht	646	21,64	126,24	100,86	17,61	0,02	-0,43	9	0,81
Zelfvertrouwen	645	16,04	113,16	74,14	17,69	-0,04	0,61	7	0,74

Tabel 6.21 toont de correlaties tussen de schalen van de vorige en de nieuwe versie van VISEON. Het aantal voor analyse beschikbare leerlingen verschilt van schaal tot schaal. De correlaties van de leerkrachtschalen van VISEON (eerste generatie) met de schaal Hyperactief/impulsief uit LK basis zijn gebaseerd op oordelen over 277 leerlingen en voor de correlaties met de overige schalen van LK basis zijn het er minimaal 464. De correlaties tussen de leerlingsschalen van VISEON (eerste generatie) en LL basis zijn gebaseerd op de antwoorden van tussen de 608 en 653 leerlingen en bij de correlaties met LL aanvullend gaat het om tussen de 356 en 414 leerlingen.

Tabel 6.21 Correlaties tussen de schalen van VISEON (eerste generatie) en VISEON 2.0

	Leerkracht				Leerling				
	Zorgvuldige werkhouding	Emotionele stabiliteit	Aangenaam versus storend gedrag	Sociaal versus teruggetrokken gedrag	Schoolbeeld	Relatie met andere leerlingen	Werkhouding	Relatie met leerkracht	Zelfvertrouwen
LK basis									
pro-sociaal gedrag	0,33	0,32	0,37	0,19					
pestgedrag	-0,27	-0,25	-0,49	0,12					
storend gedrag	-0,44	-0,23	-0,61	0,24					
teruggetrokken gedrag	-0,21	-0,50	-0,04	-0,45					
taakgericht gedrag	0,61	0,29	0,45	-0,05					
motivatie voor leren	0,58	0,34	0,40	0,10					
zelfstandigheid	0,21	0,34	-0,08	0,58					
hyperactief/impulsief	-0,48	-0,31	-0,59	0,22					
LL basis									
sociaal zelfbeeld					0,26	0,50	0,15	0,22	0,18
cognitief zelfbeeld					0,32	0,20	0,38	0,35	0,46
schoolbeleving					0,55	0,40	0,20	0,33	0,11
gepest worden					-0,35	-0,41	-0,21	-0,26	-0,05
motivatie voor leren					0,37	0,18	0,46	0,36	0,35
sociale leeromgeving					0,30	0,31	0,19	0,22	0,06
onveilige leeromgeving					-0,24	-0,27	-0,16	-0,17	-0,02
onordelijke leeromgeving					-0,34	-0,34	-0,24	-0,21	-0,09
LL aanvullend									
pro-sociaal gedrag					0,44	0,23	0,33	0,40	0,20
pestgedrag					-0,28	-0,09	-0,36	-0,29	-0,11
storend gedrag					-0,31	-0,03	-0,59	-0,37	-0,28
taakgericht gedrag					0,36	0,05	0,57	0,40	0,31
motivatie voor lezen					0,36	-0,01	0,33	0,27	0,17
motivatie voor rekenen					0,37	0,02	0,32	0,34	0,25
zelfbeeld lezen					0,23	0,02	0,27	0,19	0,35
zelfbeeld rekenen					0,04	0,00	0,24	0,07	0,45
spreekangst					-0,18	-0,27	-0,18	-0,12	-0,28
toetsangst					0,44	0,23	0,33	0,40	0,20

LK basis versus de leerkrachtlIJst van VISEON (eerste generatie)

Hiervoor hebben we aangegeven welke leerkrachtschalen van VISEON en VISEON 2.0 soortgelijke begrippen beogen te meten en derhalve een relatief sterke onderlinge samenhang zouden moeten vertonen. Inspectie van tabel 6.21 brengt aan het licht dat de correlaties tussen inhoudelijk verwante schalen tot de hoogste correlaties behoren en bovendien vrijwel zonder uitzondering hoger zijn dan die met niet of minder verwante schalen. Zo hangt de schaal *Zorgvuldige werkhouding* uit de eerste generatie VISEON relatief hoog samen met Taakgericht gedrag ($r = 0,61$) en Motivatie voor leren ($r = 0,58$) uit VISEON 2.0. Evenzo zien we een relatief sterke samenhang tussen *Emotionele stabiliteit* en Teruggetrokken gedrag ($-0,50$), tussen *Aangenaam versus storend gedrag* en Pestgedrag ($r = -0,49$), Storend gedrag ($r = -0,61$) en Hyperactief/impulsief ($r = -0,59$) en tussen *Sociaal versus teruggetrokken gedrag* en Teruggetrokken gedrag ($r = -0,45$) en Zelfstandigheid ($r = 0,58$). Het patroon van correlaties geeft aanleiding tot vertrouwen in de convergente en discriminante validiteit van sociaal-emotioneel functioneren zoals gemeten met de leerkrachtlIJst van VISEON 2.0.

LL basis en LL aanvullend versus de leerlinglijst van VISEON (eerste generatie).

Overeenkomstig de van tevoren geformuleerde verwachtingen hangt *Schoolbeeld* uit de eerste generatie VISEON duidelijk sterker samen met *Schoolbeleving* uit VISEON 2.0 ($r = 0,55$) dan met de overige schalen van deze nieuwe versie.

Hetzelfde geldt voor de samenhang tussen *Relatie met andere leerlingen* en *Sociaal zelfbeeld* ($r = 0,50$), *Schoolbeleving* ($r = 0,40$) en *Gepest worden* ($r = -0,41$). Eveneens conform de verwachtingen correleert *Werkhouding* relatief hoog met *Taakgericht gedrag* ($r = 0,57$) en *Storend gedrag* ($r = -0,59$).

In overeenstemming met het onderscheid in algemene en vakspecifieke motivaties zijn de correlaties van *Werkhouding* met *Taakgericht gedrag* ($r = 0,57$) wat hoger dan die met *Motivatie voor lezen* en *Motivatie voor rekenen* (r respectievelijk $0,33$ en $0,32$).

Aangezien de schaal *Relatie met leerkracht* geen duidelijke pendant heeft in VISEON 2.0, hebben wij met betrekking tot de convergente en discriminante validiteit geen verwachtingen geformuleerd. Dat een hoge score voor *Relatie met leerkracht* samengaat met een hoge motivatie voor leren ($r = 0,36$), meer pro-sociaal gedrag ($r = 0,40$) en meer taakgericht gedrag ($r = 0,40$).

Gezien de operationalisatie van de schaal *Zelfvertrouwen* is het niet vreemd dat deze schaal naar verhouding hoog correleert met *Cognitief zelfbeeld* ($r = 0,47$), *Zelfbeeld rekenen* ($r = 0,46$) en *Motivatie voor leren* ($r = 0,35$).

Deze resultaten ondersteunen de convergente en discriminante validiteit van sociaal-emotioneel functioneren zoals gemeten met de leerlinglijst van VISEON 2.0.

6.3.2 Relaties met de *Strength and Difficulties Questionnaire* (SDQ)

In het normeringsonderzoek is de prevalentie van pro-sociaal gedrag en probleemgedrag vastgesteld met de *Strength and Difficulties Questionnaire* (SDQ) van Goodman (1977). De SDQ is gebaseerd op DSM-classificatie (*Diagnostic and Statistical Manual of mental disorders*) van psychosociale problemen bij jongeren. Uit onderzoek blijkt dat de SDQ voor het doel van grootschalig peilingsonderzoek een betrouwbaar, efficiënt en hanteerbaar instrument is (o.a. Muris, Meesters & Van den Berg, 2003; Vogels, Crone, Hoekstra & Reijneveld, 2005; Van Roy, Veenstra & Clench-Aas, 2008). De Nederlandse versie van de SDQ wordt gezien als een indicator van de sociaal-emotionele ontwikkeling (Diepenmaat, Van Eijnden, Janssens, Loomans & Stone, 2014). In het normeringsonderzoek is alleen de leerkrachtversie van de SDQ gebruikt.

De SDQ is bedoeld om kinderen met een hoog risico op psychosociale problemen te signaleren. De vragenlijst meet de aanwezigheid van deze problemen en wat de sterke kanten van het kind genoemd wordt. De vragenlijst bevat in totaal 25 uitspraken over de aspecten Gedragsproblemen, Hyperactiviteit/aandachttekort, Problemen met leeftijdsgenoten, Pro-sociaal gedrag en Emotionele problemen. In de inleiding op de vragen is de leerkrachten verzocht hun antwoorden te baseren op het gedrag van het kind tijdens de laatste zes maanden of het huidige schooljaar. De antwoordmogelijkheden zijn 'Niet waar' (1), 'Een beetje waar' (2) en 'Zeer waar' (3). De SDQ meet de volgende aspecten van het sociaal-emotioneel functioneren:

- Gedragsproblemen (externaliserend), bijvoorbeeld 'Heeft vaak driftbuien of woede-uitbarstingen';
- Hyperactiviteit/aandachttekort (specifiek), bijvoorbeeld 'Rusteloos, overactief, kan niet lang stilzitten';
- Problemen met leeftijdsgenoten, bijvoorbeeld 'Nogal op zichzelf, neigt er toe alleen te spelen';
- Pro-sociaal gedrag, bijvoorbeeld 'Houdt rekening met gevoelens van anderen';
- Emotionele symptomen (internaliserend), bijvoorbeeld 'Klaagt vaak over hoofdpijn, buikpijn of misselijkheid'.

Op basis van theoretische noties verwachten we dat de schaal *Gedragsproblemen* uit de SDQ sterk positief zal samenhangen met de schalen *Pestgedrag* en *Storend gedrag* en sterk negatief met *Pro-sociaal gedrag* uit VISEON 2.0. Tevens verwachten we een sterke samenhang tussen *Hyperactief/aandachttekort* en *Hyperactief/impulsief* en *Storend gedrag*, tussen *Problemen met leeftijdsgenoten* en onder meer *Pro-sociaal gedrag*, tussen de gelijknamige schalen voor *Pro-sociaal gedrag* en tussen *Emotionele symptomen* en *Zelfstandigheid*.

Voor elk van de vijf schalen is een somscore verkregen door de scores over de vijf uitspraken te sommeren (theoretisch bereik van 5 tot en met 15). Tabel 6.22 toont onder meer de gemiddelden, de standaarddeviaties en de betrouwbaarheid (Cronbachs alfa). De vijf SDQ-schalen blijken voor het doel van deze validatiestudie voldoende betrouwbaar. Wat opvalt is dat de scoreverdelingen met uitzondering van Hyperactiviteit/aandachttekort extreem scheef en gepiekt zijn (wat gezien de focus op relatief ernstig probleemgedrag weinig verbazing wekt).

