

Oog voor Leergedrag

Observatie van gedrag in de klas



C.S. Verhoef-Platerink

0789933

Eerste lezer: dr. M. J. A. J. Verhallen

Tweede lezer: dr. C. A. Espin

Masterproject, Universiteit Leiden, Vakgroep Leerproblemen

18 december 2012

Oog voor Leergedrag: Observatie van gedrag in de klas

Samenvatting

Uit onderzoek is gebleken dat leergedrag en leerprestaties met elkaar samenhangen. In het licht van de huidige ambities in het Nederlands voortgezet onderwijs om de kwaliteit van het onderwijs te verbeteren en om opbrengstgericht te werken, kan inzicht in taakgerichtheid belangrijke handvatten bieden voor leerlinginterventies en klasmanagement. Hoewel observaties vergeleken met andere meetinstrumenten relatief veel tijd in beslag nemen, lijkt systematische directe observatie het meest geschikte instrument voor het meten van leergedrag. Over de vraag welke aspecten van het leergedrag voldoende convergente validiteit opleveren, is nog weinig bekend. Om dit nader te onderzoeken en om een efficiëntieslag te kunnen maken ten opzichte van bestaande observatie-systemen is het instrument Leerling Observatie in de Klas ontworpen, waarbij gekozen is voor een relatief eenvoudige opzet. Behulp van dit instrument is door middel van *momentary time sampling* de frequentie gemeten van vier typen leergedrag: actief, aan taak, niet aan taak en storend leergedrag. De metingen zijn verricht onder 22 leerlingen uit het praktijkonderwijs en 32 nieuwkomers, leerlingen die maximaal twee jaar in Nederland woonden en de Nederlandse taal nog onvoldoende beheersen om deel te kunnen nemen aan het regulier onderwijs. Aan de hand van mentorbeoordelingen is de convergente validiteit van de verschillende typen leergedrag onderzocht. Daarnaast richtte het onderzoek zich op de vraag of de typen leergedrag voldoende verband hielden met leerprestaties. Gemiddeld sterke en significante correlaties met de mentorscore zijn gevonden voor het type actief leergedrag en de samengevoegde typen positief en negatief leergedrag. De onderzochte typen leergedrag vertoonden geen significante relatie met leerprestaties. De verdelingen van het leergedrag verschilden per type les en per school zodat een onderscheid op basis van deze factoren op zijn plaats lijkt. Aanbevolen wordt om de convergente validiteit van Leerling Observatie in de Klas nader te onderzoeken met behulp van een gevalideerd meetinstrument onder leerlingen uit het regulier onderwijs.

Introductie

Een van de speerpunten van het Nederlands onderwijsbeleid op dit moment is *opbrengstgericht werken* (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011). Deze onderwijskundige aanpak is erop gericht om op een gestructureerde manier het beste uit alle leerlingen te halen. Dit geschiedt door hoge leerdoelen voor leerlingen te stellen. Om dit te verwezenlijken dient de leerontwikkeling systematisch te worden gevolgd en dient het volgsysteem nadrukkelijk te worden verbonden met het onderwijs zodat de lessen zowel inhoudelijk en als qua aanbod beter worden afgestemd op de verschillende niveaus van de leerlingen. Hieraan verbonden interventies worden ingezet op leerlingniveau, groepsniveau en schoolniveau. Dit vergt prestatiegerichtheid en veranderingsgezindheid van zowel leerlingen als leerkrachten (Visscher & Ehren, 2011).

De behoefte aan verbetering van het onderwijs in Nederland bestaat al enkele jaren. In 2007 hebben leerlingen klachten geuit over de kwaliteit van het onderwijs, naar aanleiding waarvan de Commissie Parlementair Onderzoek Onderwijsvernieuwingen is ingesteld. Deze commissie kwam tot het oordeel dat het eerder ingevoerde *nieuwe leren* – minder kennisoverdracht, meer zelfstandig leren – onvoldoende effectief was (Commissie Parlementair Onderzoek Onderwijsvernieuwingen, 2008). De negatieve ontwikkelingen lijken nog niet te zijn gekeerd. Recentelijk rapporteerde De Volkskrant over een dalend vertrouwen van ouders in het basis- en voortgezet onderwijs (De Graaf, 2012). Uit een onderzoek van Elzinga (2012) kwam naar voren dat de meeste ouders een betere inzet van leerkrachten verwachten. De meerderheid (56%) vond dat leerlingen al hard genoeg werken en zesentachtig procent van de ouders vond dat leerkrachten doortastender zouden moeten optreden om de leerlingen beter bij de les te houden. Intussen blijven de leerprestaties onder Nederlandse middelbare scholieren dalen (Commissie Parlementair Onderzoek Onderwijsvernieuwingen, 2008; Inspectie van het Onderwijs, 2011b; Inspectie van het Onderwijs, 2012; OECD, 2010).

De sleutel voor verbetering van de onderwijskwaliteit wordt gezocht in de opbrengstgerichte aanpak, waarbij de principes van de *oplossingsgerichte benadering* worden gevolgd (Ledoux, Blok, Boogaard, & Krüger, 2009; Visscher & Ehren, 2011). Deze - uit de Verenigde Staten afkomstige - benadering heeft als doel het maximale uit alle leerlingen te halen door onderwijs op een passend niveau aan te bieden en hun leerprestaties intensief te volgen (Chafouleas, Volpe, Gresham, & Cook, 2010; McIntosh et al., 2011; Tilly, 2008). Het fundament voor de opbrengstgerichte aanpak en de oplossingsgerichte benadering is kwalitatief hoog onderwijs met een leerlingvolgsysteem als belangrijk hulpmiddel (Inspectie van het Onderwijs, 2011a; McMaster & Espin, 2007). De data uit het leerlingvolgsysteem worden niet alleen gebruikt voor beslissingen op leerlingniveau. Zij dienen ook als feedback voor leerkrachten ter verbetering van de manier waarop zij lesgeven (Oomens, Van Aarsen, Van Eck, & Kieft, 2008; Tilly, 2008). Er wordt afstand genomen van de opvatting dat ondersteuning pas nodig is wanneer problemen een significante omvang aannemen (McIntosh et al.; Merrell, 2010; Visscher & Ehren, 2011). Het uitgangspunt is dat er dient te worden gehandeld zodra er enige discrepantie bestaat tussen het huidige functioneren van een leerling en het verwachte niveau

van functioneren, bepaald aan de hand van de prestaties van de betreffende klas. De aanpak is universeel: niet alleen risicoleerlingen, maar alle leerlingen worden gescreend en gevolgd (Inspectie van het Onderwijs, 2011a). Aldus wordt de kans verkleind dat leerlingen over het hoofd worden gezien of dat problemen te hoog oplopen (McIntosh et al.; Shinn, Good, & Parker, 1998). Kenmerkend voor de opbrengstgerichte aanpak en de oplossingsgerichte benadering is verder dat de ontwikkelingen van de leerlingen in de natuurlijke onderwijssetting worden gevolgd (Briesch, Chafouleas, & Riley-Tillman, 2010; Ledoux et al., 2009). Op deze wijze wordt rekening gehouden met de context waarbinnen het onderwijs plaatsvindt.

Onder basisscholen is de opbrengstgerichte aanpak zeer effectief gebleken (Ledoux et al.). Volgens een onderzoek van de Inspectie van het Onderwijs (2011a) laat 91% van de basisscholen die opbrengstgericht werken, een positieve trend ten aanzien van leerprestaties zien, tegen 56% van de basisscholen die deze vorm van werken niet gebruiken. Een mogelijke verklaring voor de aanhoudende negatieve het afnemende vertrouwen van leerlingen en ouders in het voortgezet onderwijs en de dalende leerprestaties is dat het opbrengstgericht werken op middelbare scholen nog weinig in praktijk is gebracht. In 2011 voldeed slechts een vijfde van de middelbare scholen aan de voorwaarden die de Inspectie van het Onderwijs aan het opbrengstgericht werken stelt (Ministerie van Onderwijs, Cultuur en Wetenschap, 2011; Visscher & Ehren, 2011).

Leervorderingen

In het kader van het opbrengstgericht werken gaat steeds meer aandacht uit naar het volgen en optimaliseren van leervorderingen (Inspectie van het Onderwijs, 2011b en 2012). Een effectief gebleken model hiervoor is Response to Intervention, een uitwerking van de oplossingsgerichte benadering (Hagans, 2008; McIntosh et al.). Volgens dit *Three-tiered Model* vindt de begeleiding van leerlingen op drie niveaus (*tiers*) plaats: het reguliere niveau (*tier I*), het subgroepniveau (*tier II*) en het individuele niveau (*tier III*). Leerlingen die op het reguliere klassenniveau onvoldoende functioneren, krijgen aanvullend begeleiding op *tier II* of *tier III*-niveau, afhankelijk van de discrepantie tussen hun functioneren en het verwachte functioneren. Aan de hand van wekelijkse danwel maandelijkse metingen wordt geëvalueerd of de ingezette onderwijsinterventies het gewenste effect hebben (Tilly, 2008). Voor deze wijze van meten is essentieel dat de meetinstrumenten geschikt zijn voor herhaald gebruik. Aan de hand van deze metingen kan nauwkeurig worden ingeschat of een leerling vorderingen heeft gemaakt en het verwachte niveau benadert. Onder scholen in de Verenigde Staten wordt het *Three-tiered Model* veelal in combinatie met *curriculum-based measurement* toegepast (McIntosh et al.; Tilly, 2008). Daarbij worden leervorderingen op frequente basis gemeten aan de hand van de stof die wordt onderwezen (McMaster & Espin, 2007). Aldus wordt rekening gehouden met inhoud van de lessen en de manier waarop de lessen worden gegeven. De metingen geven op een systematische wijze inzicht in de effectiviteit van het gegeven onderwijs (Ledoux et al.).

Op basis hiervan kunnen leerkrachten gefundeerd en bewust beslissen of, en zo ja, op welke wijze zij hun lessen dienen te verbeteren (Visscher & Ehren, 2011).

Leergedrag

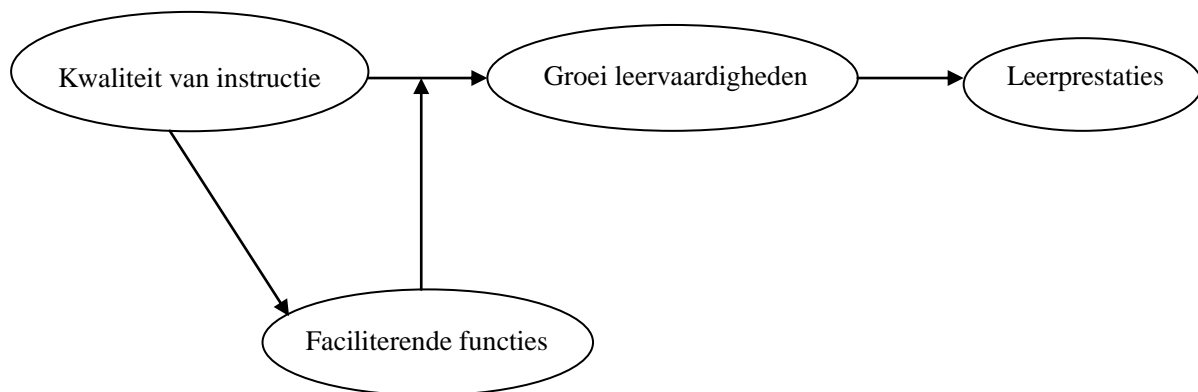
Leervordering vormt slechts één van de aspecten die zijn verbonden aan leerprestaties. Een ander aspect is leergedrag, dat wil zeggen: het gedrag dat leerlingen in de klas vertonen naar aanleiding van aangeboden lessen. Door naast het cognitief functioneren het leergedrag systematisch en gedetailleerd in kaart te brengen, verkrijgt men een completer beeld van de leerling (Shapiro, 2004a). Dit sluit aan bij de gedachte van *Functional Behavioral Assessment* (Nock & Kurtz, 2005; Steege & Brown-Chidsey, 2005). Volgens deze benadering verkrijgt men via leergedrag inzicht in de manier waarop een leerling leert, waar eventuele problemen vandaan komen en hoe deze kunnen worden aangepakt.