Tabel 6.22 Enkele beschrijvende statistische gegevens voor de SDQ

	N	Min	Max	Gem	Stddev	Scheefheid	Kurtosis	Aantal uitspraken	Alfa
Gedragsproblemen	118	5	12	5,92	1,59	2,12	4,61	5	0,73
Hyperactiviteit/aandachttekort	119	5	15	7,39	2,84	0,97	-0,27	5	0,89
Problemen met leeftijdsgenoten	119	5	15	6,18	1,75	2,02	5,39	5	0,73
Pro-sociaal gedrag	118	5	15	6,41	1,90	1,60	3,03	5	0,80
Emotionele symptomen	119	5	14	6,03	1,77	2,30	5,77	5	0,80

Tabel 6.23 toont de correlaties tussen de schalen van de SDQ en VISEON 2.0. De tabel laat zien dat de correlaties tussen inhoudelijk verwante schalen duidelijk hoger zijn dan die tussen niet of minder verwante schalen. Zoals verwacht vertoont de SDQ-schaal *Gedragsproblemen* een sterk positieve samenhang met zowel Pestgedrag als Storend gedrag (r respectievelijk 0,68 en 0,62) terwijl deze schaal sterk negatief correleert met Pro-sociaal gedrag ($r = -0,69$).

Eveneens overeenkomstig de verwachtingen zien we een sterke samenhang van *Hyperactief/aandachttekort* met Hyperactief/impulsief ($r = 0,78$) en Storend gedrag ($r = 0,69$).

Hetzelfde geldt voor de relatie tussen *Problemen met leeftijdsgenoten* en Pro-sociaal gedrag ($r = -0,54$), tussen de beide gelijknamige schalen voor Pro-sociaal gedrag ($r = -0,75$) en tussen *Emotionele symptomen* en Teruggetrokken gedrag ($r = 0,57$).

De correlaties van de SDQ-schalen met die van LL basis en LL aanvullend zijn weliswaar wat lager dan met LK basis maar nog steeds goed interpreteerbaar in het licht van wat op basis van theoretische noties verwacht mag worden. Voor de redenen waarom de leerkrachtschalen van de SDQ en VISEON 2.0 hoger met elkaar zullen samenhangen dan met de leerlingsschalen van VISEON 2.0 verwijzen wij naar de discussie aan het einde van bijlage 1.

Deze resultaten geven aanleiding tot vertrouwen in de convergente en discriminante validiteit van sociaal-emotioneel functioneren zoals gemeten met VISEON 2.0.

Tabel 6.23 Correlaties tussen de schalen van de SDQ en VISEON 2.0

	Gedrags- problemen	Hyperactiviteit/ aandachttekort	Problemen met leeftijdsgenoten	Pro-sociaal gedrag	Emotionele symptomen
LK basis					
pro-sociaal gedrag	-0,69	-0,56	-0,54	-0,75	-0,38
pestgedrag	0,68	0,53	0,45	0,70	0,33
storend gedrag	0,62	0,69	0,29	0,65	0,34
teruggetrokken gedrag	0,17	0,22	0,56	0,30	0,57
taakgericht gedrag	-0,35	-0,64	-0,07	-0,36	-0,18
motivatie voor leren	-0,47	-0,66	-0,36	-0,52	-0,39
zelfstandigheid	-0,06	-0,20	-0,25	-0,16	-0,23
hyperactief/impulsief	0,69	0,78	0,41	0,72	0,42
LL basis					
sociaal zelfbeeld	-0,09	-0,16	-0,21	-0,19	-0,06
cognitief zelfbeeld	-0,35	-0,35	-0,16	-0,27	-0,09
schoolbeleving	-0,30	-0,29	-0,21	-0,22	-0,10
gepest worden	0,36	0,22	0,37	0,26	0,30
motivatie voor leren	-0,40	-0,40	-0,14	-0,48	-0,03
sociale leeromgeving	-0,13	-0,07	-0,05	-0,13	-0,05
onveilige leeromgeving	0,07	-0,05	0,15	-0,01	0,21
onordelijke leeromgeving	0,26	0,16	0,34	0,24	0,38
LL aanvullend					
pro-sociaal gedrag	-0,37	-0,28	-0,24	-0,45	-0,11
pestgedrag	0,47	0,34	0,24	0,44	0,23
storend gedrag	0,55	0,51	0,21	0,45	0,26
taakgericht gedrag	-0,43	-0,46	-0,13	-0,47	-0,06
motivatie voor lezen	-0,19	-0,30	-0,01	-0,18	0,10
motivatie voor rekenen	-0,17	-0,21	-0,19	-0,29	-0,13
zelfbeeld lezen	-0,12	-0,17	-0,03	-0,01	0,02
zelfbeeld rekenen	0,05	0,01	-0,11	-0,06	-0,04
spreekangst	-0,02	-0,01	0,08	-0,06	-0,07
toetsangst	-0,05	-0,18	0,05	-0,16	-0,09

6.3.3 Relaties met de Kanjervragenlijst

De Kanjervragenlijst (Vliek, Riet & Weide, 2012; Vliek, 2015) is een instrument voor het volgen van het sociaal-emotioneel functioneren van leerlingen in de leeftijd van acht tot dertien jaar. De Kanjervragenlijst is in 2013 door de COTAN goedgekeurd en de Inspectie adviseert het gebruik van deze test sinds augustus 2014. Het doel is leerkrachten in staat te stellen om leerlingen beter te begrijpen, op de juiste manier te begeleiden en te ondersteunen bij het sociale functioneren in de klas. Evenmin als VISEON 2.0 is de Kanjervragenlijst gericht op het diagnosticeren en voorspellen van toekomstig probleemgedrag. In het normeringsonderzoek zijn drie van de vier schalen van de Kanjervragenlijst afgenomen bij 208 leerlingen in groep 5, 6, 7 en 8. Het gaat om de volgende schalen:

- Hulpvaardig sociaal;
- Negatieve intenties;
- Onrustig verstorend.

Hulpvaardig sociaal

Met de schaal *Hulpvaardig sociaal* wordt gemeten in hoeverre de leerling zelf het idee heeft pro-sociaal gedrag te vertonen, zich hier competent in te voelen en dit ook te willen. In de Kanjervragenlijst is het concept *Hulpvaardig sociaal* als volgt geoperationaliseerd:

- Gevoelens van sociale competentie (Ik kan anderen goed helpen, Ik kan veel);
- Willen bijdragen aan het welbevinden van een ander (Ik wil eerlijk zijn op school, Ik vind het leuk om klasgenoten blij te maken);

- Tonen van sociale verantwoordelijkheid (Ik doe aardig, Ik ben te vertrouwen, Ik help kinderen in de klas).

Negatieve intenties

Onder *Negatieve intenties* wordt verstaan: de wil om een ander iets vervelends aan te doen, of de wil tot agressie. *Negatieve intenties* is als volgt geoperationaliseerd:

- Het kind wil andere kinderen pesten of uitlachen.
- Het kind wil gemeen doen of de strijd opzoeken.
- Het kind wil de leerkracht storen of dwarsbomen.

Onrustig verstorend

Onder *Onrustig verstorend* wordt impulsief gedrag verstaan dat storend is in de klas en daarmee niet aan de verwachtingen van de leerkracht voldoet. In de Kanjervragenlijst is het concept *Onrustig verstorend* als volgt geoperationaliseerd:

- Impulsief gedrag (door de klas roepen, rare geluiden maken, niet rustig zijn);
- Het niet voldoen aan de verwachtingen van de leerkracht (brutaal zijn, een boze reactie van de leerkracht op het gedrag uitlokken, het doen van dingen die niet mogen).

Op basis van theoretische noties verwachten we onder meer een relatief sterke en positieve samenhang tussen de Kanjer-schaal *Hulpvaardig sociaal* en Pro-sociaal gedrag en een naar verhouding sterk negatieve relatie van *Hulpvaardig sociaal* met Pestgedrag en Storend gedrag. Verder verwachten we dat *Negatieve intenties* vooral zal samenhangen met Pestgedrag en Storend gedrag en dat *Onrustig verstorend* een sterke samenhang met Hyperactief/impulsief en Storend gedrag te zien zal geven.

Tabel 6.24 toont onder meer de gemiddelden, de standaarddeviaties en de betrouwbaarheid (Cronbachs alfa).

Tabel 6.24 Enkele beschrijvende statistische kenmerken voor drie schalen van de Kanjervragenlijst

	N	Min	Max	Gem	Stddev	Scheefheid	Kurtosis	Aantal uitspraken	Alfa
Hulpvaardig sociaal	200	23	40	34,43	4,17	-0,62	-0,48	10	0,82
Negatieve intenties	208	8	12	8,33	0,79	2,70	7,16	8	0,47
Onrustig verstorend	200	8	25	13,79	3,03	0,16	-0,14	8	0,27

De spreiding van de antwoorden voor *Negatieve intenties* blijkt zeer klein terwijl de verdeling zeer scheef en gepiekt is (zie ook Vliek, Riek & Weide, 2012). De betrouwbaarheid van *Hulpvaardig sociaal* is goed, maar die van *Negatieve intenties* en *Onrustig verstorend* is zeer laag. Vanwege de sterk van normaliteit afwijkende verdelingen van *Negatieve intenties* en de zeer lage betrouwbaarheid van twee van de drie schalen rapporteren we in tabel 6.25 de procentuele verdeling van de antwoorden over de antwoordmogelijkheden en de item-restcorrelatie (Rir). De Rir's van de uitspraken van *Onrustig verstorend* zijn (uiteraard) berekend na hercodering van de antwoorden op de drie positief geformuleerde uitspraken. De zeer geringe spreiding van de antwoorden op de vragen over *Negatieve intenties* vindt zijn oorzaak in het gegeven dat vrijwel alle leerlingen hier voor de antwoordmogelijkheid 'Helemaal niet waar' kiezen. Bij de schaal *Onrustig verstorend* blijken de item-restcorrelaties alle positief maar zonder uitzondering laag tot zeer laag. Omdat de betrouwbaarheden voor *Negatieve intenties* en *Onrustig en verstorend* aanmerkelijk lager zijn dan de door Vliek, Riek en Weide (2012) gerapporteerde indices moeten de uitkomsten voor wat betreft deze twee schalen met voorzichtigheid geïnterpreteerd worden.

Tabel 6.25 Procentuele verdeling van de antwoorden over de antwoordmogelijkheden en item-restcorrelaties voor de drie schalen van Kanjervragenlijst

	helemaal niet waar	een beetje waar	best waar	helemaal waar	Rir
Hulpvaardig sociaal					
Ik ben te vertrouwen	0%	4%	32%	65%	0,43
Ik kan anderen goed helpen	0%	8%	44%	47%	0,57
Ik doe aardig	0%	4%	50%	45%	0,37
Ik help kinderen in de klas	0%	14%	39%	47%	0,63
Ik kan veel	0%	16%	41%	42%	0,42
Als ik iets beloof aan mijn juf of meester, dan doe ik het ook	0%	10%	28%	62%	0,54
Ik help als iemand valt of verdrietig is	1%	16%	30%	53%	0,49
Ik vind het leuk om klasgenoten blij te maken	2%	6%	31%	61%	0,42
Ik wil eerlijk zijn op school	0%	5%	33%	62%	0,52
Ik wil dat de juf of meester trots op me is	2%	12%	25%	62%	0,57
Negatieve intenties					
Ik wil anderen uitlachen	94%	6%	0%	0%	0,34
Ik wil kinderen pesten	99%	1%	0%	0%	0,15
Ik wil ruzie	99%	1%	0%	0%	0,15
Ik wil stout zijn	95%	5%	0%	0%	0,32
Ik vind het leuk om kinderen aan het huilen te krijgen	100%	0%	0%	0%	0,00
Ik vind het leuk om gemeen te doen op school	97%	3%	0%	0%	0,22
Ik vind het leuk om kinderen bang te maken	89%	9%	1%	0%	0,24
Ik wil de juf of meester storen in de klas	94%	6%	0%	0%	0,33
Onrustig verstorend					
Ik maak rare geluiden in de klas	75%	21%	3%	2%	0,06
Ik roep door de klas	69%	26%	3%	1%	0,06
Ik doe brutaal op school	85%	14%	0%	0%	0,18
Ik doe dingen die niet mogen op school	77%	21%	2%	0%	0,00
Ik gedraag me zoals het hoort*	5%	15%	37%	44%	0,14
De juf of meester is vaak boos op mij	75%	20%	3%	2%	0,02
Ik ben rustig in de klas*	8%	22%	36%	34%	0,21
Ik let goed op in de les*	1%	15%	43%	40%	0,14

* Rir berekend na spiegelen van de scores volgens (4=1) (3=2) (2=3) (1=4).

Tabel 6.26 toont de correlaties tussen de drie schalen van de Kanjervragenlijst en VISEON 2.0. De correlaties met LK basis zijn gebaseerd op de antwoorden van minimaal 88 leerlingen en voor de correlaties met LL basis en LL aanvullend zijn dat er ten minste 136.

Tabel 6.26 Samenhang van de drie schalen van de Kanjervragenlijst met VISEON 2.0 (88 < N < 154)

	Hulpvaardig sociaal	Negatieve intenties	Onrustig verstorend
LK basis			
pro-sociaal gedrag	0,48	-0,35	-0,04
pestgedrag	-0,52	0,43	0,27
storend gedrag	-0,47	0,48	0,11
teruggetrokken gedrag	-0,12	0,07	-0,01
taakgericht gedrag	0,18	-0,35	-0,16
motivatie voor leren	0,33	-0,43	-0,05
zelfstandigheid	0,05	-0,01	0,13
hyperactief/impulsief	-0,46	0,44	0,15
LL basis			
sociaal zelfbeeld	0,21	-0,12	0,11
cognitief zelfbeeld	0,37	-0,20	0,21
schoolbeleving	0,23	-0,24	0,05
gepest worden	-0,07	0,20	0,03
motivatie voor leren	0,53	-0,36	0,26
sociale leeromgeving	0,24	-0,10	0,24
onveilige leeromgeving	-0,13	0,22	-0,08
onordelijke leeromgeving	-0,12	0,04	-0,01
LL aanvullend			
pro-sociaal gedrag	0,56	-0,19	0,21
pestgedrag	-0,38	0,40	-0,05
storend gedrag	-0,45	0,37	-0,19
taakgericht gedrag	0,64	-0,31	0,26
motivatie voor lezen	0,31	-0,18	0,17
motivatie voor rekenen	0,47	-0,23	0,20
zelfbeeld lezen	0,17	-0,15	0,18
zelfbeeld rekenen	0,16	-0,04	0,16
spreekangst	-0,13	0,03	-0,18
toetsangst	0,01	0,05	-0,12

Overeenkomstig de van tevoren geformuleerde verwachtingen correleert *Hulpvaardig sociaal* relatief sterk en positief met Pro-sociaal gedrag zoals gemeten met LK basis ($r = 0,48$) en LL basis ($r = 0,56$), maar naar verhouding sterk negatief met de leerkracht- en leerlingschaal voor Pestgedrag (r respectievelijk $-0,52$ en $-0,38$) en de leerkracht- en leerlingschaal voor Storend gedrag (r respectievelijk $-0,47$ en $-0,45$).

Niet verwacht maar toch begrijpelijk is de relatief hoge samenhang van *Hulpvaardig sociaal* met Motivatie voor leren ($r = 0,53$), Taakgericht gedrag ($r = 0,64$) en Motivatie voor rekenen ($r = 0,47$) in de leerlinglijst. Tegelijkertijd zien we een relatief sterke positieve samenhang van *Negatieve intenties* met leerkracht- en leerlingschaal voor Pestgedrag (r respectievelijk $0,43$ en $0,40$) en de leerkracht- en leerlingschaal voor Storend gedrag (r respectievelijk $0,48$ en $0,37$).

De veronderstelde sterke relatie tussen enerzijds *Onrustig verstorend* en anderzijds Hyperactief/impulsief en Storend gedrag blijft uit, maar dat is waarschijnlijk toe te schrijven aan de zeer lage betrouwbaarheid waarmee *Onrustig verstorend* in onze responsgroep gemeten is.

Ondanks de soms scheve verdeling en lage betrouwbaarheid van de Kanjervragenlijst in onze responsgroep, geven de resultaten aanleiding tot vertrouwen in de convergente en discriminante validiteit van sociaal-emotioneel functioneren zoals gemeten met VISEON 2.0.

6.3.4 Relaties met twee schalen uit de Schoolvragenlijst (SVL)

In het normeringsonderzoek zijn twee schalen van de Schoolvragenlijst (Smits & Vorst, 1990) afgenomen bij 206 leerlingen uit groep 5, 6, 7 en 8:

- Sociale vaardigheid;
- Zelfvertrouwen bij proefwerken.

Anders dan het label van de schaal doet vermoeden, hebben de uitspraken van de SVL-schaal *Sociale vaardigheid* een grotere gelijkenis met gangbare conceptualisaties en operationalisaties van spreekangst (vgl. Kuhlemeier, Van den Bergh & Rijlaarsdam, 2002) dan met de VISEON-schalen uit het domein Sociaal gedrag. Voor de uitspraken van de SVL-schaal *Zelfvertrouwen bij proefwerken* geldt dat deze enige gelijkenis vertonen met die van Toetsangst uit VISEON 2.0. We verwachten dan ook dat *Sociale vaardigheid* sterker met Spreekangst zal correleren dan met andere schalen van VISEON 2.0 en hetzelfde geldt voor de samenhang tussen *Zelfvertrouwen bij proefwerken* en Toetsangst.

De uitspraken van de beide SVL-schalen en de verdeling van antwoorden van de leerlingen over de antwoordmogelijkheden zijn weergegeven in tabel 6.27.

Tabel 6.27 Procentuele verdeling van de antwoorden op de uitspraken over Sociale vaardigheid en Zelfvertrouwen bij proefwerken (rijpercentages optellend tot honderd procent)

	dat is zo	dat weet ik niet	dat is niet zo
Sociale vaardigheid			
Ik durf best in een volle klas iets aan de juf of meester te vragen.	91	7	2
Ik vind het vervelend om vóór de klas te moeten komen.	11	13	75
Als ik namens mijn klas iets moet zeggen tegen de juf of meester, dan durf ik dat best.	68	27	5
Als een juf of meester boos op mij is, voor iets dat ik niet gedaan heb, dan durf ik dat best te zeggen.	83	13	4
Ik vind het eng om een kamer binnen te gaan waar veel mensen zijn.	7	14	80
Als de juf of meester iets aan mij vraagt in de klas, voel ik me verlegen.	8	8	84
Als we in de klas met zijn allen ergens over praten, durf ik best te zeggen wat ik er van denk.	80	14	7
Ik word verlegen, als plotseling iedereen in de klas naar me kijkt.	26	22	52
Zelfvertrouwen bij proefwerken			
Als ik een proefwerk gemaakt heb, denk ik meestal dat ik het wel goed gedaan heb.	73	21	6
Vlak voordat ik een proefwerk krijg, ben ik bang dat ik door de zenuwen iets zal vergeten.	25	15	59
Ik ben erg bang voor proefwerken.	2	8	89
Als ik mijn werk goed geleerd heb en ik krijg een beurt voor een cijfer, voel ik me meestal zeker van mezelf.	74	22	4
Als de juf of meester zegt dat we een beurt krijgen voor een cijfer, dan ben ik bang	10	22	68
Bij een proefwerk ben ik meestal rustig en kan ik met aandacht werken.	78	11	11
Als we de cijfers terugkrijgen van een proefwerk, voel ik me erg zenuwachtig.	30	12	58
Als ik mijn best heb gedaan, lever ik mijn werk in met een gevoel dat ik het wel voldoende heb gemaakt.	82	14	3

Na spiegeling van positief geformuleerde uitspraken is voor beide schalen een somscore verkregen door de scores over de acht uitspraken te sommeren (theoretisch bereik van 8 tot en met 24). Tabel 6.28 toont onder meer de gemiddelden, de standaarddeviaties en de betrouwbaarheid (Cronbachs alfa).

Tabel 6.28 Enkele beschrijvende kenmerken voor de schalen Sociale vaardigheid en Zelfvertrouwen bij proefwerken van de SVL

	N	Min	Max	Gem	Stddev	Scheef- heid	Kurtosis	Aantal items	Alfa
Sociale vaardigheid	206	16	28	25,42	2,62	-1,32	1,45	8	0,66
Zelfvertrouwen bij proefwerken	206	14	28	24,90	3,20	-1,22	1,38	8	0,76

In de SVL-handleiding wordt vermeld dat de SVL kan worden afgenomen vanaf groep 6 van het basisonderwijs, terwijl wij de schalen ook aan leerlingen van groep 5 hebben voorgelegd. Een hier niet gerapporteerde analyse per groep laat zien de gemiddelden, standaarddeviaties en betrouwbaarheden nauwelijks van groep tot groep verschillen, zodat samenvoeging van de gegevens van groep 5 met die van groep 6, 7 en 8 gerechtvaardigd is.

Tabel 6.29 toont de correlaties tussen de SVL en VISEON 2.0. De correlaties met LL basis en LL aanvullend zijn gebaseerd op de antwoorden van minimaal 138 leerlingen en voor de correlaties met LK basis zijn het er minimaal 88.