Volgens Cobb (1972), DiPerna (2006) en Greenwood (2002) bieden metingen een betere weergave van de werkelijkheid naarmate leergedrag specifieker in kaart wordt gebracht. In de literatuur is taakgerichtheid - dat wil zeggen: de mate van actieve deelname aan lessen - gebleken als operationalisatie van het construct leergedrag (o.a. Carroll, 1989; Cobb, 1972; DiPerna, 2006; Greenwood, Delquadri, & Hall, 1984; Greenwood, 1991; Wentzel, 1993). Taakgericht gedrag krijgt extra betekenis wanneer het wordt gezien in relatie tot tijd (DiPerna, 2006; Greenwood, 1991; Morgan, Farkas, Tufis, & Sperling, 2008; Shapiro, 2004a). De rol van tijd komt duidelijk naar voren in het *Model of School Learning* van Carroll (1963), een theoretisch model voor het meten en het veranderen van gedrag in de klas (Millman, Bieger, Klag, & Pine, 1983; Shapiro, 2004a). Volgens dit model worden leerprestaties hoofdzakelijk bepaald door vijf factoren: (1) de tijd die een leerling nodig heeft om iets te leren, (2) de tijd die de leerling daarvoor krijgt, (3) de tijd die de leerling aan het leren besteedt, (4) de kwaliteit van de instructie en (5) het vermogen om de instructie te begrijpen. Het idee achter dit model is dat een leerling beter zal presteren, naarmate hij of zij de beschikbare tijd beter benut. Het uitgangspunt daarbij is dat de kwaliteit van de instructie is afgestemd op het vermogen van de leerling om de instructie te begrijpen (Carroll, 1989; Millman et al., 1983).

Relatie tot Leerprestaties

Onderzoek heeft uitgewezen dat leergedrag en leerprestaties met elkaar samenhangen: beter presterende leerlingen vertonen over het algemeen meer positief leergedrag dan slechter presterende leerlingen (Algozzine, Wang, & Violette, 2011; Greenwood et al., 1984; Hattie, 2009). DiPerna (2006) kenmerkt taakgerichtheid, naast leervaardigheid, motivatie en sociale vaardigheid, als een van de vier faciliterende functies in het leerproces. In navolging van Greenwood (1996) beschouwt DiPerna (2006) taakgerichtheid als een product van de instructies van de leerkracht. Volgens het model van DiPerna (2006) versterken de faciliterende functies de mate waarin

leervaardigheden tot leerprestaties leiden. Hoe hoger de kwaliteit van instructie, hoe sterker deze invloed is (Figuur 1).



Figuur 1. *Invloed van faciliterende functies op leerprestaties (DiPerna, 2006)*

DiPerna (2006) en Greenwood (1996) gaan ervan uit dat leergedrag het niveau van leerprestaties beïnvloedt. Een longitudinaal onderzoek van Morgan et al. (2008) onder 11.515 basisschoolleerlingen bevestigde dit. Leerlingen van gemiddeld 5,5 jaar die onvoldoende taakgericht waren, hadden anderhalf jaar later drie keer zo vaak problemen op leesgebied als hun taakgerichte leeftijdsgenoten. Uit het onderzoek van Morgan et al. bleek voorts dat leerprestaties andersom ook het leergedrag kunnen beïnvloeden. Benedengemiddelde lezers vertoonden na anderhalf jaar twee keer zo vaak een gebrekkige taakgerichtheid als de andere leerlingen. Gelet op dit bidirectionele verband pleiten Morgan et al. ervoor om de begeleiding van leerlingen niet alleen te richten op leervorderingen, maar tevens op leergedrag.

Genoemde onderzoeken van Algozzine et al. (2011), Greenwood et al. en Morgan et al. hebben betrekking op basisschoolkinderen. Het is mogelijk dat leergedrag van middelbare scholieren zich op een andere wijze tot leerprestaties verhoudt. Onder middelbare scholieren lijken invloedrijke factoren, zoals motivatie en huiswerk, andere rol in het leerproces te spelen dan onder leerlingen op de basisschool (Cooper, Robinson, & Patall, 2006; Lepper, Henderlong Corpus, & Iyengar, 2005; Muhlenbruck, Cooper, Nye, & Lindsay, 2000).

Leergedrag is wellicht een belangrijke factor voor schoolsucces in groepen die meer begeleiding nodig hebben. Een voorbeeld van een dergelijke groep binnen het middelbaar onderwijs is de groep *nieuwkomers*. Deze leerlingen wonen maximaal twee jaar in Nederland en beheersen de Nederlandse taal nog onvoldoende om deel te kunnen nemen aan het regulier onderwijs (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006). Naast grote veranderingen, waaronder emigratie en schoolwisselingen, en soms traumatische ervaringen hebben nieuwkomers vaker dan gemiddeld te maken met een didactische achterstand ten opzichte van leeftijdsgenoten die in Nederland zijn opgegroeid (Oomens, Van Aarsen, & Kooij, 2012). Deze factoren kunnen het leerproces beïnvloeden

(Hattie, 2009). Van migrantenkinderen in de Verenigde Staten is bekend dat zij gemiddeld lagere leerprestaties behalen dan hun leeftijdsgenoten (Elbers, 2010). Mogelijk geldt dit ook voor nieuwkomers in Nederland. Het onderwijs voor nieuwkomers verschilt in meerdere opzichten van het reguliere onderwijs. Ten eerste is erop gericht om de taalachterstand van de leerlingen in te halen (Ministerie van Onderwijs, Cultuur en Wetenschap, 2006). Daarnaast is er een grote variatie aan cultuur en leeftijd en stromen leerlingen gedurende het gehele leerjaar in en uit. Gebleken is dat deze afwijkende omstandigheden een intensievere inzet van didactische en communicatieve leerkrachtcompetenties vereisen (Van Vijfeijken & Van Schilt-Mol, 2012).

Vergelijkbaar met de bovenstaande groep zijn leerlingen in het praktijkonderwijs. Tot deze vorm van speciaal onderwijs worden leerlingen toegelaten die moeite hebben met leren en naar verwachting geen VMBO-diploma zullen kunnen behalen. Het praktijkonderwijs richt zich vooral op de begeleiding van de leerlingen voor deelname aan het arbeidsproces (Inspectie, 2012). Evenals in het onderwijs voor nieuwkomers ontbreekt de reguliere examen- en diplomastructuur (Onderwijsraad, 2010).

Interventies

De mate waarin een leerling taakgericht gedrag vertoont, biedt belangrijke handvatten voor de inzet van interventies. Met het juiste inzicht in het leergedrag kunnen interventies doeltreffender worden ingezet en kan een leerkracht effectiever lesgeven. De leerkracht is minder tijd kwijt aan het bewaken van de orde, heeft minder stress in de omgang met leerlingen en houdt meer tijd over voor het bieden van onderwijs (Goei & Kleinen, 2009; Muscott, Mann, & LeBrun, 2008). Soms lijkt een leerling voldoende taakgericht gedrag te vertonen maar bestaat het vermoeden dat de instructie beter zou moeten worden afgestemd op het leervermogen van de leerling. In dat geval ligt een cognitieve onderwijsinterventie voor de hand, zoals het geven van een aanvullende instructie voor specifieke deelvaardigheden (Burns, Coddington, Boice, & Lukito, 2010). Toont een leerling te weinig taakgerichtheid tijdens de les, dan komen gedragsgerichte interventies aan de orde (Shapiro, 2004a). Een voorbeeld van een effectief gebleken interventie is *self-monitoring* (Wood, Murdock, Croning, Dawson, & Kirby, 1998; Ratvon, 2008; Briesch & Chafouleas, 2009). Hierbij monitoren leerlingen hun eigen gedrag in de klas en leren zij hun gedrag aan te passen zodat het taakgerichte gedrag toeneemt en het niet-taakgerichte gedrag afneemt. Bij een - weliswaar kleine - steekproef van vier zorgleerlingen registreerden Wood et al. (1998) na toepassing van self-monitoring een significante toename niet aan taakgericht gedrag als ook een lichte verbetering van de leerprestaties.

Ter verbetering van leergedrag kan tevens worden gedacht aan effectiever klassenmanagement (Inspectie van het Onderwijs, 2011b; Visscher & Ehren, 2011). Daarbij zetten leerkrachten hun tijd, instructies en activiteiten op zodanige wijze in dat leerlingen optimaal profiteren van het geboden onderwijs (Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). Door bijvoorbeeld lessen beter op de verschillende niveaus van leerlingen af te stemmen of instructies in kleinere

groepen te geven, kunnen leerlingen worden gestimuleerd om zich actiever op te stellen en neemt de kans op betere leerprestaties toe (DiPerna, 2006; Greenwood, 1991; Greenwood, 1996; Hart, Massetti, Fabiano, Pariseau, & Pelham, 2011).

Systematische Directe Observatie

De meest gebruikte methode voor het meten van leergedrag is systematische directe observatie (Greenwood, Horton, & Utley, 2002; Shapiro & Heick, 2004; Volpe, DiPerna, Hintze, & Shapiro, 2005). Bij systematische directe observatie worden indicatoren van leergedrag in een alledaagse onderwijssetting op systematische wijze geobserveerd en gekwantificeerd (Shapiro, 2004a). De data bestaan uit de frequentie waarin bepaalde typen leergedrag voorkomen.

Van het scala aan bestaande meetinstrumenten sluit systematische directe observatie het beste aan op de oplossingsgerichte benadering (Brown-Chidsey, 2005; McIntosh et al.; Shapiro, 2004a; Tilly, 2008). Zo worden de gedragingen van de leerling in interactie met de omgeving geobserveerd terwijl de les plaatsvindt. Daardoor worden invloeden van het lokaal, de soort les, de leerkracht en de medeleerlingen in het geobserveerde leergedrag verweven (Nock & Kurtz, 2005). Gestandaardiseerde en genormeerde observatiesystemen, zoals de Academic Engaged Time Code (Walker & Severson, 1990) en de Direct Observation Form (Achenbach, 1986), zijn om die reden minder geschikt. Op basis van deze systemen worden observatiedata vergeleken met de scores van een grote normgroep. Omgevingsfactoren, zoals interactie met de leerkracht, worden kunstmatig gelijk gehouden terwijl deze factoren bij de probleemoplossende benadering juist de kern van de assessment vormen. Anders dan bij bijvoorbeeld de Attention-Deficit Hyperactivity Disorder School Observatie Code (Gadow, Sprafkin, & Nolan, 1996) en Classroom Observation Code (Abikoff & Gittelman, 1985), gaat het bij systematische directe observatie niet zozeer om het signaleren van klinisch afwijkend leergedrag, maar om het screenen van leergedrag op schoolbreed niveau. Men meet immers om alle leerlingen te kunnen volgen in hun ontwikkeling om vanuit de betekenis van het getoonde gedrag beter te kunnen begrijpen hoe zij functioneren (Steege & Brown-Chidsey, 2005). Gunstig is dan ook dat met systematische directe observatie grote en kleine veranderingen in kwaliteit of omvang kunnen worden geïdentificeerd (Briesch et al., 2010; Volpe et al., 2005). Systematische directe observatie kan bovendien herhaaldelijk plaatsvinden zonder dat dit tot praktische of psychometrische bezwaren leidt (Briesch et al., 2010).

Ten opzichte van vragenlijsten en andere meetinstrumenten waarbij informatie via derden wordt ingewonnen, is het voordeel van systematische directe observatie dat er in hoge mate objectief en onafhankelijk van de perceptie van de leerkracht wordt gemeten (Briesch et al., 2010; Nock & Kurtz, 2005). De meetresultaten zijn bovendien specifiek en daarmee beter geschikt voor het volgen van individuele leerlingen over langere termijn (Hintze & Matthews, 2004). De registratie van het gedrag vindt op directe wijze plaats, dat wil zeggen: op het moment dat het betreffende gedrag zich voordoet. Hiermee wordt eventuele ruis door retrospectie voorkomen (Christ, Riley-Tillman, &

Chafouleas, 2009). Bij de afname van vragenlijsten bestaat er een reële kans op *bias* door sociale wenselijkheid of vooroordelen (Nock & Kurtz, 2005). Dit nadeel bestaat ook bij direct behavior rating. Daarbij wordt een leerkracht geacht de frequentie van de typen gedragingen aan te geven op een ordinale beoordelingsschaal. Briesch et al. (2010) hebben een vergelijkend onderzoek verricht naar de generaliseerbaarheid van systematische directe observatie en *direct behavior rating*. Hieruit kwam naar voren dat de uitkomsten van direct behavior rating aanzienlijk werden vertroebeld door de perceptie van de leerkracht: 20% van de error variantie werd veroorzaakt door leerkrachtfactoren. Bij systematische directe observatie was de invloed van de leerkracht verwaarloosbaar.

Systematische directe observatie heeft enkele zwakke kanten die een nadere beschouwing verdienen. Ten eerste nemen observaties veel tijd in beslag (Chafouleas et al., 2010; Nock & Kurtz, 2005). De duur van de meest gangbare observatiesystemen varieert van 15 tot 32 minuten per leerling (Volpe et al.). Ten tweede is nog weinig bekend over de vraag welke aspecten van leergedrag voldoende convergente validiteit van systematische directe observatie opleveren (Chafouleas et al.; Hintze & Matthews, 2004, Nock & Kurtz, 2005).

Indicatoren van Leergedrag

De convergente validiteit van systematische directe observatie wordt voor een groot deel bepaald door de selectie van de gedragingen die worden geobserveerd (Volpe et al.). Wanneer men leerprestaties wil verhogen door de aanpak van leergedrag, zullen de geobserveerde gedragingen een nauwkeurige en realistische weergave moeten zijn van leergedrag dat verband houdt met leerprestaties (Hintze, 2005). Leergedrag in de zin van taakgerichtheid kan nader worden gespecificeerd als actief en aan taak leergedrag - samen positief leergedrag - en niet aan taak en storend leergedrag - samen negatief leergedrag. Een onderzoek van Riley-Tillman, Chafouleas, Sassu, Chanese en Glazer (2008), waarbij systematische directe observatiemetingen werden vergeleken met metingen via direct behavior rating, wees uit dat aan taak en storend leergedrag valide indicatoren zijn. Actief leergedrag lijkt significant verband te houden met leerprestaties (Greenwood et al., 1984; Shapiro, 2004a). Door actief aan de les deel te nemen lijken leerlingen significant beter te profiteren van geboden lessen (Greenwood, 1996). Deze vorm van positief leergedrag dient te worden onderscheiden van aan taak leergedrag, zoals het aankijken van de leerkracht. Greenwood et al. ondervonden dat dit type leergedrag geen significant verband hield met leerprestaties (DiPerna, 2006; Shapiro, 2004a). Men zou om deze reden kunnen overwegen om aan taak leergedrag, de passieve vorm van positief leergedrag, niet te meten. Om verschillende typen leergedrag in relatie tot elkaar te kunnen analyseren kan de mate van aan taak leergedrag echter waardevolle informatie bieden (Shapiro, 2004a). Bovendien verhoogt het registreren van aan taak leergedrag de nauwkeurigheid van de meting doordat het construct op specifiekere wijze wordt geoperationaliseerd (Hintze, 2005). Leerlingen die veel negatief leergedrag vertonen en weinig taakgericht werken, presteren vaker zwak (Morgan et al.). Negatief leergedrag dat tevens storend is, dat wil zeggen ongehoorzaam, vijandig of

agressief, lijkt een nog sterker verband te houden met zwakke leerprestaties dan niet aan taak leergedrag, de passieve vorm van negatief leergedrag (Arnold, 1997; DiPerna, 2006; Greenwood, 1996).

Veel gehanteerde observatie-instrumenten voor systematische directe observatie zijn de BOSS (Shapiro, 2004b), de MS-CISSAR (Greenwood et al., 2002) en de SECOS (Saudargas, 1997). De MS-CISSAR en de SECOS omvatten beide een relatief groot aantal categorieën. De MS-CISSAR hanteert naast drie categorieën voor leergedrag (taakgericht gedrag, neutraal aan taak gedrag en negatief gedrag) vijf categorieën ter nadere aanduiding van de omgeving en drie categorieën voor de specificatie van de leerkracht. De SECOS omvat zes typen toestanden en zes typen gebeurtenissen. Door het grote aantal gedragscategorieën zijn de MS-CISSAR en de SECOS complex in afname en interpretatie, hetgeen de efficiëntie beperkt en de kans op error vergroot (Hintze, 2005; Shapiro, 2004a).

Bij de BOSS wordt uitgegaan van een eenvoudiger systeem met vijf gedragscategorieën: actief, passief, niet-betrokken motorisch, niet-betrokken verbaal en niet-betrokken passief leergedrag. De voorgeschreven duur van vijftien aaneengesloten minuten is echter lang terwijl het aantal lessen, gedurende welke leerlingen worden geobserveerd, is beperkt. Ter verhoging van de efficiëntie en de betrouwbaarheid lijkt het logischer om het aantal meetmomenten te verhogen en de observatieduur per meting te verkorten (Hintze, 2005).

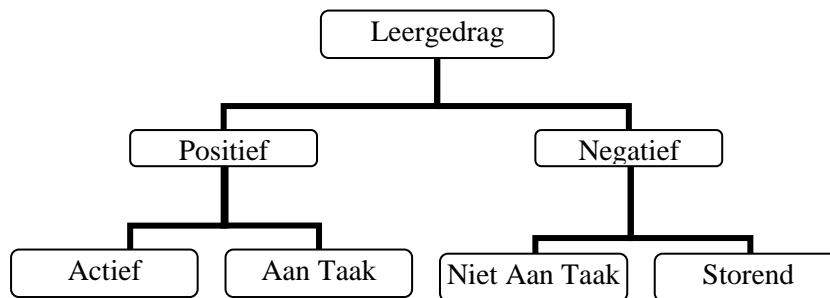
Voor de registratie van twee typen leergedrag, actief en passief leergedrag, hanteert de BOSS *momentary time sampling*. De hiermee data bestaan uit momentopnamen per tijdsinterval. De drie overige typen leergedrag worden gemeten met behulp van *interval recording*, waarbij de observatie gedurende het gehele tijdsinterval voortduurt. Het nadeel van de laatstgenoemde manier van meten is dat leerlingen langer achter elkaar worden geobserveerd, waardoor de kans op ongewenst beïnvloeding van het leergedrag door de observatie wordt vergroot (Hintze, 2005). *Momentary time sampling* lijkt derhalve meer valide en betrouwbare data op te leveren (Saudargas, 1997).

Het kan zinvol zijn om een onderscheid naar type les te maken. Taakgerichtheid tijdens rekenlessen kan een andere betekenis hebben dan taakgerichtheid tijdens taallessen (Cobb, 1972; Greenwood, Horton, & Utley, 2002). Bij de BOSS wordt per type les nog een onderscheid gemaakt naar type instructie, zoals klassikale instructie of zelfstandig werken (Shapiro, 2004b). Het bezwaar van deze grote mate van specificiteit is dat er meer meetmomenten vereist zijn om per variabele over voldoende valide en betrouwbare data te kunnen beschikken, hetgeen de efficiëntie van het meetinstrument beperkt (Messick, 1995).

Huidige Onderzoek

Voor het huidige onderzoek is de Leerling Observatie in de Klas (LOK) ontwikkeld. Met dit systeem wordt gestreefd naar een valide, betrouwbare en efficiënte wijze van observeren die aansluit bij de principes van de oplossingsgerichte benadering. De typen leergedrag die worden gemeten, zijn

actief, aan taak, niet aan taak en storend leergedrag. De geobserveerde typen leergedrag zijn in Figuur 2 weergegeven.



Figuur 2. Typen geobserveerd leergedrag

De eerste vraag die is onderzocht, luidt: *Welke typen leergedrag hebben een voldoende convergente validiteit?* Dit is onderzocht aan de hand van de deelvraag: *Welke van de typen geobserveerd leergedrag vertonen voldoende verband met mentorbeoordelingen van leergedrag?* In de tweede plaats is de volgende vraag beantwoord: *Welke typen leergedrag houden voldoende verband met leerprestaties?* Leidend hiervoor was de deelvraag: *In hoeverre houden de verschillende typen geobserveerde leergedrag verband met leerprestaties?*

Aangezien uit eerder onderzoek is gebleken dat leerlingen zich tijdens rekenlessen anders gedragen dan tijdens lessen Nederlands, is voor beide onderzoeksvragen onderzocht of de gevonden correlaties verschilden per type les. Voorts is rekening worden gehouden met het feit dat de geobserveerde leerlingen afkomstig waren van twee verschillende scholen. Verwacht werd dat de observatiedata overeen zouden komen met de beoordelingen van de mentoren. Daarnaast was de verwachting dat de typen actief, niet aan taak en storend leergedrag een sterk verband zouden vertonen met leerprestaties.

Methode

Onderzoeksgroep

Het huidige onderzoek komt voort uit een initiatief van een scholengroep die onderzoek wilde laten verrichten naar manieren om de leerontwikkeling van leerlingen te optimaliseren. De onderzoeksgroep bestond uit 54 leerlingen uit het tweede leerjaar van het voortgezet onderwijs. Deze groep bestond voor 44% uit jongens en voor 56% uit meisjes. De gemiddelde leeftijd was 15,23 jaar ($SD = 1,49$). De leerlingen zaten in vijf verschillende klassen, verdeeld over twee vestigingen van de scholengroep in een grote stad in de Randstad. Op de ene vestiging kregen twee van deze klassen

Praktijkonderwijs op indicatie van de Regionale Verwijzingscommissie Voortgezet Onderwijs. In de regel wordt een dergelijke indicatie verleend wanneer een leerling laag begaafd is of een lichte verstandelijke beperking heeft (IQ 60-77) en aan het einde van de basisschool een achterstand heeft op minstens twee van de volgende vier terreinen: rekenen, begrijpend lezen, technisch lezen en spellen (Rijksoverheid, 2012). De overige drie klassen op de andere vestiging bestonden uit *nieuwkomers* ($n = 32$). Van 22 nieuwkomers was het non-verbale IQ bekend ($M = 80,95$, $SD = 9,12$). Eén leerling had een IQ van 55. De rest had een IQ van 70 of hoger ($MAX = 99$). In Tabel 1 zijn de kenmerken van de leerlingen van het praktijkonderwijs en de nieuwkomers weergegeven.

Tabel 1

Onderzoeksgroep per onderwijstype

	<i>N</i>	Jongen	Meisje	M_{leeftijd}	SD_{leeftijd}	MIN	MAX
Praktijkonderwijs	22	13	9	14,73	0,70	13,42	16,08
Nieuwkomers	32	11	21	15,58	1,78	12,92	19,58
Totaal	54	24	30	15,23	1,49	12,92	19,58

De klassen die in het kader van dit onderzoek voor observatie in aanmerking kwamen, zijn door de scholengroep voorgedragen. Voorgaand aan de observaties zijn de ouders van de betreffende leerlingen hiervan op de hoogte gesteld door middel van een Nederlandse en een Engelse brief. Wanneer ouders bezwaar maakten tegen deelname aan het onderzoek, werd hun kind niet geobserveerd.

Onderzoeksofzet

Het onderzoek is verricht op basis van een 2 x 2 design. Van iedere leerling zijn in totaal vijf minuten van een geobserveerde rekenles en vijf minuten van een geobserveerde les Nederlands geselecteerd. Zo kon er worden geanalyseerd op basis van data, verzameld tijdens lessen rekenen, lessen Nederlands of beide. Om invloeden, samenhangend met het meetmoment, te beperken zijn *at random* data uit de eerste, danwel de tweede observatie week geselecteerd (Tabel 2).

Voor de analyse van het verband met leerprestaties is een kleinere subgroep onderzocht aangezien alleen van de nieuwkomers de rapportcijfers bekend waren. Van een van de leerlingen ontbrak het rapportcijfer voor rekenen. Ten behoeve van de vergelijkbaarheid van de leerlingen zijn de data van deze leerling weggelaten, voor zover het verband met rapportcijfers werd geanalyseerd. De hierna resterende subgroep bevatte 31 leerlingen.

Tabel 2

Aantallen leerlingen per les en observatieweek (N = 54)

School	Rekenen		Nederlands	
	Week 1	Week 2	Week 1	Week 2
Praktijkonderwijs	11	11	12	10
Nieuwkomers	16	16	16	16
Totaal	27	27	28	26

Gedragsobservatie

Met het instrument LOK wordt observeerbaar leergedrag in kaart gebracht door middel van systematische directe observatie. Gedurende één minuut werd iedere tien seconden genoteerd welke van de vier verschillende typen leergedrag zich voordeed. Na iedere minuut werd een volgende leerling geobserveerd. Op deze wijze zijn van alle leerlingen gedurende tien afzonderlijke minuten data verzameld. Vijf van deze minuten maakten deel uit van een rekenles. De overige vijf minuten waren onderdeel van een les Nederlands.

De observaties zijn verricht door masterstudenten die vooraf een videotraining kregen. Aan de hand van een film werd geoefend met het identificeren van de vier verschillende typen leergedrag. Waar meerdere interpretaties mogelijk bleken, werden concrete afspraken gemaakt. Een leerling die nog niet daadwerkelijk schreef, maar wel zijn of haar pen naar het papier toe bewoog om te gaan schrijven, werd bijvoorbeeld aangemerkt als zijnde actief. De training werd voortgezet tot een vooraf bepaald niveau van interbeoordelaarsbetrouwbaarheid was behaald. Wenselijk werd een correlatie van 0,70 aangehouden (Leary, 2011).

Indicatoren van Leergedrag

Actief leergedrag omvatte taakgericht gedrag. Hiervan was sprake wanneer de leerling mondeling, schriftelijk of motorisch reageerde als respons op de instructie van de leerkracht. Dit betrof onder meer het beantwoorden van een vraag van de leerkracht of het praten tegen een medeleerling over de lesstof wanneer dat was toegestaan.

Onder *aan taak leergedrag* werd verstaan passief leergedrag met zichtbare interesse voor de lesstof, zoals stil lezen, kijken naar het bord of luisteren naar de leerkracht of een medeleerling, passend binnen de gegeven instructies.

Niet aan taak leergedrag was aan de orde wanneer een leerling gericht was op iets anders dan hetgeen de leerkracht hem of haar had opgedragen, zonder dat storend was voor anderen. Voorbeelden hiervan zijn staren of irrelevante informatie lezen.