Tabel 6.29 Samenhang tussen twee schalen van de Schoolvragenlijst en VISEON 2.0 (88 < N < 153)

	Sociale vaardigheid	Zelfvertrouwen bij proefwerken
LK basis		
pro-sociaal gedrag	-0,08	-0,19
pestgedrag	0,14	0,20
storend gedrag	0,18	0,26
teruggetrokken gedrag	-0,05	0,07
taakgericht gedrag	-0,01	0,08
motivatie voor leren	-0,07	-0,06
zelfstandigheid	0,03	0,03
hyperactief/impulsief	0,13	0,22
LL basis		
sociaal zelfbeeld	0,22	0,08
cognitief zelfbeeld	0,16	0,30
schoolbeleving	0,22	0,15
gepest worden	-0,09	-0,12
motivatie voor leren	0,24	0,15
sociale leeromgeving	0,10	0,14
onveilige leeromgeving	-0,02	-0,12
onordelijke leeromgeving	-0,06	0,04
LL aanvullend		
pro-sociaal gedrag	0,14	-0,04
pestgedrag	0,01	-0,12
storend gedrag	0,08	-0,04
taakgericht gedrag	0,08	0,14
motivatie voor lezen	-0,05	-0,02
motivatie voor rekenen	0,20	0,11
zelfbeeld lezen	0,05	0,08
zelfbeeld rekenen	0,30	0,30
spreekangst	-0,49	-0,30
toetsangst	-0,34	-0,56

Zoals verwacht correleert *Sociale vaardigheid* duidelijk hoger met Spreekangst ($r = -0,49$) dan met alle andere schalen van VISEON 2.0 en hetzelfde geldt voor de correlatie tussen *Zelfvertrouwen bij proefwerken* en Toetsangst ($r = -0,56$). Deze resultaten vormen een ondersteuning voor de discriminante validiteit van spreek- en toetsangst zoals gemeten met VISEON 2.0.

6.4 Relevante verschillen tussen groepen

6.4.1 Geslacht

Eerder constateerden wij dat de afnamemodus en de jaargroep niet of nauwelijks van belang zijn voor het niveau van het sociaal-emotioneel functioneren van de leerlingen (zie paragraaf 4.2.1 en 4.2.2). Het geslacht van de leerling blijkt daarentegen wel een belangrijkste 'voorspeller' van het sociaal-emotionele functioneren. In de Peiling Sociale Competentie (Kuhlemeier, Van Boxtel & Van Til, 2002) bleek de invloed van geslacht zelfs beduidend groter dan die van leerlinggewicht, doubleren, gezinssamenstelling, sociaal-economische status, geboorteland van de ouders, thuistaal en religieuze of levensbeschouwelijke affiliatie. Daarbij bleek het sociaal-emotioneel functioneren van jongens overwegend minder gunstig dan van meisjes. De verschillen in het voordeel van meisjes waren het grootst bij externaliserende problematiek en relatief klein bij zelfbeeld en internaliserende problematiek. Deze resultaten zijn in overeenstemming met prevalentie-onderzoek (o.a. Scholte & Van der Ploeg, 2007) waarbij is vastgesteld dat externaliserende

problematiek (ADHD, ASS, antisociale en gedragsstoornissen) bij jongens veel vaker voorkomt dan bij meisjes en dat de verschillen qua internaliserende problematiek (angst, depressie) doorgaans veel kleiner zijn. Eerder zagen we dat de schalen Pro-sociaal gedrag, Pestgedrag en Storend gedrag sterk samenhangen met de SDQ-schalen Gedragsproblemen en Hyperactiviteit/aandachttekort die externaliserend probleemgedrag meten (zie paragraaf 6.3.2). Wij verwachten dan ook dat jongens naar verhouding minder gunstige scores behalen op de schalen Pro-sociaal gedrag, Pestgedrag, Storend gedrag, Hyperactief/impulsief gedrag en Taakgericht gedrag. Tegelijkertijd verwachten we relatief kleine of afwezige sekseverschillen bij de meer aan internaliserende problematiek gerelateerde schalen: Teruggetrokken gedrag, Zelfstandigheid, Spreekangst en Toetsangst.

Tabel 6.30 laat zien dat de resultaten volledig overeenkomen met de van tevoren geformuleerde verwachtingen. Jongens behalen duidelijk minder gunstige resultaten voor Pro-sociaal gedrag, Pestgedrag, Storend gedrag, Hyperactief/impulsief gedrag en Taakgericht gedrag, terwijl de sekseverschillen bij Teruggetrokken gedrag, Zelfstandigheid, Sociaal zelfbeeld, Cognitief zelfbeeld, Spreekangst en Toetsangst duidelijk kleiner of afwezig zijn.

De vakgebied-specifieke vormen van zelfbeeld en motivatie gedragen zich zoals op basis van eerder onderzoek verwacht mag worden: meisjes zijn als groep duidelijk meer gemotiveerd voor lezen dan jongens, terwijl jongens een positiever beeld van hun eigen competentie in rekenen hebben (Kuhlemeier, Van Boxtel & Van Til, 2002; Marsh, 1986).

Tabel 6.30 Verschillen tussen jongens en meisjes in sociaal-emotioneel functioneren

	Jongen		Meisje		Effect-grootte	Variantie-analyse	
	Gem	Stddev	Gem	Stddev		<i>p</i>	<i>Eta squared</i>
LK basis							
pro-sociaal gedrag	23,14	4,50	25,07	4,24	-0,44	< 0,001	0,046
pestgedrag	7,91	2,65	6,85	1,82	0,48	< 0,001	0,052
storend gedrag	10,47	3,57	9,29	3,44	0,34	< 0,001	0,028
teruggetrokken gedrag	12,95	4,00	13,16	4,13	-0,05	0,406	0,001
taakgericht gedrag	18,94	4,51	21,12	4,15	-0,51	< 0,001	0,060
motivatie voor leren	21,01	4,35	22,34	3,85	-0,33	< 0,001	0,026
zelfstandigheid	19,11	4,15	19,70	4,31	-0,14	0,030	0,005
LL basis							
sociaal zelfbeeld	20,19	2,74	20,34	2,81	-0,05	< 0,001	0,001
cognitief zelfbeeld	22,83	3,06	23,22	3,05	-0,13	< 0,001	0,004
schoolbeleving	23,97	3,59	24,84	3,11	-0,26	< 0,001	0,017
gepest worden	10,95	3,32	9,54	2,76	0,47	< 0,001	0,051
motivatie voor leren	21,81	3,34	22,08	3,37	-0,08	< 0,001	0,002
sociale leeromgeving	21,91	3,24	22,22	3,26	-0,10	< 0,001	0,002
onveilige leeromgeving	12,91	3,78	11,96	3,60	0,26	< 0,001	0,016
onordelijke leeromgeving	14,11	3,74	13,48	3,60	0,17	< 0,001	0,007
LL aanvullend							
pro-sociaal gedrag	20,62	3,29	22,16	2,94	-0,49	< 0,001	0,058
pestgedrag	9,87	2,76	8,44	1,66	0,65	< 0,001	0,091
storend gedrag	11,70	3,62	10,02	2,87	0,52	< 0,001	0,063
taakgericht gedrag	21,16	3,37	22,69	3,06	-0,48	0,006	0,054
motivatie voor lezen	13,20	4,00	15,28	3,98	-0,52	0,030	0,064
motivatie voor rekenen	13,41	4,39	13,03	4,36	0,09	0,035	0,002
zelfbeeld lezen	14,93	3,20	15,50	3,27	-0,18	0,040	0,008
zelfbeeld rekenen	15,27	3,45	13,60	3,68	0,47	0,087	0,052
spreekangst	15,11	4,90	16,08	5,76	-0,18	0,239	0,008
toetsangst	13,38	5,80	15,49	6,46	-0,35	0,284	0,029

6.4.2 Stad versus platteland

Over de relatie tussen sociaal-emotioneel functioneren en mate van verstedelijking is nog weinig bekend. We hebben dan ook geen uitgesproken verwachtingen over eventuele verschillen tussen leerlingen uit meer en minder stedelijke gebieden. Toch vinden we het belangrijk om hier de uitkomsten van een analyse van de invloed van verstedelijking te presenteren. De gebruiker van VISEON 2.0 kan deze informatie immers gebruiken bij de interpretatie van de resultaten. In de analyse is een onderscheid gemaakt tussen sterk tot zeer sterk verstedelijkte gebieden (stad) en niet tot matig verstedelijkte gebieden (platteland). De resultaten zijn weergegeven in tabel 6.31. De verschillen tussen leerlingen uit de stad en het platteland zijn over het algemeen afwezig of hooguit klein en meestal niet significant. De belangrijkste uitkomsten zijn dat leerlingen die op het platteland wonen volgens hun leerkrachten wat meer teruggetrokken gedrag vertonen ($d = -0,38$) en dat plattelandskinderen naar eigen zeggen een wat hoger sociaal zelfbeeld hebben ($d = -0,50$) en hun leeromgeving als ordelijker percipiëren ($d = 0,43$)

Tabel 6.31 *Versillen tussen leerlingen in stedelijke en landelijke omgeving in sociaal-emotioneel functioneren*

	Stad		Platteland		Effect-grootte	Variantie-analyse		
	Gem	Stddev	Gem	Stddev		<i>p</i>	<i>Eta squared</i>	
LK basis								
pro-sociaal gedrag	24,66	4,48	23,82	4,44	0,19	0,057	0,009	
pestgedrag	7,31	2,45	7,37	2,22	-0,03	0,130	0,000	
storend gedrag	9,71	3,59	9,97	3,55	-0,07	0,313	0,001	
teruggetrokken gedrag	12,05	3,94	13,57	3,97	-0,38	< 0,001	0,035	
taakgericht gedrag	20,27	4,58	19,96	4,25	0,07	0,994	0,001	
motivatie voor leren	22,10	4,36	21,52	3,93	0,14	0,848	0,005	
zelfstandigheid	20,06	4,30	18,99	4,12	0,25	< ,000	0,016	
LL basis								
sociaal zelfbeeld	10,73	4,34	13,03	4,77	-0,50	< 0,001	0,059	
cognitief zelfbeeld	20,18	2,81	20,29	2,78	-0,04	0,433	0,000	
schoolbeleving	23,08	3,07	22,96	3,05	0,04	0,419	0,000	
gepest worden	23,99	3,71	24,60	3,16	-0,18	< 0,001	0,007	
motivatie voor leren	10,60	3,34	10,10	3,04	0,16	< 0,001	0,006	
sociale leeromgeving	22,02	3,53	21,91	3,25	0,03	0,492	0,000	
onveilige leeromgeving	21,59	3,30	22,30	3,19	-0,22	< 0,001	0,011	
onordelijke leeromgeving	13,49	4,06	11,90	3,38	0,43	< 0,001	0,042	
LL aanvullend								
pro-sociaal gedrag	14,15	3,81	13,60	3,59	0,15	0,002	0,005	
pestgedrag	21,14	3,35	21,50	3,08	-0,11	0,189	0,003	
storend gedrag	9,48	2,52	8,91	2,21	0,24	0,005	0,014	
taakgericht gedrag	11,36	3,73	10,54	3,04	0,24	0,005	0,014	
motivatie voor lezen	21,46	3,45	22,18	3,15	-0,22	0,011	0,011	
motivatie voor rekenen	14,33	4,05	14,06	4,18	0,07	0,445	0,001	
zelfbeeld lezen	13,43	4,19	12,92	4,42	0,12	0,171	0,003	
zelfbeeld rekenen	15,32	3,21	15,13	3,27	0,06	0,496	0,001	
spreekangst	14,13	3,80	14,59	3,55	-0,13	0,132	0,004	
toetsangst	16,24	5,38	14,85	5,07	0,27	0,002	0,017	