Leergedrag werd als *storend* beschouwd wanneer verbaal of motorisch gedrag een inbreuk maakte op de leeromgeving van zichzelf of van anderen. Hierbij kan worden gedacht aan het maken van storende geluiden, praten voor zijn of haar beurt, weigeren om een instructie op te volgen, verbaal of fysiek agressief gedrag vertonen.

De samengevoegde variabelen *positief* en *negatief* leergedrag werden van elkaar onderscheiden aan de hand van de vraag of een leerling zich conform de instructies van de leerkracht gedroeg. Positief leergedrag is gevormd door een optelling van de percentages actief en aan taak leergedrag. Negatief leergedrag bestond uit de som van de percentages niet aan taak en storend leergedrag.

De frequenties van de vier verschillende typen leergedrag werden omgezet in percentages. Wanneer bij een leerling gedurende één minuut observeren bijvoorbeeld twee keer storend leergedrag en vier keer actief leergedrag werd geregistreerd, leverde dit een percentage op van 33,3% storend leergedrag en 66,6% actief leergedrag.

Mentorscore

Enkele weken na het observeren hebben de mentoren van de klassen de leerlingen beoordeeld op leergedrag aan de hand van een 5-punts Likert-schaal. Deze mentoren, twee mannen en drie vrouwen, waren alle docent Nederlands van de betreffende leerlingen. Hen werd gevraagd het leergedrag van hun leerlingen te beoordelen met 1 (slecht), 2 (minder goed), 3 (goed) of 4 (zeer goed). In de instructie voor de mentoren is leergedrag gedefinieerd als “het gedrag dat is verbonden met en invloed heeft op het leren, zoals het luisteren naar instructies en het zelfstandig kunnen werken”.

Rapportcijfers

Alleen van de groep nieuwkomers kon over de rapportcijfers worden beschikt. Gebruikt zijn de rapportcijfers voor rekenen en Nederlands die ten tijde van de observatie het meest recent waren.

Procedure

Alle leerkrachten en leerlingen waren vooraf op de hoogte gesteld van de observaties. De leerkrachten waren gevraagd om zo min mogelijk aandacht aan de observatoren te schenken. De lessen werden gegeven zoals de leerkrachten dat gewend waren te doen. Bij de meeste rekenlessen werd grotendeels zelfstandig gewerkt. Bij lessen Nederlands gaf de leerkracht veelal klassikale instructies of werd er interactief met de gehele klas of in groepjes gewerkt. Enkele leerlingen deden een toets tijdens de observaties. Voor de dataverzameling namen de observatoren plaats in het lokaal schuin voor de te observeren leerlingen zodat de gezichten van de leerlingen goed zichtbaar waren en de observatoren weinig opvielen. Om ervoor zorg te dragen dat de leerlingen aan de juiste anonieme leerlingcodes werden gekoppeld, werd een leerlingcodelijst met foto's gehanteerd. Twee observatoren observeerden ieder een helft van de klas. Daarbinnen werden de leerlingen om de beurt een minuut lang geobserveerd. Iedere tien seconden kregen de observatoren via een koptelefoon een pieptoon te horen. Het gedrag dat de te observeren leerling op dat moment vertoonde, werd op het formulier gescoord. Het kijken naar de leerlingen geschiedde zo onopvallend mogelijk om beïnvloeding door de observator te minimaliseren. Tijdens een vijfde van de lessen observeerde een interbeoordelaar

gelijktijdig met een van de twee observatoren zodat de interbeoordelaarsbetrouwbaarheid kon worden gecontroleerd.

Data-analyse

Ter beantwoording van de onderzoeksvragen zijn de observatiedata geanalyseerd aan de hand van Pearson correlatietoetsen, univariate variantie-analyses en GLM herhaalde metingen. Aangezien deze toetsen een normaalverdeling van de variabelen veronderstellen, is voor alle variabelen gecontroleerd of de scheefheid en de gepiekttheid kleiner waren dan $|3,00|$. Voor zover dat niet het geval was, is op de betreffende variabele een logaritmische transformatie toegepast.

Om te voorkomen dat het werkelijke beeld van de onderzoeksgroep zou worden vertekend, is deze groep op uitbijters gecontroleerd. Een leerling werd als uitbijter beschouwd wanneer de frequentie van het leergedrag een waarde had die anderhalf tot drie keer de interkwartielafstand van de mediaan verwijderd was. Om de invloed van de uitbijters te kunnen inschatten, zijn de analyses zowel met als zonder uitbijters uitgevoerd.

Aan de hand van correlatietoetsen is onderzocht hoe de afzonderlijke typen leergedrag correleerden met de mentorscore danwel de rapportcijfers. De uitkomsten zijn tweezijdig getoetst, waarbij de grens van significantie werd gesteld op $p = 0,05$. Voor de bepaling van de sterkten van de correlaties is de tabel van Cohen (1992) aangehouden. Op basis hiervan werd een correlatie als sterk beschouwd bij $r = 0,50$, gemiddeld sterk bij $r = 0,30$ en zwak bij $r = 0,10$. Bij $r < 0,10$ werd een verband als zeer zwak beschouwd. De convergente validiteit werd als voldoende beschouwd bij een correlatie van $r \geq 0,30$. Effecten van de GLM herhaalde metingen en de variantie-analyses, zijn eveneens geïnterpreteerd aan de hand van de tabel van Cohen (1992), groot bij partial $\eta^2 = 0,14$, gemiddeld bij partial $\eta^2 = 0,06$ en klein bij partial $\eta^2 = 0,01$. De effecten golden als voldoende groot bij partial $\eta^2 \geq 0,06$.

Voor een bestudering van de relatie van de samengevoegde typen leergedrag met de mentorscore en de rapportcijfers is een univariate variantie-analyse gebruikt. Aangezien de percentages *positief* en *negatief leergedrag* volledig met elkaar samenhangen, is bij de bespreking van de analyses volstaan met de vermelding van de uitkomsten voor *positief leergedrag*.

Resultaten

Data-inspectie

Voor de analyse van het verband tussen de geobserveerde typen *leergedrag* en de toegekende mentorscores is onderzocht of de variabelen normaal waren verdeeld. Dit was niet het geval bij de variabelen *storend leergedrag* op basis van beide lessen ($z_{\text{scheefheid}} = 3,27$) en *storend leergedrag* op basis van de rekenlessen ($z_{\text{scheefheid}} = 3,94$). Voor de analyses met dit type leergedrag, met en zonder onderscheid naar type les, is de variabele *storend leergedrag* logaritmisch getransformeerd. De

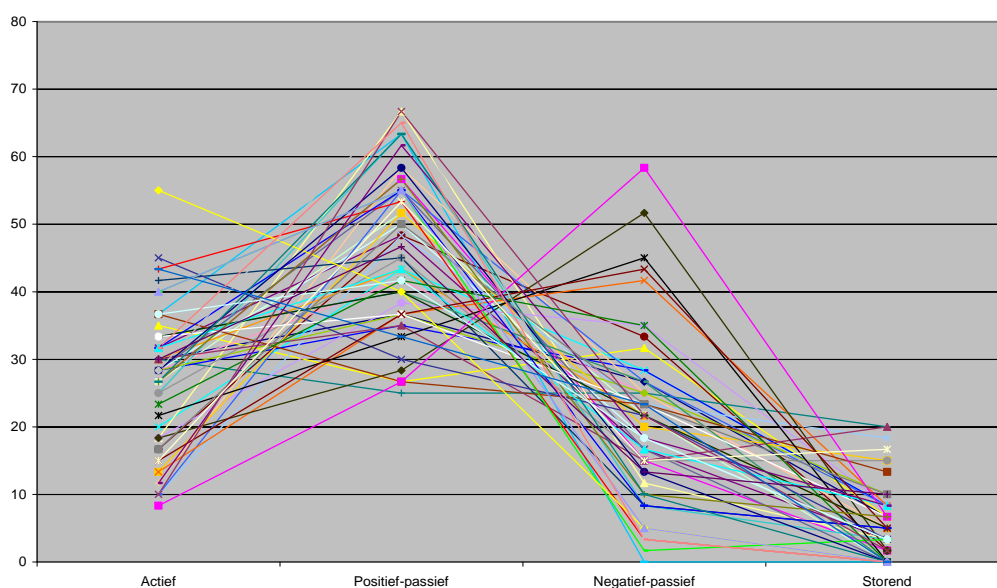
variabelen *actief*, *aan taak* en *niet aan taak leergedrag* en de samengevoegde variabelen *positief* en *negatief leergedrag* waren wel normaal verdeeld.

Aan de hand van boxplots voor de vier typen leergedrag zijn twee uitbijters gesignaleerd. Deze betroffen een jongen en een meisje uit twee verschillende klassen, van wie de frequentie van *niet aan taak leergedrag* anderhalf tot drie maal de interkwartielafstand van de mediaan was verwijderd. Het zojuist genoemde meisje bleek tevens de enige uitbijter te zijn ten aanzien van de samengevoegde typen leergedrag. Van dit meisje lagen de frequenties van *positief* en *negatief leergedrag* op een afstand van anderhalf tot drie maal de interkwartielafstand van de mediaan.

Aangezien de kans bestond dat de gevonden uitbijters een vertekend beeld van de werkelijkheid zouden geven, zijn alle analyses met en zonder de uitbijters gedaan. Weglating van de uitbijters bracht geen noemenswaardige verandering in de normaalverdelingen met zich mee. *Storend leergedrag* bleef als enige variabele niet normaal verdeeld.

Beschrijving Leergedrag

Vier typen leergedrag. In Tabel 3 zijn de beschrijvende gegevens van de vier typen leergedrag weergegeven. De verdelingen worden verder geïllustreerd in de onderstaande lijngrafiek (Figuur 3). *Aan taak leergedrag* kwam gemiddeld het grootste deel van de tijd voor ($M = 46,51\%$, $SD = 11,80$). De leerlingen vertoonden gemiddeld minder vaak *actief leergedrag* ($M = 26,98\%$, $SD = 10,36$) en *niet aan taak leergedrag* ($M = 20,86\%$, $SD = 12,44$). *Storend leergedrag* werd gemiddeld slechts 5,65% van de tijd vertoond ($SD = 5,48$). Slechts vier procent van leerlingen vertoonden meer dan 10% van de tijd *storend leergedrag*. Van de vier typen leergedrag vertoonde de variabele *niet aan taak leergedrag* de grootste variatie (MIN = 0,00%; MAX = 58,33%). De variabele *storend leergedrag* varieerde het minst (MIN = 0,00%; MAX = 20,00%).



Figuur 3. Percentages vier typen leergedrag per leerling ($N = 54$).

Samengevoegde typen leergedrag. Wanneer de typen leergedrag werden samengevoegd, bleek dat de geobserveerde leerlingen gemiddeld 73,49% van de tijd *positief leergedrag* (actief en aan taak) vertoonde ($SD = 14,25$). *Negatief leergedrag* (niet aan taak en storend gedrag) werd 26,51% van de tijd vertoond ($SD = 14,25$).

Tabel 3

Gemiddelde percentages (standaard deviaties) vier typen leergedrag en positief leergedrag (N = 54)

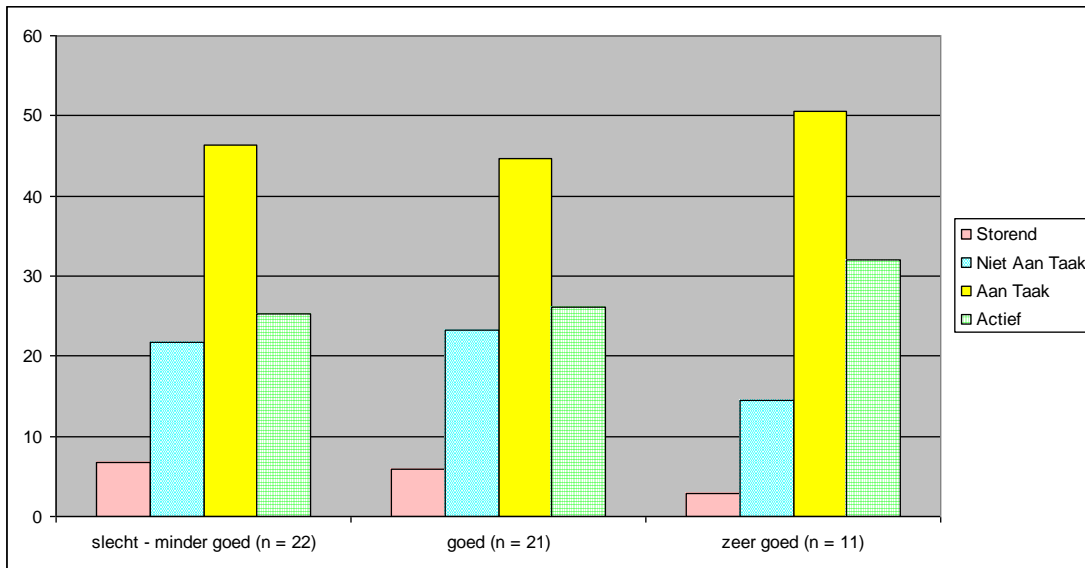
Type leergedrag	M (SD)
Actief	26,98% (10,36)
Aan Taak	46,51% (11,80)
Niet Aan Taak	20,86% (12,44)
Storend	5,65% (5,48)
Positief	73,49% (14,25)
Negatief	26,51% (14,25)

Mentorscore. De mentorscores in de gradaties *zeer goed*, *goed* en *minder goed* werden aan 11 leerlingen, 21 leerlingen respectievelijk 18 leerlingen toegekend. De mentorscore *slecht* werd aan vier leerlingen gegeven. Vier van de vijf mentoren kenden hun leerlingen mentorscores toe in minimaal drie van de vier gradaties. Eén mentor bleek zijn leerlingen uitsluitend de mentorscores *goed* en *minder goed* te hebben toegekend.

Aangezien GLM herhaalde metingen en variantieanalyses min of meer gelijke groepsgroottes vereisen, zijn voor deze toetsen de leerlingen met een mentorscore *slecht* samengevoegd met de leerlingen die een mentorscore *minder goed* hadden behaald. Wanneer er een onderscheid naar school werd gemaakt, werden de leerlingen met een mentorscore *zeer goed* om dezelfde reden samengevoegd met de leerlingen, aan wie een mentorscore *goed* was toegekend.

Convergente Validiteit

Vier typen leergedrag. Figuur 4 laat de gemiddelde percentages zien van de typen leergedrag per mentorscore. De uitkomsten van de correlatietoetsen zijn opgenomen in Tabel 4. Het verband tussen de mentorscore en *actief leergedrag* was volgens de tabel van Cohen (1992) gemiddeld sterk en significant ($r = 0,30$, $p = 0,03$). Het verband met *storend leergedrag* was zwakker en marginaal significant ($r = -0,25$, $p = 0,07$). De variabelen *aan taak leergedrag* ($r = 0,09$, $p = 0,53$) en *niet aan taak leergedrag* ($r = -0,21$, $p = 0,14$) leverden een zwak en insignificant verband op. Wanneer de twee uitbijters werden weggelaten, werd ook het verband met de variabele *niet aan taak leergedrag* significant ($r = -0,29$, $p = 0,04$). De verschillen bleven in dezelfde mate significant.



Figuur 4. Percentages vier typen leergedrag per mentorscore.

Tabel 4

Correlaties percentages vier typen leergedrag en positief leergedrag met mentorscore (N = 54)

Type leergedrag	Algemeen	Rekenen	Nederlands
	<i>r</i> (sig.)	<i>r</i> (sig.)	<i>r</i> (sig.)
Actief	0,30* (0,03)	0,29* (0,04)	0,04 (0,77)
Aan Taak	0,09 (0,53)	-0,01 (0,95)	0,13 (0,35)
Niet Aan Taak	-0,21 (0,14)	-0,15 (0,29)	-0,16 (0,25)
Storend	-0,25 (0,07)	-0,31* (0,03)	0,03 (0,85)
Positief	0,29* (0,03)	0,25 (0,07)	0,17 (0,22)
Negatief	-0,29* (0,03)	-0,25 (0,07)	-0,17 (0,22)

* $p < 0,05$ (tweezijdig)

Om te toetsen of de verdeling van de typen leergedrag verschilde per mentorscore is een GLM herhaalde metingen-toets uitgevoerd met mentorscore als between subject factor en de vier typen gedragingen als within subjectfactor. Hieruit bleek dat per leerling er significante verschillen waren tussen de vier typen leergedrag, $F(3,153) = 135,42$, $p < 0,001$, partial $\eta^2 = 0,73$. Er was geen interactie effect met mentorscore. Als between subject factor bleek mentorscore evenmin significant. Wel bleek bij de LSD Post Hoc analyses dat leerlingen die de mentorscore *zeer goed* hadden gekregen, in de verdeling van de typen leergedrag significant verschilden van leerlingen die van hun mentor een *slecht-minder goed* score voor hun leergedrag hadden gekregen (mean difference = 0,84; $p = 0,05$). Na weglating van de twee uitbijters veranderden de resultaten niet.

Samengevoegde typen leergedrag. Het verband tussen de mentorscore en *positief leergedrag* was positief, gemiddeld sterk en significant ($r = 0,29$, $p = 0,03$) (Tabel 4). Wanneer de

uitbijter wordt weggelaten, nam de sterkte van de correlatie toe en bleef het verband significant ($r = 0,33, p = 0,02$). De univariate variantie-analyse met mentorscore als fixed factor en *positief leergedrag* als afhankelijke variabele liet zien dat de mentorscore een marginaal significant effect had op het percentage *positief leergedrag*, $F(2,51) = 3,04, p = 0,06$, partial $\eta^2 = 0,11$. LSD Post Hoc analyses lieten zien dat kinderen die de mentorscore *slecht-minder goed* hadden gekregen niet significant verschilden van kinderen met de mentorscore *goed*. Beide groepen kinderen verschilden echter significant in positief leergedrag van de groep kinderen die een mentorscore *zeer goed* kregen (mean difference = 11,06 ; $p = 0,03$ en mean difference = 11,78; $p = 0,03$). Zonder de uitbijter was de uitkomst van de variantie-analyse vrijwel gelijk.

Beschrijving Leergedrag per Type Les en per Type School

Onderscheid naar type les. De frequentie waarmee de vier typen leergedrag tijdens de lessen rekenen en Nederlands werden vertoond, is weergegeven in Tabel 5. *Actief leergedrag* bleek tijdens de lessen rekenen gemiddeld 30,06% van de tijd voor te komen ($SD = 19,81$) en tijdens de lessen Nederlands gemiddeld 23,89% van de tijd ($SD = 13,23$). Tijdens de rekenlessen kwam *aan taak leergedrag* gemiddeld 37,04% van de tijd voor ($SD = 19,84$). Tijdens de lessen Nederlands bedroeg dit percentage 55,99% ($SD = 17,06$). *Niet aan taak* werd tijdens de rekenlessen gemiddeld 25,25% ($SD = 20,02$) van de tijd vertoond en tijdens de lessen Nederlands gemiddeld 16,48% van de tijd ($SD = 13,51$). De gemiddelde frequentie *storend leergedrag* was tijdens de rekenlessen 7,65% ($SD = 8,89$) en tijdens de lessen Nederlands 3,64% ($SD = 5,62$).

Tabel 5

Gemiddelde percentages (standaard deviaties) vier typen leergedrag en positief leergedrag met en zonder onderscheid naar type les (N = 54)

Type leergedrag	Leergedrag	Leergedrag	Leergedrag
	<i>M (SD)</i>	tijdens Rekenen <i>M (SD)</i>	tijdens Nederlands <i>M (SD)</i>
Actief	26,98% (10,36)	30,06% (19,81)	23,89% (13,23)
Aan Taak	46,51% (11,80)	37,04% (19,84)	55,99% (17,06)
Niet Aan Taak	20,86% (12,44)	25,25% (20,02)	16,48% (13,51)
Storend	5,65% (5,48)	7,65% (8,89)	3,64% (5,62)
Positief	73,49% (14,25)	67,10% (22,37)	79,88% (16,17)
Negatief	26,51% (14,25)	32,90% (22,37)	20,12% (16,17)

Bij samenvoeging van de variabelen bleek dat *positief leergedrag* tijdens de rekenlessen gemiddeld 67,10% van de tijd voorkwam ($SD = 22,37$) en tijdens de lessen Nederlands 79,88% van de

tijd ($SD = 16,17$). *Negatief leergedrag* had tijdens de rekenlessen een gemiddelde frequentie van 32,90% ($SD = 22,37$). Tijdens de lessen Nederlands bedroeg deze frequentie gemiddeld 20,12% ($SD = 16,17$). Evenals bij de vier typen leergedrag was de spreiding van de variabelen *positief* en *negatief leergedrag* groter bij de rekenlessen dan bij de lessen Nederlands.

Onderscheid naar type school. In Tabel 6 is beschreven hoe vaak de vier typen leergedrag en de samengevoegde typen leergedrag gemiddeld voorkwamen bij een onderscheid naar school. Onder de leerlingen in het praktijkonderwijs kwam *actief leergedrag* gemiddeld 23,48% van de tijd voor ($SD = 8,28$) en onder de nieuwkomers gemiddeld 29,38% van de tijd ($SD = 11,07$). Het gemiddelde percentage, waarin *aan taak leergedrag* voorkwam, lag voor de leerlingen in het praktijkonderwijs op 46,59% ($SD = 11,45$). Voor de nieuwkomers lag dit percentage op 46,46% ($SD = 12,22$). *Niet aan taak leergedrag* kwam onder de leerlingen in het praktijkonderwijs 23,11% van de tijd voor ($SD = 9,22$) en onder de nieuwkomers 19,32% van de tijd ($SD = 14,18$). Het gemiddelde percentage van de tijd, waarin de leerlingen in het praktijkonderwijs *storend leergedrag* vertoonden, was 6,82% ($SD = 5,34$). Onder de nieuwkomers bedroeg dit percentage 4,84% ($SD = 5,51$). *Positief leergedrag* kwam onder de leerlingen in het praktijkonderwijs gemiddeld 70,08% van de tijd voor ($SD = 12,14$). Onder de nieuwkomers lag dit gemiddelde op 75,83% ($SD = 15,28$). Voor de leerlingen in het praktijkonderwijs was de gemiddelde frequentie van *negatief leergedrag* 29,92% ($SD = 12,14$). Voor de nieuwkomers lag deze op 24,17% ($SD = 15,28$).

Tabel 6

Gemiddelde percentages (standaard deviaties) vier typen leergedrag en positief leergedrag met en zonder onderscheid naar type school

Type leergedrag	Leergedrag	Leergedrag	Leergedrag
	<i>M (SD)</i>	Praktijkonderwijs (<i>n</i> = 22) <i>M (SD)</i>	Nieuwkomersonderwijs (<i>n</i> = 32) <i>M (SD)</i>
Actief	26,98% (10,36)	23,48% (8,28)	29,38% (11,07)
Aan Taak	46,51% (11,80)	46,59% (11,45)	46,46% (12,22)
Niet Aan Taak	20,86% (12,44)	23,11% (9,22)	19,32% (14,18)
Storend	5,65% (5,48)	6,82% (5,34)	4,84% (5,51)
Positief	73,49% (14,25)	70,08% (12,14)	75,83% (15,28)
Negatief	26,51% (14,25)	29,92% (12,14)	24,17% (15,28)

Effect Type Les en Type School

Verband met mentorscore bij onderscheid naar type les. Wanneer er een onderscheid werd gemaakt naar type les en de typen leergedrag afzonderlijk van elkaar werden geanalyseerd, bleek het verband tussen de mentorscore en *actief leergedrag* tijdens de rekenlessen ongeveer even sterk en significant ($r = 0,29$, $p = 0,04$) als *actief leergedrag* op basis van beide lessen ($r = 0,30$, $p = 0,03$) (Tabel 7). Ten opzichte van *storend leergedrag* op basis van beide lessen ($r = -0,25$, $p = 0,07$) werd het verband met *storend leergedrag* tijdens de rekenlessen sterker en significanter ($r = -0,31$, $p = 0,03$). Bij *aan taak* en *niet aan taak leergedrag* tijdens de rekenlessen en de afzonderlijke vier typen leergedrag tijdens de lessen Nederlands leverde het onderscheid naar type les een insignificant verband op. Zonder de uitbijters blijven de significanties ongeveer gelijk.

Door het onderscheid naar type les werd de correlatie tussen de mentorscore en de samengevoegde variabele *positief leergedrag* tijdens de rekenlessen werden zwakker, met een marginale significantie ($r = 0,25$, $p = 0,07$), vergeleken met de correlatie zonder onderscheid naar type les ($r = 0,29$, $p = 0,03$). Wanneer de uitbijter werd weggelaten, werd deze correlatie met *positief leergedrag* tijdens de rekenlessen iets sterker en significant ($r = -0,27$, $p = 0,05$). De correlaties van *positief leergedrag* tijdens de lessen Nederlands waren met en zonder de uitbijter insignificant. Ook de uitkomsten van de variantie-analyses met een onderscheid naar type les waren insignificant. Dit veranderde niet na weglating van de uitbijter.