6.4.3 Leerlinggewicht

Anders dan in het cognitieve domein is naar de relatie tussen sociaal-emotioneel functioneren en leerlinggewicht nog weinig onderzoek gedaan. We hebben dan ook geen uitgesproken verwachtingen over eventuele verschillen tussen leerlingen met een verschillend leerlinggewicht. Toch presenteren we hier de analyseresultaten aangezien de gebruiker deze informatie kan gebruiken bij de interpretatie van de resultaten. In het normeringsonderzoek is het leerlinggewicht niet op individueel niveau opgevraagd. Wel kennen we het percentage gewichtenleerlingen per school, dit wil zeggen het percentage leerlingen met een gewicht van 0,30 of 1,20. In de analyse is een onderscheid gemaakt tussen scholen met hooguit tien procent gewichtenleerlingen en scholen met meer dan tien procent gewichtenleerlingen. De resultaten zijn weergegeven in tabel 6.32. De verschillen tussen leerlingen van scholen met minder en meer dan tien procent gewichtenleerlingen zijn over het algemeen afwezig of hooguit klein en vaak niet significant. Kinderen van scholen met meer dan tien procent gewichtenleerlingen vertonen volgens hun leerkrachten wat minder teruggetrokken gedrag ($d = -0,33$) en bovendien zijn ze wat zelfstandiger ($d = -0,33$) dan leerlingen van scholen met hooguit tien procent gewichtenleerlingen. Daarnaast hebben de leerlingen van scholen met meer dan tien procent gewichtenleerlingen naar eigen zeggen een positiever sociaal zelfbeeld ($d = 0,36$), percipiëren zij hun leeromgeving als minder ordelijk ($d = -0,46$) en vertonen zij meer verstrend gedrag ($d = -0,39$).

Tabel 6.32 Verschillen tussen leerlingen van scholen met tot 10% en meer dan 10% gewichtenleerlingen

	Tot 10% gewichtenleerlingen		Meer dan 10% gewichtenleerlingen		Effect- grootte	Variantie- analyse		
	Gem	Stddev	Gem	Stddev		p	Eta <i>squared</i>	
LK basis								
pro-sociaal gedrag	23,98	4,47	24,55	4,47	-0,13	0,057	0,004	
pestgedrag	7,26	2,10	7,50	2,67	-0,10	0,130	0,002	
storend gedrag	9,95	3,55	9,70	3,60	0,07	0,313	0,001	
teruggetrokken gedrag	13,38	4,00	12,06	3,93	0,33	< 0,000	0,025	
taakgericht gedrag	20,09	4,32	20,09	4,52	0,00	0,994	0,000	
motivatie voor leren	21,76	4,01	21,81	4,33	-0,01	0,848	0,000	
zelfstandigheid	18,96	4,15	20,36	4,24	-0,33	< 0,000	0,025	
LL basis								
sociaal zelfbeeld	12,51	4,68	10,82	4,57	0,36	< 0,000	0,029	
cognitief zelfbeeld	20,35	2,73	20,03	2,89	0,11	0,021	0,003	
schoolbeleving	22,92	3,06	23,16	3,04	-0,08	0,112	0,001	
gepest worden	24,55	3,23	24,08	3,61	0,14	0,005	0,004	
motivatie voor leren	10,02	2,91	10,78	3,54	-0,24	< 0,000	0,013	
sociale leeromgeving	21,77	3,30	22,31	3,43	-0,16	0,001	0,006	
onveilige leeromgeving	22,34	3,22	21,48	3,21	0,27	< 0,000	0,016	
onordelijke leeromgeving	11,88	3,34	13,59	4,12	-0,46	< 0,000	0,047	
LL aanvullend								
pro-sociaal gedrag	13,74	3,61	13,89	3,81	-0,04	0,437	0,000	
pestgedrag	21,43	3,07	21,17	3,48	0,08	0,376	0,001	
storend gedrag	8,87	2,19	9,80	2,63	-0,39	< 0,000	0,032	
taakgericht gedrag	10,73	3,27	11,21	3,56	-0,14	0,121	0,004	
motivatie voor lezen	21,95	3,30	21,74	3,29	0,06	0,496	0,001	
motivatie voor rekenen	14,20	4,22	14,10	3,91	0,02	0,799	0,000	
zelfbeeld lezen	12,97	4,39	13,50	4,18	-0,12	0,181	0,003	
zelfbeeld rekenen	15,17	3,31	15,29	3,09	-0,04	0,686	0,000	
spreekangst	14,45	3,66	14,28	3,67	0,05	0,594	0,000	
toetsangst	15,01	5,13	16,39	5,38	-0,26	0,004	0,015	

7 Samenvatting

In deze wetenschappelijke verantwoording staat VISEON 2.0 centraal, een digitaal instrument voor het in kaart brengen en volgen van het sociaal-emotioneel functioneren van leerlingen in het basisonderwijs. Met VISEON 2.0 kunnen leerkrachten informatie verzamelen over in totaal 26 aspecten van het sociaal-emotioneel functioneren van leerlingen. Daarvan worden er acht bevraagd in het basisgedeelte van de leerkrachtvragenlijst (in deze verantwoording benoemd als 'LK basis'), acht in het basisgedeelte van de leerlinglijst ('LL basis') en tien in het aanvullend gedeelte van de leerlinglijst ('LL aanvullend'). Enerzijds hebben de aspecten betrekking op het individueel functioneren van de leerlingen. Daarbij wordt onderscheid gemaakt naar vier domeinen: zelfbeeld, welbevinden, sociaal gedrag en werkhouding. Anderzijds is er in het instrument ook aandacht voor de (sociale, veilige en ordelijke) leeromgeving van de groep als voorwaarde voor het sociaal-emotioneel functioneren.

In hoofdstuk 2 is de meetpretentie van VISEON 2.0 nader beschreven en zijn doelgroep, gebruiksdoel en functie van het instrument toegelicht. Daarbij is ruim aandacht geschonken aan twee perspectieven die leidend waren bij de constructie van het instrument: een ontwikkelingspsychologisch en onderwijskundig perspectief. Daarnaast is uiteengezet hoe VISEON 2.0 aansluit bij VISEON voor kleuters. Het ontwikkelingspsychologisch perspectief is vooral geïntroduceerd met de bedoeling om, mede door het benoemen van centrale ontwikkelingstaken, op het spoor te komen van aspecten van sociaal-emotioneel functioneren die belangrijk zijn voor kinderen in groep 3 tot en met 8 van het basisonderwijs. Waar deze ontwikkelingstaken konden worden gekoppeld aan doelen in het onderwijs (het onderwijskundig perspectief) vormden deze de initiële basis voor de keuze van dimensies en aspecten van sociaal-emotioneel functioneren. In dit hoofdstuk vindt de lezer ten slotte in detail een beschrijving van de dimensies, domeinen en aspecten die in VISEON 2.0 zijn ondergebracht.

In hoofdstuk 3 deden we verslag van het constructie-onderzoek en presenteerden we het uiteindelijk samengestelde instrumentarium. Eerst is aandacht besteed aan de inhoudelijke en psychometrische uitgangspunten voor instrumentconstructie. Vervolgens is verslag gedaan van de beide proefafnames waarin het concept-instrumentarium is onderzocht op onder meer begrijpelijkheid, hanteerbaarheid (o.a. afnametijd), betrouwbaarheid en validiteit. Daarna volgde een beknopte beschrijving van de inhoud, opbouw, afname en scoring van de lijsten voor leerkrachten en leerlingen. Meer informatie is te vinden in de gebruikershandleiding (Cito, 2015). Tot slot presenteerden we het uiteindelijk samengestelde instrumentarium, inclusief een aantal beschrijvende statistische gegevens op schaal- en itemniveau. Er blijkt sprake van homogene schalen met goede item-restcorrelaties, zonder hinderlijke overlap van uitspraken binnen en tussen schalen. De scores op enkele schalen zijn in overeenstemming met de verwachtingen erg scheef verdeeld; het betreft met name schalen zoals Pestgedrag en Storend gedrag die betrekking hebben op probleemgedrag dat in het basisonderwijs weinig voorkomt.

In hoofdstuk 4 is het normeringsonderzoek en de daarop gebaseerde normering beschreven. In paragraaf 4.1 stond de opzet en uitvoering van het normeringsonderzoek centraal. In het normeringsonderzoek zijn de vragenlijsten zowel digitaal als op papier afgenomen bij leerlingen van verschillende jaargroepen. In paragraaf 4.2 bespraken we de vraag in hoeverre het geoorloofd is om de via verschillende afnamemodi en bij verschillende jaargroepen verzamelde gegevens tot één normgroep samen te voegen. Samenvoeging tot één normgroep voor allen is alleen geoorloofd als er geen betekenisvolle verschillen zijn tussen de papieren en digitale afname en tussen de onderscheiden jaargroepen. Zoals op grond van het literatuuronderzoek van Kuhlemeier, Straat en Van der Molen (2015) verwacht werd, bracht de vergelijking geen noemenswaardige verschillen tussen de papieren en digitale afname aan het licht. Dit gold zowel voor het niveau van het sociaal-emotioneel functioneren, de betrouwbaarheid van de meting, de validiteit in termen van *Differential Item Functioning* als de invloed van de afnamemodus op de landelijke normen. De jaargroep bleek evenmin van invloed op het sociaal-emotioneel functioneren. De effectgroottes voor de verschillen tussen opeenvolgende jaargroepen bleken namelijk verwaarloosbaar klein. Daarnaast

vertoonden de effectgroottes voor de verschillen tussen opeenvolgende jaargroepen een grillig patroon, zonder dat er sprake was van een min of meer gelijkmatige toe- of afname van het sociaal-emotioneel functioneren van de lagere naar de hogere jaargroepen. De algehele conclusie was dat er geen aanleiding is voor afzonderlijke normering per afnamemodus of per jaargroep. De onder verschillende afnamecondities bij verschillende jaargroepen verzamelde gegevens mogen tot één normgroep worden samengevoegd. In paragraaf 4.3 besteedden we aandacht aan de representativiteit van de normeringssteekproef ten opzichte van de landelijke populaties van scholen en leerlingen. De normering van de schalen van LK basis, LL basis en LL aanvullend is gebaseerd op respectievelijk 1007, 1970 en 624 leerlingen. De responsgroep van *scholen* bleek voldoende representatief ten opzichte van de populatie van scholen voor basisonderwijs in Nederland. De responsgroep van *leerlingen* was voldoende representatief met betrekking tot verstedelijking en leerlinggewicht, maar bij regio, schoolgrootte en denominatie zagen we lichte tot soms forse afwijkingen van het landelijke beeld (met name bij LL aanvullend). Een geëigend middel om een niet volledig representatieve responsgroep meer representatief te maken is weging naar een proportionele afspiegeling van de populatie. Weging is echter alleen zinvol als de achtergrondkenmerken aan de hand waarvan de representativiteit bepaald is ertoe doen (dit wil zeggen: variantie in het sociaal-emotioneel functioneren verklaren). Deze kenmerken bleken echter niet of nauwelijks van belang voor het sociaal-emotioneel functioneren van de leerlingen. Weging naar een proportionele afspiegeling zou dus geen of hooguit een zeer klein effect hebben op de schattingen van het sociaal-emotioneel functioneren (en de normering). Om deze reden is besloten weging achterwege te laten.