Tabel 7

Correlaties percentages vier typen leergedrag en positief leergedrag met mentorscore (N = 54)

	Algemeen	Rekenen	Nederlands
Type leergedrag	r (sig.)	r (sig.)	r (sig.)
Actief	0,30* (0,03)	0,29* (0,04)	0,04 (0,77)
Aan Taak	0,09 (0,53)	-0,01 (0,95)	0,13 (0,35)
Niet Aan Taak	-0,21 (0,14)	-0,15 (0,29)	-0,16 (0,25)
Storend	-0,25 (0,07)	-0,31* (0,03)	0,03 (0,85)
Positief	0,29* (0,03)	0,25 (0,07)	0,17 (0,22)
Negatief	-0,29* (0,03)	-0,25 (0,07)	-0,17 (0,22)

* $p < 0,05$ (tweezijdig)

Verband met mentorscore bij onderscheid naar type school. Wanneer de leerlingen werden verdeeld naar school, bleek het verband tussen de mentorscore en *actief leergedrag* van de leerlingen in het praktijkonderwijs zeer zwak en insignificant ($r = 0,17$, $p = 0,45$) (Tabel 8). Onder de leerlingen in het praktijkonderwijs was het verband met *aan taak leergedrag* ($r = -0,41$, $p = 0,06$) en *niet aan taak leergedrag* ($r = -0,38$, $p = 0,08$) marginaal significant. Alleen *storend leergedrag* van de

leerlingen van het praktijkonderwijs leverde een significante correlatie met de mentorscore op ($r = -0,45$, $p = 0,04$). Onder de nieuwkomers waren geen van de verbanden tussen de mentorscore en de vier typen leergedrag significant (Tabel 8). Vooral *storend leergedrag* van de nieuwkomers leverde een opvallend lage correlatiecoëfficiënt op ($r = -0,05$, $p = 0,77$).

Wanneer er een onderscheid naar type school werd gemaakt, bleek dat de correlatie van positief leergedrag van leerlingen in het praktijkonderwijs sterk en significant was ($r = 0,50$, $p = 0,02$). Onder de nieuwkomers was de correlatie van positief leergedrag juist zwak en insignificant ($r = 0,11$, $p = 0,55$). Het weglaten van de uitbijter leidt niet tot een noemenswaardig verschil in significantie.

Tabel 8

Correlaties percentages vier typen leergedrag en positief leergedrag met mentorscore per school

	Algemeen	Praktijkonderwijs ($n = 22$)	Nieuwkomers ($n = 32$)
Type leergedrag	r (sig.)	r (sig.)	r (sig.)
Actief	0,30* (0,03)	0,17 (0,45)	0,30 (0,10)
Aan taak	0,09 (0,53)	0,41 (0,06)	-0,13 (0,47)
Niet aan taak	-0,21 (0,14)	-0,38 (0,08)	-0,08 (0,66)
Storend	-0,25 (0,07)	-0,45* (0,04)	-0,05 (0,77)
Positief	0,29* (0,03)	0,50* (0,02)	0,11 (0,55)
Negatief	-0,29* (0,03)	-0,50* (0,02)	-0,11 (0,55)

* $p < 0,05$ (tweezijdig)

Effect type les en type school. Om te toetsen of de verdeling van de typen leergedrag verschilde per mentorscore en om te onderzoeken of het type les en het type school hierop van invloed waren, is een GLM herhaalde metingen toets uitgevoerd met mentorscore en type school als between subject factoren en de vier typen gedragingen in de twee typen lessen als within subjectfactor. Om te kleine cellen te voorkomen is de mentorscore gehercodeerd naar een mentorscore met een categorie *slecht-minder goed* en een categorie *goed-zeer goed*. Bij een onderscheid naar type les bleek er geen significant verband te bestaan tussen de verdeling van de typen leergedrag en de mentorscore. Het type les had wel een significant effect op het leergedrag van de leerlingen, $F(1,50) = 31,21$, $p < 0,001$, partial $\eta^2 = 0,38$. De verdelingen van de typen leergedrag per leerling per les verschilden wel significant van elkaar, $F(3,48) = 751,63$, $p < 0,001$, partial $\eta^2 = 0,98$. Een significant interactie-effect werd gevonden tussen type les en leergedrag, $F(3,48) = 7,29$, $p < 0,001$, partial $\eta^2 = 0,31$. Wanneer er een onderscheid werd gemaakt naar type les, had het type school geen significant effect op de verdeling van de typen leergedrag. Na weglating van de twee uitbijters veranderden de resultaten niet.

Om te toetsen of de positief of negatief leergedrag verschilde per les is een GLM herhaalde metingen met mentorscore en type school als between subject factoren en type leergedrag (positief en negatief) in de twee type lessen als within subjectfactor gedaan. Om te kleine cellen te voorkomen is mentorscore gehercodeerd waarbij de categorieen slecht en minder goed werden samengevoegd en de categorieen goed en zeer goed. Hieruit bleek dat type les geen effect had op het positieve en negatieve leergedrag van de leerlingen, Er was een significant effect van leergedrag, $F(1,50) = 140,74$, $p = 0,001$, partial $\eta^2 = 0,74$. Tevens was er een significant interactie effect van type leergedrag en type les, $F(1,50) = 9,15$, $p < 0,004$, partial $\eta^2 = 0,16$. Er was geen significant effect van school of van mentorscore. Na weglating van de twee uitbijters veranderden de resultaten niet.

Subgroep Leerlingen met Rapportcijfers

Data-inspectie. De analyse van het verband tussen leergedrag en leerprestaties is verricht onder een subgroep van 31 nieuwkomers, zijnde de leerlingen van wie de rapportcijfers voor rekenen en Nederlands bekend waren. Met betrekking tot deze groep vertoonden de variabelen *storend leergedrag* op basis van de rekenlessen ($z_{\text{scheefheid}} = 3,92$) en *storend leergedrag* op basis van de lessen Nederlands ($z_{\text{scheefheid}} = 3,04$) een bovenmatige scheefheid. Om deze reden is *storend leergedrag* bij de statistische analyses voor deze subgroep logaritmisch getransformeerd.

Aan de hand van boxplots voor de vier typen leergedrag zijn vier uitbijters gesignaleerd binnen de subgroep van leerlingen van wie beide rapportcijfers bekend waren. Hiervan waren er reeds twee als uitbijter binnen de gehele onderzoeksgroep gesignaleerd. Deze jongen en dit meisje scoorden samen met nog een jongen opvallend hoog op de variabele *niet aan taak leergedrag*. De vierde uitbijter had een opvallend hoge score op de variabele *storend leergedrag*. Het meisje dat als uitbijter op *niet aan taak leergedrag* gold, bleek tevens een uitbijter te zijn ten aanzien van de samengevoegde variabelen *positief* en *negatief leergedrag*. Voor alle uitbijters was de afstand tot de mediaan anderhalf tot drie keer de interkwartielafstand.

Na weglating van de uitbijters vertoonde de variabele *niet aan taak leergedrag*, naast de variabele *storend leergedrag*, een zodanige scheefheid ($z_{\text{scheefheid}} = 3,10$) dat ook hiervoor een logaritmische transformatie nodig was.

Beschrijving Leergedrag Subgroep Leerlingen met Rapportcijfers

Typen leergedrag. Met betrekking tot de subgroep van leerlingen met rapportcijfers zijn de vier typen leergedrag en de samengevoegde typen leergedrag, met en zonder onderscheid naar type les, weergegeven in Tabel 9. Ten opzichte van de beschrijving van het leergedrag van de gehele onderzoeksgroep waren er geen opvallende verschillen.

Tabel 9

Gemiddelde percentages vier typen leergedrag per type les voor subgroep rapportcijfers (N = 31)

Type leergedrag	Leergedrag	Leergedrag	Leergedrag
	<i>M (SD)</i>	tijdens Rekenen <i>M (SD)</i>	tijdens Nederlands <i>M (SD)</i>
Actief	29,25 (11,23)	37,10 (21,41)	21,40 (13,19)
Aan Taak	46,67 (12,36)	31,83 (19,36)	61,51 (17,89)
Niet Aan Taak	19,09 (14,35)	24,19 (22,59)	13,98 (13,40)
Storend	5,00 (5,53)	6,88 (9,81)	3,12 (4,47)
Positief	75,91 (15,52)	68,92 (25,11)	82,90 (15,24)
Negatief	24,09 (15,52)	31,08 (25,11)	17,10 (15,24)

Rapportcijfers. In vergelijking tot de rapportcijfers voor rekenen ($M = 6,06$, $SD = 1,56$) lagen de rapportcijfers voor Nederlands gemiddeld hoger terwijl de spreiding van de rapportcijfers voor Nederlands kleiner was ($M = 6,77$, $SD = 0,95$). Ook verschilde de verhouding tussen behaalde voldoende en behaalde onvoldoende. Bij rekenen hadden 13 leerlingen een rapportcijfer lager dan 5,5 behaald terwijl dat bij Nederlands slechts een leerling betrof.

Verband met Rapportcijfers

Vier typen leergedrag. Tussen de vier typen leergedrag enerzijds en de rapportcijfers voor rekenen en voor Nederlands anderzijds is geen significant verband gebleken (Tabellen 10a en 10b). Ten opzichte van zowel het rapportcijfer rekenen als het rapportcijfer Nederlands waren de correlaties niet sterker dan $r = 0,10$. Het weglaten van de vier uitbijters leverde geen verbetering op. Het verband met de rapportcijfers bleef insignificant.

Samengevoegde typen leergedrag. De samenvoeging tot de variabelen *positief* en *negatief leergedrag* leidde tot correlaties van $r = 0,02$ ($p = 0,90$) met betrekking tot het rapportcijfer rekenen en $r = 0,04$ ($p = 0,82$) met betrekking tot het rapportcijfer Nederlands (Tabellen 10a en 10b). Ook na het weglaten van de uitbijter bleven de correlaties zeer zwak en insignificant.

Onderscheid naar type les. Noch de verbanden tussen de vier typen leergedrag op basis van de rekenlessen en het rapportcijfer voor rekenen, noch die tussen de vier typen leergedrag op basis van de lessen Nederlands en het rapportcijfer voor Nederlands, waren significant (Tabellen 10a en 10b). De correlaties met de vier typen leergedrag op basis van de rekenlessen veranderden, met en zonder de uitbijters, nauwelijks door het onderscheid naar type les. Dit lag anders bij het leergedrag op basis van lessen Nederlands. Daar nam de correlatie met *actief leergedrag* in sterkte en significantie toe ($r = 0,25$, $p = 0,18$) ten opzichte van *actief leergedrag* op basis van beide lessen

($r = 0,05$, $p = 0,77$). Het weglaten van de uitbijters had een kleine, positieve invloed op de sterkte en de significantie van dit verband ($r = 0,27$, $p = 0,17$). Ook de correlatie met *storend leergedrag* op basis van lessen Nederlands ($r = 0,24$, $p = 0,20$) nam toe in sterkte en significantie ten opzichte van *storend leergedrag* op basis van beide lessen ($r = -0,04$, $p = 0,85$). Opvallend daarbij was dat dit verband positief was, ook wanneer de uitbijters werden weggelaten.

Tussen de variabelen *positief* en *negatief leergedrag* enerzijds en de rapportcijfers rekenen en Nederlands anderzijds veranderden de correlaties eveneens weinig door het onderscheid naar type les. Zowel voor rekenen ($r = 0,04$, $p = 0,84$), als voor Nederlands ($r = 0,02$, $p = 0,91$) bleven de correlaties zeer zwak en insignificant. Ook hier leidde het weglaten van de uitbijter niet tot een noemenswaardig verschil in significantie.

Tabel 10a. *Correlaties percentages vier typen leergedrag met rapportcijfer rekenen (N = 31)*

	Leergedrag	Leergedrag tijdens Rekenen
Type leergedrag	r (sig.)	r (sig.)
Actief	-0,08 (0,69)	-0,09 (0,64)
Aan Taak	0,10 (0,60)	0,15 (0,43)
Niet Aan Taak	0,06 (0,76)	-0,06 (0,74)
Storend	0,08 (0,69)	-0,04 (0,84)
Positief	0,02 (0,90)	0,04 (0,84)

* $p < 0,05$ (tweezijdig)

Tabel 10b. *Correlaties percentages vier typen leergedrag met rapportcijfer Nederlands (N = 31)*

	Leergedrag	Leergedrag tijdens Nederlands
Type leergedrag	r (sig.)	r (sig.)
Actief	0,05 (0,77)	0,25 (0,18)
Aan Taak	<0,01 (0,98)	-0,16 (0,38)
Niet Aan Taak	0,02 (0,93)	-0,09 (0,65)
Storend	-0,04 (0,85)	0,24 (0,20)
Positief	0,04 (0,82)	0,02 (0,91)

* $p < 0,05$ (tweezijdig)

Verband mentorscores met rapportcijfers. De correlaties tussen de rapportcijfers enerzijds en de mentorscores anderzijds waren sterk en significant. In relatie tot de rapportcijfers voor rekenen was de correlatie $r = .46$ ($p < .01$). Met betrekking tot de rapportcijfers voor Nederlands was de correlatie $r = .66$ ($p < .01$).

Discussie

Dit onderzoek heeft zich gericht op de vier typen *actief*, *aan taak*, *niet aan taak* en *storend leergedrag* en de samengevoegde typen *positief* en *negatief leergedrag*. Ten aanzien van deze typen leergedrag zijn de volgende vragen onderzocht: Hebben de typen geobserveerd leergedrag een voldoende convergente validiteit? Welke typen geobserveerd leergedrag houden verband met leerprestaties? Wat is het effect op de correlaties wanneer er een onderscheid wordt gemaakt naar les of school?

Met dit onderzoek is getracht een bijdrage te leveren aan de ontwikkeling van een observatiesysteem dat zich onderscheidt in praktische hanteerbaarheid en als screeningsinstrument kan worden ingezet voor scholen die hun leerlingen beter willen laten presteren.

Bij de analyses bleken de uitbijters geen structurele invloed op de uitkomsten te hebben. Met uitzondering van de correlatie tussen de mentorscore en *niet aan taak leergedrag* op basis van beide lessen bleven de correlaties met de mentorscore en met de rapportcijfers na weglating van de uitbijters in dezelfde mate significant. Hieronder zal dan ook worden uitgegaan van de onderzoeksgroep met behoud van de uitbijters.

Convergente Validiteit

Op basis van eerder onderzoek (Riley-Tillman et al., 2008, Volpe et al.) was de verwachting dat de frequenties waarin *actief*, *aan taak*, *niet aan taak* en *storend leergedrag* waren geobserveerd, een significante relatie zouden vertonen met de gegeven mentorscores. Een nauw verband met de mentorscores zou indiceren dat het gebruikte observatiesysteem voldoende in staat zou zijn om een representatief beeld van het construct leergedrag in kaart te brengen. Aan deze verwachting werd slechts voor een aantal typen leergedrag voldaan.

Per leerling bleken de verdelingen van de vier typen leergedrag significant van elkaar te verschillen. De resultaten impliceren dat *actief leergedrag* valide indicator van leergedrag is. *Actief leergedrag* leverde een gemiddeld sterke en significante correlatie met de mentorscore op ($r = 0,30$, $p = 0,03$). De correlatie met *storend leergedrag* was marginaal significant, doch zwak ($r = -0,25$, $p = 0,07$). De correlaties met *aan taak* en *niet aan taak leergedrag* waren niet significant en niet eenduidig qua richting. Anders dan verwacht, ging meer *aan taak leergedrag* of minder *niet aan taak leergedrag* niet altijd samen met een hogere mentorscore. Wanneer de verdelingen van de vier typen leergedrag in samenhang met elkaar aan de mentorscore in drie gradaties werden gerelateerd, bleek dat de verdelingen van het leergedrag per mentorscore niet significant van elkaar verschilden.

Positief en *negatief leergedrag* lijken redelijk valide indicatoren van leergedrag te zijn. De samengevoegde variabelen vertoonden een gemiddeld sterk en significant verband met de mentorscore in vier gradaties ($r = 0,29$, $p = 0,03$). Het effect van de mentorscore in drie gradaties op *positief* en *negatief leergedrag* was gemiddeld groot en marginaal significant. De mentorscore *zeer goed* bleek een voldoende onderscheidend vermogen te hebben ten aanzien van de mate van *positief* en *negatief leergedrag*. De mentorscores *slecht-minder goed* en *goed* vertoonden echter onvoldoende samenhang met de samengevoegde typen leergedrag. Voor het screenen van leerlingen met slecht, minder goed en goed leergedrag lijken de mentorscores onderscheidend vermogen te missen.

Bij een onderscheid naar type les bleken de verdelingen van de vier typen leergedrag significant van elkaar te verschillen. Het type les en de vier met elkaar samenhangende typen leergedrag hadden een significant interactie-effect op de mentorscore. Het onderscheid naar type les bleek een positieve invloed te hebben op het verband tussen de mentorscores en de vier afzonderlijke typen leergedrag tijdens de rekenlessen. De correlatie met *actief leergedrag* was marginaal significant ($r = 0,29$, $p = 0,04$) en de correlatie met *storend leergedrag* significant ($r = -0,31$, $p = 0,03$). Daarentegen had het onderscheid naar type les een ongunstige invloed op het verband met de vier typen leergedrag tijdens de lessen Nederlands. Geen van deze typen leergedrag hield significant verband met de mentorscore. Ook de samenhang tussen de mentorscore en de samengevoegde variabelen *positief* en *negatief leergedrag* bleek sterker te zijn ten aanzien van het leergedrag tijdens de rekenlessen ($r = 0,25$, $p = 0,07$) dan ten aanzien van het leergedrag tijdens de lessen Nederlands ($r = 0,17$, $p = 0,22$). De verdeling van de samengevoegde typen leergedrag interacteerde met het type les. Een mogelijke verklaring voor het verschil in effect is dat de leerlingen tijdens de rekenlessen in grotere mate zelfstandig werkten, hetgeen zich kan hebben geuit in een grotere variatie van het leergedrag. Tijdens de rekenlessen leken de leerkrachten minder invloed uit te oefenen op het gedrag van de leerlingen in de klas dan tijdens de lessen Nederlands. Mogelijk benaderde hierdoor het leergedrag tijdens de rekenlessen in sterkere mate het leergedrag, zoals dat werd gezien vanuit de perceptie van de mentoren.

De relaties tussen leergedrag en de mentorscores lijken te verschillen tussen de leerlingen in het praktijkonderwijs en de nieuwkomers. Wanneer de verdelingen van de typen leergedrag per school met elkaar werden vergeleken, bleken de correlaties met de typen *aan taak* ($r = 0,41$, $p = 0,06$), *niet aan taak* ($r = -0,38$, $p = 0,08$) en *storend leergedrag* ($r = -0,45$, $p = 0,04$) significant dan wel marginaal significant te zijn voor de leerlingen in het praktijkonderwijs terwijl voor de nieuwkomers geen van de vier typen leergedrag een significante correlatie met de mentorscore bleek te hebben. Ook de correlatie met *positief* en *negatief leergedrag* was significant ten aanzien van de leerlingen in het praktijkonderwijs ($r = 0,50$, $p = 0,02$) en niet significant ten aanzien van de nieuwkomers ($r = 0,11$, $p = 0,55$). Deze resultaten lijken erop te wijzen dat de twee groepen leerlingen afzonderlijk van elkaar dienen te worden beschouwd. Het leergedrag van nieuwkomers lijkt voor leerkrachten moeilijker in te schatten dan dat van leerlingen in het praktijkonderwijs.

Een mogelijke reden dat de analyses minder bewijs voor de validiteit van de gekozen typen leergedrag bieden dan verwacht, is dat de mentorscore onvoldoende valide en betrouwbaar was. Inherent aan de gebruikte beoordelingsschaal is dat er relatief veel ruimte is voor een subjectief oordeel (Briesch et al, Nock & Kurtz, 2005). Tijdens het huidige onderzoek zijn er enkele aanwijzingen gevonden die impliceren dat de mentorscores zijn beïnvloed door *bias*. Zo waren de mentorbeoordelingen waren veelal positief. De mentorscore *slecht* werd minder vaak toegekend dan men op basis van de percentages geobserveerd negatief leergedrag zou verwachten. Een mentor leek zijn 11 leerlingen zeer voorzichtig te beoordelen door alleen een mentorscore *minder goed* en *goed* toe te kennen. Bovendien was het verband tussen de mentorscores en de rapportcijfers in grote mate sterker en signifikanter dan het verband tussen het geobserveerde leergedrag en de rapportcijfers. Hierdoor rijst de vraag of de mentorscores niet te veel zijn bepaald door de rapportcijfers en te weinig zijn gestoeld op het werkelijk vertoonde leergedrag. Gelet op de genoemde punten lijkt het waarschijnlijk dat de mentorscore minder geschikt was als toetsingskader voor dit onderzoek.

Verband met Leerprestaties

Bestaande literatuur deed vermoeden dat de variabelen *actief*, *niet aan taak* en *storend leergedrag* een sterk verband zouden vertonen met de rapportcijfers (Arnold, 1997; DiPerna, 2006; Greenwood et al, 1984; Morgan et al, Shapiro, 2004a). Gelet op het doel om leerlingen op leergedrag te screenen en zo nodig te verbeteren zodat de leerprestaties kunnen worden geoptimaliseerd, is dit aspect van groot belang voor de bruikbaarheid van het instrument (DiPerna, 2006).

Teleurstellend was dan ook de uitkomst dat geen van de onderzochte typen leergedrag een significant verband vertoonden met de rapportcijfers. Wanneer de typen leergedrag werden samengevoegd tot *positief* en *negatief leergedrag*, werd het verband met de rapportcijfers zwakker. Het onderscheid naar type les leek te leiden tot een verbetering van de correlatie tussen de vier typen leergedrag op basis van de lessen Nederlands en het rapportcijfer Nederlands. Ook dit verband bleef echter insignificant.

Het is niet uitgesloten dat rapportcijfers voor de onderzochte subgroep minder indicatief zijn voor het construct leerprestaties dan in de eerdergenoemde onderzoeken geldt (Arnold, 1997; DiPerna, 2006; Greenwood et al, 1984; Morgan et al, Shapiro, 2004a). Een aanwijzing hiervoor is dat slechts een van de 31 nieuwkomers onvoldoende scoorde op Nederlands terwijl deze groep zich kenmerkte door een lage beheersing van de Nederlands taal. Dit lijkt verklaarbaar wanneer men bedenkt dat het nieuwkomersonderwijs erop is gericht om leerlingen voor te bereiden op een overstap naar het reguliere onderwijs en niet gebonden is aan het reguliere toetsingskader (Inspectie, 2012; Ministerie van Onderwijs, Cultuur en Wetenschap, 2006).

Type Les

Een belangrijke bevinding van dit onderzoek is dat het voor een waarheidsgetrouw beeld van het leergedrag van belang lijkt om een onderscheid te maken naar type les. De 54 geobserveerde leerlingen lieten significant verschillende patronen in leergedrag zien terwijl deze patronen per type les van elkaar verschillen. Zo varieerde het leergedrag meer tijdens de rekenlessen dan tijdens de lessen Nederlands. Ook bleken de leerlingen tijdens de rekenlessen meer negatief leergedrag te vertonen dan tijdens de lessen Nederlands. In dat kader lijken de bevindingen in de literatuur dat specifiekere meten een hogere validiteit oplevert, te worden bevestigd (Cobb, 1972; Greenwood, 2002; DiPerna, 2006). Dit lijkt te impliceren dat leergedrag tijdens rekenlessen een andere betekenis heeft ten opzichte van de wijze waarop een leerling presteert, dan leergedrag tijdens lessen Nederlands. Het verschil in patronen zou kunnen samenhangen met de cognitieve vaardigheden die wel voor rekenen en niet voor Nederlands vereist zijn, en andersom (Cobb, 1972). Een alternatieve verklaring is dat leerkrachten per type les verschillende soorten instructies geven en in een verschillende mate invloed uitoefenen op het gedrag van hun leerlingen.

Conclusie

Alles overziend, wijzen de uitkomsten van dit onderzoek erop dat *actief leergedrag* en *positief* en *negatief leergedrag* een redelijke convergente validiteit bezitten. De onderzochte typen leergedrag blijken geen verband te houden met leerprestaties. Wel is gebleken dat het type les een belangrijke rol speelt ten aanzien van de manier waarop leergedrag en leerprestatie met elkaar samenhangen.

Beperkingen

Afgezien van de mogelijk verschillende relatie tussen leergedrag en mentorscores tussen leerlingen in het praktijkonderwijs en nieuwkomers rijst de vraag in hoeverre deze leerlingen vergelijkbaar zijn met leerlingen in het regulier onderwijs. De leeftijd van de onderzoeksgroep varieerde zeer sterk, hetgeen het leergedrag kan hebben beïnvloed. Bovendien indiceerde de spreiding van de non-verbale IQ-scores van de leerlingen dat de leervermogens relatief grote verschillen vertoonden. Onder meer door het ontbreken van een landelijk toetsingskader verschilde mogelijk ook het soort onderwijs dat de onderzochte groepen leerlingen kregen, ten opzichte van het reguliere onderwijs. Ten aanzien van de nieuwkomers speelden daarnaast nog verschillen in cultuur en taal een mogelijk complicerende rol. De kans bestaat dat de combinatie van de genoemde factoren het leergedrag en het verband met leerprestaties op een zodanige wijze beïnvloedt dat de resultaten van dit onderzoek slechts beperkt gelden voor leerlingen in het regulier onderwijs.

Een andere beperking van het huidige onderzoek is dat de betrouwbaarheid niet is onderzocht. De interbeoordelaarsbetrouwbaarheid werd geacht voldoende te zijn. Tijdens de voorbereiding op de observaties zijn de observatoren getraind tot de observaties de gewenste correlaties vertoonden.

Getracht is om de betrouwbaarheid te vergroten door de observaties *at random* te verspreiden over verschillende lessen. Desondanks is het mogelijk dat de test-hertestbetrouwbaarheid van de data onvoldoende was, met een beperking van de validiteit tot gevolg (Hintze, 2005). Dit gold des te sterker voor de analyses met een onderscheid naar type les, waarbij de correlaties werden berekend op basis van de helft van de totale observatieminuten.