In paragraaf 4.4 stond de eigenlijke normeringprocedure centraal: de omzetting van de ruwe scores in percentielscores en de drie niveauscores N (Niet afwijkend; 0-80%), A (Afwijkend; 80-95%) en Zeer afwijkend (Z; 95-100%). Tevens rapporteerden we de afkappunten op de ruwe somscoreschaal tussen de signalen N en A en tussen A en Z. De omzettingstabellen zijn vanwege hun omvang niet opgenomen in deze verantwoording maar op aanvraag verkrijgbaar bij de eerste auteur van deze verantwoording.

Het betrouwbaarheidsonderzoek voor VISEON 2.0 is gerapporteerd in hoofdstuk 5. Het had betrekking op drie aspecten van betrouwbaarheid: interne consistentie, interbeoordelaarsbetrouwbaarheid en test-hertestbetrouwbaarheid.

Op de eerste plaats hebben wij de interne consistentie van de afzonderlijke schalen in het instrument vastgesteld in termen van Cronbachs alfa. Deze coëfficiënt bereikt voor alle 26 schalen waarden die hoger zijn dan 0,80, wat als hoog aangemerkt kan worden, zeker in relatie tot het gebruiksdoel van het instrument (op basis van VISEON 2.0 worden geen belangrijke, irreversibele beslissingen genomen). De bijbehorende betrouwbaarheidsintervallen zijn smal en de standaardmeetfouten zijn klein.

In aanvulling op de betrouwbaarheid in termen van interne consistentie werd in een beperkt aanvullend onderzoek de interbeoordelaarsbetrouwbaarheid geschat door beide leerkrachten in een duobaan de leerkrachtlijst te laten invullen voor in totaal 290 leerlingen. De percentages exacte overeenstemming per antwoordmogelijkheid (op itemniveau) liggen voor de schalen van de leerkrachtlijst tussen 67% en 88%. Wanneer men verschillen in één scorepunt op de vierpuntsschalen acceptabel acht, liggen de percentages overeenstemming zelfs tussen 97% en 99%; de Kappa's variëren tussen 0,43 en 0,67 wat volgens de vuistregel van Landis en Koch (1997) als *moderate* tot *substantial* geïnterpreteerd mag worden.

De correlaties tussen de schaalscores zoals vastgesteld bij de eerste en tweede afname liggen, zoals gebruikelijk, lager dan de interne consistentiecoëfficiënten. Voor de meeste schalen liggen de waarden tussen 0,66 en 0,81, waarbij de schaal Motivatie voor leren een uitzondering vormt met een waarde van 0,53. Om de hoogte van de waarden te kunnen interpreteren zochten we vergelijkingsmateriaal in de vorm van soortgelijke instrumenten. De gevonden waarden zijn lager dan de waarden die we aantreffen bij de kleuterversie en de eerste versie van VISEON. In laatstgenoemd instrument wordt gebruikgemaakt van bipolaire schalen die in de regel minder scheef verdeeld waren dan de scores op de leerkrachtsschalen van VISEON 2.0. Mogelijk zijn de correlaties beïnvloed door verdelingskenmerken. De VISEON 2.0-correlaties waren ongeveer even hoog als bij Zien! en hoger dan bij de SDQ (Hart & Ziel).

Het onderzoek naar de test-hertestbetrouwbaarheid werd uitgevoerd op de twee in de gebruikershandleiding aanbevolen afnamemomenten november en april (Cito, 2015). De in november verzamelde scores vertonen een middelmatig tot sterke samenhang met de scores voor dezelfde leerlingen in april ($.36 < r < .72$). De scores zijn dus relatief stabiel over de onderzochte periode van vijf maanden. Daarbij bleek

het oordeel van de leerkracht wat stabiel over tijd dan dat van de leerling ($.59 < r < .72$).

De geconstateerde stabiliteit was echter niet zo hoog dat we moeten uitgaan van sociaal-emotioneel functioneren als een stabiel en onveranderbaar persoonskenmerk. De geconstateerde middelmatige tot sterke samenhang geeft aanleiding tot de veronderstelling dat er zich gedurende vijf maanden veranderingen in het sociaal-emotionele functioneren van individuele leerlingen kunnen voordoen die mogelijk extra aandacht vereisen. Tegelijkertijd is de samenhang niet zo laag dat er twijfel opkomt of er beide momenten wel dezelfde eigenschappen gemeten worden.

Een ongeveer even hoge stabiliteit is ook gevonden in de vergelijking met andere instrumenten voor sociaal-emotioneel functioneren over een ongeveer even lange periode. Opgemerkt wordt dat de gerapporteerde stabiliteitscoëfficiënten de meetnauwkeurigheid in sterke mate onderschatten. Voor een juiste schatting van de meetnauwkeurigheid is onderzoek nodig met een aanzienlijke korter interval van bijvoorbeeld drie à vier weken.

In hoofdstuk 6 is verslag gedaan van diverse analyses met betrekking tot de begripsvaliditeit; criteriumvaliditeit is gezien de gebruiksdoelen niet van toepassing.

De factoranalyses van de interne structuur lieten zien dat vrijwel alle 26 schalen niet alleen uni-dimensioneel maar ook factor-zuiver zijn. De sterkte en richting van de correlaties tussen de schalen zoals deze voortkwamen uit de gevonden factoroplossingen zijn in overeenstemming met theoretische noties. Nagegaan is in hoeverre de gevonden factorstructuur invariant is over drie groepen leerlingen, te weten jongens en meisjes, deelnemers aan de papieren en digitale afname (afnamemodus) en jongere en oudere leerlingen (jaargroep). Daartoe zijn de correlaties tussen de schalen in de ene groep vergeleken met die in de andere groep. De verschillen tussen de correlaties in de vergeleken groepen zijn over het algemeen zeer klein. Uit de uitgevoerde DIF-analyses komt naar voren dat de uitspraken voor deze drie groepen dezelfde betekenis hebben.

De afwezigheid van gemiddelde verschillen tussen jaargroepen betekent dat VISEON 2.0 niet geschikt is voor het vaststellen van de progressie van het sociaal-emotioneel functioneren van groep 3 naar groep 8 in de populatie. Hierin verschilt VISEON 2.0 overigens niet van soortgelijke instrumenten zoals Zien!, de SDQ en de Kanjervragenlijst. Wel vormen de gevonden stabiliteitscoëfficiënten over een periode van vijf maanden een ondersteuning voor de bruikbaarheid van VISEON 2.0 voor het traceren van 'werkelijke' veranderingen in het sociaal-emotioneel functioneren van individuele leerlingen of klassen (bijvoorbeeld als gevolg van ingrijpende gebeurtenissen in de leerling zelf, de thuissituatie of de leeromgeving).

De convergente en discriminante validiteit werd onderzocht aan de hand van correlaties met (onderdelen van) andere instrumenten voor sociaal-emotioneel functioneren. Het betreft VISEON (eerste generatie), de *Strength and Difficulties Questionnaire* (SDQ), de Kanjervragenlijst en de Schoolvragenlijst (SVL). Overeenkomstig van tevoren opgestelde verwachtingen bleken de correlaties van VISEON 2.0 met inhoudelijk verwante schalen van andere instrumenten telkens beduidend hoger dan die met inhoudelijk niet of minder verwante schalen. De uitkomsten van de analyses geven hiermee geen enkele aanleiding om de convergente en discriminante validiteit van VISEON 2.0 in twijfel te trekken.

Tot slot analyseerden we verschillen in gemiddelden voor relevante groepen leerlingen. De gevonden verschillen tussen jongens en meisjes waren volledig in lijn met theoretische notities en resultaten van eerder onderzoek (o.a. Kuhlemeier, Van Boxtel & Van Til, 2012). Ten aanzien van de invloed van verstedelijking en leerlinggewicht hadden wij geen uitsproken verwachtingen. De verschillen tussen leerlingen uit de stad en van het platteland (verstedelijking), tussen jongere en oudere leerlingen (jaargroep) en tussen leerlingen van scholen met meer of minder dan tien procent gewichtsléerlingen bleken over het algemeen afwezig of hooguit klein.

Op basis van de gepresenteerde resultaten blijkt VISEON 2.0 een bruikbaar, betrouwbaar en valide instrument voor het beoordelen en volgen van het sociaal-emotioneel functioneren van leerlingen in het basisonderwijs.

Literatuur

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*, 213-232.
- Barry, T. D., Lyman, R., & Klinger, L. G. (2002). Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology*, *40*, 259–283.
- Boekaerts, M., & Simons, P.R.J. (1995). *Leren en instructie, psychologie van de leerling en het leerproces*. Assen: Van Gorcum.
- Boxtel, H. van, Hop, M., & Hemker, B. (2014). *WISEON voor kleuters. Volginstrument voor sociaal-emotioneel functioneren. Wetenschappelijke verantwoording*. Arnhem: Cito.
- Bos, I. (2015). Q&A Gijs Huitsing. Kind kan pester en slachtoffer tegelijk zijn. *Didactief, januari*, 34.
- Broer, N.A., Haverhals, B., & Bruin, H.L. de (2012). *Verantwoording pedagogisch expertsysteem Zien! voor het primair onderwijs*. Gouda: Driestar Onderwijsadvies.
- Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Cito (2015). *WISEON 2.0. Handleiding*. Arnhem: Cito.
- Cito (z.j.). *Computerprogramma LOVS*. Arnhem: Cito.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37- 46.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (revised edition). Hillsdale, NJ: Lawrence Erlbaum.
- Coleman, J.S. (1968). The mathematical study of change. In: M.M. Blalock, Jr., & A.B. Blalock (Eds.), *Methodology in social research* (pp. 428-478). New York: McGraw-Hill.
- Cronbach, L.J. (1970). *Essentials of psychological testing*. Third Edition. New York: Harper & Row.
- Coplan, R.J., & Prakash, K. (2003). Spending time with teacher: characteristics of preschoolers who frequently elicit versus initiate interactions with teachers. *Early Childhood Research Quarterly*, *18*, 143–58.
- Dam, G. ten, & Volman, M. (1999). *Scholen voor sociale competentie. Een pedagogisch-didactische benadering*. Lisse: Swets & Zeitlinger Publishers.
- Dam, G. ten, Geijssel, F., Ledoux, G., Reumerman, R., Keunen, M., & Visser, A. (2010). *Handleiding burgerschap meten voor het primair onderwijs*. Amsterdam: Rovict BV.
- Dam, G. ten, Geijssel, F., Reumerman, R., & Ledoux, G. (2010). Burgerschapscompetenties: de ontwikkeling van een meetinstrument. *Pedagogische Studiën*, *87*, 5, 313-333.