Het aantal data dat voor dit onderzoek voorhanden was, was beperkt. Als gevolg hiervan was het voor de GLM herhaalde metingen en de variantie-analyses noodzakelijk om leerlingen met verschillende mentorscores samen te voegen. Hierdoor gaven de onderzoeksresultaten een minder specifiek beeld. Vanwege de geringe mate waarin *storend leergedrag* is geregistreerd, bleek de systematiek van *momentary time sampling* minder geschikt te zijn voor het meten van dit type leergedrag. In vergelijking tot de andere typen leergedrag is *storend leergedrag* vaak korter van duur en is de kans groter dat *storend leergedrag* te beperkt in beeld wordt gebracht (Shapiro, 2004a). Mogelijk leent *interval recording* zich beter voor de meting van dit soort leergedrag.

Het is mogelijk dat de uitkomsten van dit onderzoek zijn vertroebeld door het ontbreken van een onderscheid naar type instructie. Een signaal hiervoor is dat het leergedrag significant verschilde per type les. Dezelfde typen gedragingen kunnen verschillende betekenissen hebben, afhankelijk van de bedoeling van de les (Shapiro, 2004a). De verwachtingen van de leerkracht zullen tijdens een toets of een spreekbeurt anders zijn dan tijdens interactieve les. Vanuit deze redenering is het mogelijk dat leergedrag beter kan worden begrepen wanneer dit wordt geplaatst in het kader van de instructie die op dat moment geldt.

Aanbevelingen

Het lijkt de moeite waard om verder onderzoek te verrichten naar de validiteit van de indicatoren *actief, aan taak, niet aan taak* en *storend* gedrag en de samengevoegde variabelen *positief* en *negatief leergedrag*. Daarbij wordt aanbevolen om geobserveerd leergedrag te vergelijken met data die zijn gemeten met behulp van een gevalideerd meetinstrument. Gelet op de uitkomsten van dit onderzoek verdient het aanbeveling om rekening te houden met het type les en het type instructie. Om een voldoende representatief beeld van de indicatoren per subcategorie te verkrijgen doet men er verstandig aan om een grotere steekproef te nemen en een groter aantal observatieminuten te verzamelen. Interessant zou het zijn om te bezien of *storend leergedrag* op een meer valide wijze wordt gemeten wanneer dit door middel van *interval recording* geschiedt. Tot slot lijkt het raadzaam om bij een vervolgonderzoek de steekproef te trekken onder leerlingen die onderwijs krijgen volgens de reguliere examen- en diplomastructuur.

Referenties

- Abikoff, H., & Gittelman, R. (1985). Classroom Observation Code: A modification of the Stony Brook Code. *Psychopharmacology Bulletin*, *21*, 901-909.
- Achenbach, T. M. (1986). *The Direct Observation Form of the Child Behavior Checklist* (rev. ed.). Burlington, VT: University of Vermont, Department of Psychiatry.
- Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions*, *13*, 3-16.
- Arnold, D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high risk boys: A preliminary evaluation of patterns and mechanisms. *Journal of applied developmental Psychology*, *18*, 317-330.
- Briesch, A. M., & Chafouleas, S. M. (2009). Review and analysis of literature on self-management interventions to promote appropriate classroom behaviors (1988–2008). *School Psychology Quarterly*, *24*, 106–118.
- Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavior assessment methods to estimate academic engagement: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review*, *39*, 408-421.
- Brown-Chidsey, R. (2005). Introduction to problem-solving assessment. In R. Brown-Chidsey, *Assessment for intervention: A problem solving approach* (pp. 3-9). New York/London: The Guildford Press.
- Burns, M. K., Coddling, R. S., Boice, & C. H., Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *39*, 69-83.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, *64*, 723-733.
- Carroll, J. B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational researcher*, *18*: 26-31.
- Chafouleas, S. M., Volpe, R. J., Gresham, F. M., & Cook, C. R. (2010). School-based behavioral assessment within problem solving models: Current status and future directions. *School Psychology Review*, *39*, 343-349.
- Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009). Foundation for the development and use of direct behavior rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, *34*, 201-213.
- Cobb, J. A. (1972). Relationship of discrete classroom behavior to fourth-grade academic achievement. *Journal of Education Psychology*, *63*, 74-80.
- Cohen, J. (1992). Quantitative methods in psychology: A power primer. *Psychological Bulletin*, *112*, 155-159.
- Commissie Parlementair Onderzoek Onderwijsvernieuwingen (2008). *Eindrapport onderzoek onderwijsvernieuwingen*. 's-Gravenhage: Sdu.

- Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of educational research, 76*, 1–62.
- De Graaf, P. (2012). Peiling kwaliteit van het onderwijs: Ouders hebben ook wel redenen tot klagen. *De Volkskrant*, 23 augustus 2012, p. 11.
- DiPerna, J. C. (2006). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the schools, 43*, 7-17.
- Elbers, E. (2010). Learning and social interaction in culturally diverse classrooms. In K. Littleton, C. P. Vood, & J. Kleine Staarman, *International handbook of psychology in education*. Bingley: Emerald Group Publishing.
- Elzinga, A. (2012). Open brief aan de (nieuwe) Minister van Onderwijs. *J/M voor ouders, 9*, 72-75.
- Gadow, K. D., Sprafkin, J., & Nolan, E. E. (1996). *ADHD School Observation Code*. Stony Brook: Checkmate Plus.
- Goei, S. & Kleijnen, R. (2009). *Omgang met zorgleerlingen met gedragsproblemen*. Zwolle: Christelijke Hogeschool Windesheim.
- Greenwood, C. R. (1991). Longitudinal analysis of time engagement and academic achievement in at-risk and non-risk students. *Exceptional Children, 57*, 521–535.
- Greenwood, C.R. (1996). The case for performance-based models of instruction. *School Psychology Quarterly, 11*, 283–296.
- Greenwood, C.R., Delquadri, J., & Hall, R.V. (1984). Opportunity to respond and student academic performance. In W. Heward, T. Heron, D. Hill, & J. Trap-Porter (Eds.), *Behavior Analysis in Education* (pp. 58–88). Columbus: Merrill.
- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review, 31*, 328–349.
- Hagans, K. S. (2008). A response-to-intervention approach to decreasing early literacy differences in first graders from different socioeconomic backgrounds: Evidence for the intervention validity of DIBELS. *Assessment for Effective Intervention, 34*, 35-42.
- Hart, K. C., Massetti, G. M., Fabiano, G. A., Pariseau, M. E., & Pelham Jr, W. E. (2011). Impact of group size on classroom on-task behavior and work productivity in children with ADHD. *Journal of Emotional and Behavioral Disorders, 19*, 55–64.
- Hattie, J. (2009). *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. Oxford: Taylor & Francis.
- Hintze, J. M., & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review, 33*, 258-270.
- Hintze, J. M. (2005). Psychometrics of direct observation. *School Psychology Review, 34*, 507-519.
- Inspectie van het Onderwijs (2011a). *Opbrengsten – Maak er werk van*. Utrecht: Inspectie van het Onderwijs.

- Inspectie van het Onderwijs (2011b). *De staat van het onderwijs: Onderwijsverslag 2009-2010*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2012). *De staat van het onderwijs: Onderwijsverslag 2010-2011*. Utrecht: Inspectie van het Onderwijs.
- Kamps, D. M., Leonard, B. R., Dugan, E. P., & Boland, B. (1991). The use of ecobehavioral assessment to identify naturally occurring effective procedures in classrooms serving students with autism and other developmental disabilities. *Journal of Behavioral Education, 4*, 367-397.
- Leary, M. (2011). *Introduction to behavioral research methods*. Boston: Pearson.
- Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). *Opbrengstgericht werken over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
- Lepper, M. R., Henderlong Corpus, J., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*, 184-196.
- McIntosh, K., MacKay, L. D., Andreou, Th., Brown, J. A., Mathews, S., Gietz, C., & Bennett, J. L. (2011). Response to intervention in Canada: Definitions, the evidence base, and future directions. *Canadian Journal of School Psychology, 26*, 18.
- McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing: A literature review. *The Journal of Special Education, 41*, 68-84.
- Merrell, K. W. (2010). Better methods, better solutions: Developments in school-based behavioral assessment. *School Psychology Review, 39*, 422-426.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 30*, 741-749.
- Millman, J., Bieger, G. R., Klag, P. A., & Pine, C. K. (1983). Relation between perseverance and rate of learning: A test of Carroll's model of school learning. *American Educational Research Journal, 20*, 425-434.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2006). *Regeling Leerplusarrangement voortgezet onderwijs en nieuwkomers*. Den Haag: Ministerie van OCW.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2011). *Actieplan Beter Presteren: Opbrengstgericht en ambitieus*. Den Haag: Ministerie van OCW.
- Morgan, P. L., Farkas, F., Tufis, P. A., & Sperling, R. A. (2008). Are reading and behavior problems risk factors for each other? *Journal of Learning Disabilities, 41*, 417-436.
- Muhlenbruck, L., Cooper, H., Nye, B., & Lindsay, J. J. (2000). Homework and achievement: Explaining the different strengths of relation at the elementary and secondary school levels. *Social Psychology of Education, 3*, 295-317.

- Muscott, H. S., Mann, E. L., & LeBrun, R. (2008). Implementation of schoolwide positive behavior support on student discipline and academic achievement. *Journal of Positive Behavior Interventions, 10*, 190-205.
- Nock, M. K., & Kurtz, S. M. S. (2005). Direct behavioral observation in school settings: Bringing science to practice. *Cognitive and Behavior Practice, 12*, 359-370.
- OECD (2010). *PISA 2009 Results: What students know and can do – student performance in reading, mathematics and science* (Deel I). Parijs: OECD.
- Onderwijsraad, (2010). *De school en leerlingen met gedragsproblemen*. Den Haag: De Onderwijsraad.
- Oomens, M., Van Aarsen, E., Van Eck, P., & Kieft, M. (2008). *Opbrengstgericht werken door scholen voor primair en voortgezet onderwijs*. Utrecht: Oberon.
- Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes*. New York: The Guildford Press.
- Riley-Tillman, T. C., Chafouleas, S. M., Sassu, K. A., Chanese, J. A. M., & Glazer, A. D. (2008). Examining the agreement of direct behavior ratings and systematic direct observation data for on-task and disruptive behavior. *Journal of Positive Behavior Interventions, 10*, 136-143.
- Saudargas, R. A. (1997). *State-Event Classroom Observation System (SECOS). Observation manual*. Knoxville: University of Tennessee.
- Shapiro, E. S. (2004a). *Academic skills problems: direct assessment and intervention* (4e dr.). New York: The Guildford Press.
- Shapiro, E. S. (2004b). *Academic skills problems workbook* (rev.). New York: The Guilford Press.
- Shapiro, E. S., & Heick, P. F. (2004). School psychologist assessment practices in the evaluation of students referred for social/behavioral/emotional. *Psychology in the Schools, 41*, 551-561.
- Shinn, M. R., Good, R. H., & Parker, C. (1998). Noncategorical special education services with students with severe achievement deficits. In D. J. Reschly, W. D. Tilly, & J. P. Grimes (Eds.), *Functional and noncategorical identification and intervention in special education* (pp. 81-106). Des Moines: State of Iowa Department of Education.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, M., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and Treatment of Children, 31*, 351-380.
- Steege, M. W., & Brown-Chidsey, R. (2005). Functional behavioral assessment: The cornerstone of effective problem solving. In R. Brown-Chidsey, *Assessment for intervention: A problem solving approach* (pp. 131-154). New York/London: The Guildford Press.
- Tilly, W. D. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology V* (pp. 17–36). Bethesda, MD: National Association of School Psychologists.

- Van Vijfeijken, M. M., & Van Schilt-Mol, T. (2012). *Nieuwkomers in het basisonderwijs: Onderzoek naar benodigde competenties van leerkrachten, intern begeleiders en schoolleiders die werken met nieuwkomers*. Eindhoven: IVA beleidsonderzoek en advies.
- Visscher, A., & Ehren, M. (2011). *De eenvoud en complexiteit van Opbrengstgericht Werken: Analyse in opdracht van de Kenniskamer van het Ministerie van Onderwijs, Cultuur en Wetenschap*. Enschede: Universiteit Twente..
- Volpe, R. J., DiPerna, J. C., Hintze, J. M., & Shapiro, E. S. (2005). Observing students in classroom settings: A review of seven coding schemes. *School Psychology Review*, 34, 454-474.
- Walker, H. M., & Severson, H. H. (1990). *Systematic screening for behavior disorders: Users guide and administration manual*. Longmont: Sopris West.
- Wentzel, K.R. (1993). Does being good make the grade? Social behavior and academic competence in middle school. *Journal of Educational Psychology*, 85, 357-364.
- Wood, S. J., Murdock, J. Y., Cronin, M. E., Dawson, N. M., & Kirby, P. C. (1998). Effects of self-monitoring on on-task behaviors of at-risk middle school students. *Journal of Behavioral Education*, 8, 263-279.