- Dam, G. ten, Volman, M., Westerbeek, K., Wolfram, P., & Ledoux, G., m.m.v. Peschar, J. (2003). *Sociale competentie langs de meetlat*. Den Haag: Transferpunt Onderwijsachterstanden.
- Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.
- Deci, E.L., & Ryan, R.M. (1991). A motivational approach to self: Integration in personality. In: R. Dienstbier (Ed.), *Nebraska symposium on motivation: Vol. 38. Perspectives on motivation* (pp. 237–288). Lincoln: University of Nebraska Press.
- Deci, E.L., & Ryan, R.M. (2000). The “What” and “Why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11, 4, 227-268.
- Diepenmaat, A., Eijdsen, M. van, Janssens, J., Loomans, E., & Stone, L. (2014). *Verantwoording SDQ leerkrachtvragenlijst*. Amsterdam/Nijmegen: GGD/Radboud Universiteit.
- Dijkstra, A.B. (2012). *Sociale opbrengsten van onderwijs*. Amsterdam: Research Institute Child Development and Education (CDE).
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie; herdruk mei 2010)*. Utrecht: NIP.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.
- Goudena, P. (1999). Een veilige omgeving waarin kinderen zich kunnen ontwikkelen. In: K. Bakker, M. Pannebakker & J. Sijnders (Red.), *Kwetsbaar en competent. Sociale participatie van kwetsbare jeugd. Theorie, beleid en praktijk* (pp. 55 – 66). Utrecht: NIZW Uitgeverij.
- Greven, J., & Letschert, J. (2006). *Kerndoelen primair onderwijs*. Den Haag: Ministerie van OCW.
- Haney, P., & Durlak, J.A. (1998). Changing self-esteem in children and adolescents: A meta-analytic review. *Journal of Clinical Child Psychology*, 27, 423-433.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford.
- Hattie, J.A. (1992). *Self-concept*. Hillsdale NJ.: Lawrence Erlbaum Associates.
- Havighurst, R. (1948, 1972). *Developmental tasks and education*. New York: McKay.
- Heise, R. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review*, 34, 1, 93-100.
- Heuvelmans, T. (2013). *TiaPlus Users Manual*. Arnhem: Cito.
- Huitsing, G. (2014). *A social network perspective on bullying* (Academisch proefschrift). Groningen: Rijksuniversiteit Groningen.
- Inspectie van het Onderwijs (2010). *Analyse en waardering van opbrengsten primair onderwijs*. Utrecht: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (2011). *Toezichtkader PO/VO 2009 (Versie 13 juli 2009; Aanpassing 17 mei 2011)*. Utrecht: Inspectie van het Onderwijs.

Johnson, M.H. (2011). *Developmental cognitive neuroscience: an introduction*. Chichester (UK): Wiley-Blackwell.

Kamphuis, F., Rosier, W., Sluijter, C., Lubbe, M. van der, Putten, K. van, & Meijer, J. (2004). *VISEON: VolgInstrument voor Sociaal-Emotionele Ontwikkeling. Verantwoording*. Arnhem: Cito.

Kuhlemeier, H. (2015). *De samenstelling van de schalen voor het normeringsonderzoek VISEON 2.0*. Interne publicatie. Arnhem: Cito.

Kuhlemeier, H., Bergh, H. van den, & Rijlaarsdam, G. (2002). The dimensionality of speaking and writing apprehension: a multilevel factor analysis of situational, task and school effects. *British Journal of Educational Psychology*, 72, 464-482.

Kuhlemeier, H., Boxtel, H. van, & Til, A. van (2012). *Balans van de sociale opbrengsten in het basisonderwijs. PPOON reeks nummer 48*. Cito: Arnhem.

Kuhlemeier, H., Straat, H., & Molen, P. van der (2015). *Prestaties op papieren en digitale examens: Wat is het verschil? Verslag van een literatuurstudie*. Arnhem/Utrecht: Cito/CvTE.

Ladd, G.W., Burgess, K.B. (199). Charting the relationship trajectories of aggressive, withdrawn, and aggressive/withdrawn children during early grade school. *Child Development*, 70, 910-29.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1, 159-174.

Ledoux, G., Geijssel, F., Dam, G. ten, & Reumerman, R. (2011). Burgerschapscompetenties van jongeren in Nederland. *Pedagogische Studien*, 88, 1, 3-22.

Marsh, H.W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 1, 129-149.

Marsh, H.W. (1990). A multidimensional, hierarchical self-concept: theoretical and empirical justification. *Educational Psychology Review*, 2, 77-172.

Marsh, H.W., & Shavelson, R.J. (1985). Self-concept: Its multifaceted hierarchical structure. *Educational Psychologist*, 20, 107-125

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42, 2, 109-142.

Meij, H. (2011). *De basis van opvoeding en ontwikkeling*. Utrecht: Nederlands Jeugdinstituut, geraadpleegd op <http://www.nji.nl/nl/Normale-ontwikkeling-pdftekst.pdf>.

Ministerie van OCW (2006). *Kerndoelen primair onderwijs*. Den Haag: Ministerie van OCW.

Mönks, F.J. & Knoers, A.M.P. (2009). *Ontwikkelingspsychologie. Inleiding tot de verschillende deelgebieden*. Assen: Van Gorcum.

Muris, P., Meesters, C., & Berg, F. van den (2003). The strengths and difficulties questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1-8.

Nijntjes, W., & Diepen, M. van (2008). *Maatschappelijke taken in het onderwijs*. Amsterdam: TSO Nipo.

Oosterlaan, J., & Veerman, J.W. (2008). Achtergronden en gebruik van vragenlijsten voor het vaststellen van emotionele en gedragsproblemen bij kinderen. In: T. Kievit, J.A. Tak & J.D. Bosch (Red.), *Handboek psychodiagnostiek voor de hulpverlening aan kinderen* (pp. 255-300). Utrecht: De Tijdstroom.

Riksen-Walraven, J.M.A. (1989). Meten in perspectief. Een levensloopmodel als achtergrond bij het meten en beïnvloeden van gedrag en interacties. *Tijdschrift voor Orthopedagogiek*, 23, 16-33.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem. Different concepts, different outcomes. *American Sociological Review*, 60, 1, 141-156.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem. Different concepts, different outcomes. *American Sociological Review*, 60, 1, 141-156.

Roy, B. van, Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *Journal of Child Psychology and Psychiatry*, 49, 1304-12.

Rubin, K.H., & Asendorpf, J. (1993) (Eds.), *Social withdrawal, inhibition and shyness in childhood*. Hillsdale, NJ: Erlbaum.

Rubin, K.H., Hymel, S., & Mills, R.S.L. (1989). Sociability and social withdrawal in childhood: stability and outcomes. *Journal of Personality*, 57, 237-55.

Rubin, K.H., Wojslawowicz, J.C., Rose-Krasnor, L., Booth-LaForce, C.L., & Burgess, K.B. (2006). The best friendships of shy/withdrawn children: prevalence, stability, and relationship quality. *Journal of Abnormal Child Psychology*, 34, 139-53.

Rubin, K.H., Coplan, R.J., & Bowker, J.C. (2009). Social withdrawal in childhood. *Annual Review of Psychology*, 60, 141-171.

Rudasill, K.M., Rimm-Kaufman, S.E., Justice, L.M., & Pernce, K. (2006). Temperament and language skills as predictors of teacher-child relationship quality in preschool. *Early Educational and Development*, 17, 271-91

Sameroff, A. (2010). A unified theory of development: A dialectic integration of nature and nurture, *Child Development*, 81, 1, 6-22.

Shavelson, R.J., & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of educational psychology*, 74, 3-17.

Shavelson, R.J., Hubner, J.J. & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46, 407-44.

Smits, J.A.E., & Vorst, H.C.M. (1990). *Schoolvragenlijst (SVL)*. Nijmegen/Lisse: Berkhout/Swets & Zeitlinger.

- Spinath, B., & Spinath, F.M. (2005). Longitudinal analysis of the link between learning motivation and competence beliefs among elementary school children. *Learning and Instruction, 15*(2), 87-102.
- Steinmayr, R., & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*(1), 80-90.
- Stone, L.L., Otten, R., Engels, R.C. M. E., Vermulst, A.A., & Janssens, J.M.A.M. (2010). Psychometric properties of the parent and teacher versions of the Strength and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13*, 254-274.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2919684/>
- Vliek, L. (2015). *Effects of Kanjertraining (Topper Training) on Emotional Problems, Behavioural Problems and Classroom Climate* (Proefschrift). Utrecht: Digitale Klerken.
- Vliek, L., Riet, B., & Weide, G. (2012). *Verantwoording van de Kanjervragenlijst*. Almere: Instituut voor Kanjertrainingen BV.
- Vogels, A.G.C., Crone, M.R., Hoekstra, F., & Reijneveld, S.A. (2005). *Drie vragenlijsten voor het opsporen van psychosociale problemen bij kinderen van zeven tot twaalf jaar*. Leiden: TNO.
- Vijverberg, G. (2009). *De invloed van persoonskenmerken op leerprestaties*. Intern rapport. Arnhem: Cito.
- Zeidner, M. (1998). *Test anxiety: The state of the art*. New York: Plenum.

Bijlage 1

Correlaties tussen de schalen van LK basis, LL basis en LL aanvullend

Ter bestudering van de interne structuur van de meting van het sociaal-emotioneel functioneren rapporteerden we in paragraaf 6.1.3 de samenhang tussen de schalen van VISEON 2.0. We deden dat alleen binnen elk van de drie onderdelen van VISEON 2.0 afzonderlijk (LK basis, LL basis en LL aanvullend). De correlaties tussen schalen van verschillende onderdelen zijn vanwege de omvang niet in de hoofdtekst opgenomen. In deze bijlage doen we verslag van de samenhang tussen de schalen van verschillende onderdelen, met inbegrip van de relaties tussen de schalen voor leerkrachten en leerlingen.

Correlaties tussen LL basis en LL aanvullend

De correlaties tussen de acht schalen van LL basis en de tien schalen van LL aanvullend zijn weergegeven in tabel 1.

Tabel 1 Correlaties tussen de schalen van LL basis en LL aanvullend

	LL basis							
	sociaal zelfbeeld	cognitief zelfbeeld	school-beleving	gepest worden	motivatie voor leren	sociale leer-omgeving	onveilige leer-omgeving	onordelijke leer-omgeving
LL aanvullend								
pro-sociaal gedrag	0,39	0,32	0,36	-0,21	0,46	0,47	-0,18	-0,24
pestgedrag	-0,19	-0,20	-0,23	0,58	-0,28	-0,23	0,46	0,37
storend gedrag	-0,17	-0,38	-0,22	0,36	-0,40	-0,19	0,29	0,37
taakgericht gedrag	0,26	0,48	0,28	-0,17	0,60	0,27	-0,21	-0,20
motivatie voor lezen	0,11	0,25	0,29	-0,15	0,31	0,22	-0,21	-0,17
motivatie voor rekenen	0,09	0,29	0,20	-0,07	0,39	0,11	-0,07	-0,17
zelfbeeld lezen	0,14	0,27	0,17	-0,07	0,25	0,13	-0,06	-0,05
zelfbeeld rekenen	0,13	0,40	0,06	0,02	0,23	0,04	-0,01	-0,06
spreekangst	-0,27	-0,28	-0,19	0,17	-0,20	-0,26	0,17	0,23
toetsangst	-0,15	-0,29	-0,14	0,21	0,02	-0,19	0,21	0,25

Geheel overeenkomstig de verwachtingen hangt Motivatie voor leren uit LL basis sterk samen met Taakgericht gedrag uit LL aanvullend ($r = 0,60$).

Tussen enerzijds Pestgedrag uit LL aanvullend en Gepest worden en Onveilige leeromgeving uit LL basis zien we een middelmatig sterke samenhang (r respectievelijk 0,58 en 0,46). De relatief hoge correlatie tussen Pestgedrag en Gepest worden is in overeenstemming met onderzoek naar pestgedrag waaruit naar voren komt dat leerlingen die zelf pesten vaak ook slachtoffer zijn van pestgedrag door anderen (o.a. Bos, 2015).

Pro-sociaal gedrag vertoont een middelmatig positieve samenhang met Motivatie voor leren ($r = 0,46$) en Sociale leeromgeving ($r = 0,47$) en een zwakke samenhang met Sociaal zelfbeeld ($r = 0,39$), Cognitief zelfbeeld ($r = 0,32$) en Schoolbeleving ($r = 0,36$).

Conform de eis van discriminante validiteit hangen de beide leergebiedspecifieke schalen voor zelfbeeld en de schalen voor spreek- en toetsangst niet of nauwelijks samen met de schalen van LL basis. Een begrijpelijke uitzondering is de relatief hoge correlatie tussen Zelfbeeld rekenen en Cognitief zelfbeeld ($r = 0,40$).

Correlaties tussen de schalen van enerzijds LK basis en anderzijds LL basis en LL aanvullend

De correlaties tussen de acht schalen van LK basis en de achttien schalen van LL basis en LL aanvullend zijn weergegeven in tabel 2.

Tabel 2 Correlaties tussen de schalen van enerzijds LK basis en anderzijds LL basis en LL aanvullend

	LK basis							
	pro- sociaal gedrag	pest- gedrag	storend gedrag	terug- getrokken gedrag	taak- gericht gedrag	motivatie voor leren	zelfstandig- heid	hyper- actief/im- pulsief
LL basis								
sociaal zelfbeeld	0,21	-0,15	-0,08	-0,17	0,09	0,10	0,15	-0,10
cognitief zelfbeeld	0,16	-0,15	-0,19	-0,09	0,27	0,35	0,15	-0,25
schoolbeleving	0,22	-0,23	-0,16	-0,10	0,17	0,19	0,08	-0,19
gepest worden	-0,20	0,32	0,21	0,09	-0,12	-0,09	0,01	0,21
motivatie voor leren	0,12	-0,13	-0,16	-0,05	0,20	0,25	0,09	-0,22
sociale leeromgeving	0,13	-0,10	-0,07	-0,06	0,05	0,12	0,05	-0,13
onveilige leeromgeving	-0,11	0,18	0,17	0,05	-0,12	-0,11	0,05	0,16
onordelijke leeromgeving	-0,12	0,14	0,16	0,12	-0,11	-0,12	-0,04	0,20
LL aanvullend								
pro-sociaal gedrag	0,26	-0,19	-0,17	-0,06	0,08	0,20	0,10	-0,19
pestgedrag	-0,25	0,40	0,34	-0,09	-0,22	-0,21	0,09	0,33
storend gedrag	-0,22	0,33	0,50	-0,08	-0,35	-0,30	0,06	0,45
taakgericht gedrag	0,19	-0,23	-0,35	0,09	0,30	0,29	-0,01	-0,29
motivatie voor lezen	0,15	-0,21	-0,28	0,05	0,21	0,28	0,02	-0,27
motivatie voor rekenen	0,16	-0,17	-0,18	-0,07	0,14	0,29	0,14	-0,18
zelfbeeld lezen	0,05	-0,06	-0,09	-0,10	0,20	0,23	0,14	-0,12
zelfbeeld rekenen	0,06	-0,02	-0,03	-0,21	0,14	0,25	0,22	-0,07
spreekangst	-0,04	-0,08	-0,13	0,14	-0,10	-0,09	-0,11	-0,05
toetsangst	-0,05	-0,08	-0,05	0,14	-0,10	-0,09	-0,10	-0,02

Het oordeel van de leerkracht over het sociaal-emotioneel functioneren blijkt over het algemeen hooguit zwak samen te hangen met dat van de leerling zelf. Zelfs bij schalen van de leerkrachtlijst die een pendant in de leerlinglijst hebben, zien we hooguit middelmatige correlaties tussen de antwoorden van leerkrachten en leerlingen. Bij deze gelijklopende schalen bedragen de correlaties tussen de antwoorden van leerkracht en leerlingen 0,26 voor Pro-sociaal gedrag, 0,40 voor Pestgedrag, 0,50 voor Storend gedrag, 0,30 voor Taakgericht gedrag en 0,25 voor Motivatie voor leren. Onderzoek heeft laten zien dat antwoorden van leerkrachten en leerlingen over het algemeen zwakker samenhangen dan antwoorden van leerkrachten of leerlingen onderling (o.a. Achenbach, McConaughy & Howell, 1987; Vlieg, Riet & Weide, 2012). Desalniettemin is de samenhang tussen de leerkracht- en leerlingsschalen met enige regelmaat lager dan men op basis van inhoudelijk-theoretische overwegingen zou verwachten. Wel kunnen we (gelukkig) constateren dat de richting (het algebraïsche teken) van de correlaties in overeenstemming is met de verwachtingen en in geen enkel geval als contra-intuïtief beschouwd moet worden.

Discussie

De bevinding dat de samenhang tussen de oordelen van leerkracht en leerling bij vergelijkbare begrippen duidelijk hoger is dan bij niet of minder verwante begrippen vormt een ondersteuning voor de convergente en discriminante validiteit. Dit neemt niet weg dat de samenhang tussen de oordelen vaak zwakker is dan men zou verwachten op basis van de gedachte dat alle schalen betrekking hebben op sociaal-emotioneel functioneren. Echter, zoals Oosterlaan en Veerman (2008) overtuigend uiteenzetten, hoeven discrepanties tussen de waarnemingen van leerkrachten en leerlingen niet noodzakelijkerwijs als een validiteitsprobleem gezien te worden. Discrepanties tussen de oordelen van beide typen respondenten kunnen ook een bron van aanvullende informatie vormen. Oordelen kunnen uiteenlopen omdat de situaties waarin leerkracht en

leerling gedrag waarnemen verschillen. Een deel van het bevroegde gedrag speelt zich immers buiten het zicht van de leerkracht af. Wie de beste informant is, hangt ook af van de aard van het onderzoeksobject. Zo is internaliserende problematiek voor de leerkracht vaak niet direct waarneembaar zodat de leerling daarover het beste zelf kan rapporteren. Voor hyperactief-impulsief gedrag is daarentegen de leerkracht (of de ouders) de beste informant. Leerkracht en leerling kunnen problemen bewust of onbewust over- of onderrapporteren, bijvoorbeeld omdat zij verschillende referentiekaders hanteren of omdat de school baat heeft bij een bepaalde uitslag. Jonge kinderen zijn soms minder goed dan de leerkracht in staat om eigen gedrag te vergelijken met dat van medeleerlingen. Onderzoek laat zien dat leerkrachten en leerlingen inderdaad niet altijd hetzelfde gedrag observeren en daardoor tot andere oordelen komen. Zo toonde Huitsing (2014) aan dat leerkrachten en leerlingen verschillen in hun waarneming van pestgedrag. Leerkrachten rapporteren minder pestgedrag tussen jongens en meisjes dan leerlingen zelf, en juist méér pestgedrag tussen meisjes onderling. Leerlingen zelf zeggen dat jongens zowel jongens als meisjes pesten en dat meisjes vooral onderling pesten. Leerkrachten en leerlingen zijn het er vaak wel over eens dat een kind wordt gepest of dat het andere kinderen pest, maar ze noemen niet noodzakelijk dezelfde slachtoffer-pestrelaties. Verder zijn leerkrachten en leerlingen het meer met elkaar eens over wie de pester is dan over wie het slachtoffer is. Tot slot kunnen leerlingen een incident 'opkloppen', terwijl leerkrachten te gemakkelijk denken dat de ruzie beslecht is. Ongetwijfeld vormen de verschillende waarnemingen en referentiekaders een mogelijke verklaring voor verschillen tussen de scores van leerkracht en leerling. Met betrekking tot de vraag hoe hiermee om te gaan volgen wij het advies van Oosterlaan en Veerman (2008): "Zolang de validiteit van verschillende informatie met betrekking tot verschillende gedragsproblemen niet eenduidig is vastgesteld, is het raadzaam de informatie van verschillende informanten gelijk te wegen en geen informatie verloren te laten gaan (...) (p. 272)". Ogenschoonlijk tegenstrijdige informatie kan echter bij leerkrachten vragen oproepen die de validiteit van VISEON 2.0 ten onrechte in diskrediet kunnen brengen. In de handleiding hebben we de gebruiker er daarom op geattendeerd dat zijn of haar oordeel kan afwijken van dat van de leerling en aan de hand van praktijkvoorbeelden uitgelegd hoe hij of zij daarmee om kan gaan.

Cito helpt je inzicht te krijgen in je ontwikkeling en mogelijkheden. Door kennis, vaardigheden en competenties objectief meetbaar te maken en de ontwikkeling er van te volgen, kun je het beste uit jezelf halen, verantwoorde keuzes maken en beter richting geven aan je toekomst. Cito draagt daaraan bij door wereldwijd werk te maken van goed en eerlijk toetsen, vanuit de kernwaarden kundig, toonaangevend, integer, innovatief en betrokken.

Cito

Amsterdamseweg 13
Postbus 1034
6801 MG Arnhem
T (026) 352 11 11
www.cito.nl

Fotografie: Ron Steemers