

**PREVENTIE VAN GEDRAGSPROBLEMEN VAN LEERLINGEN IN HET
BASISONDERWIJS: EFFECTEN VAN TAAKSPEL EN TAAKSPEL-PLUS-PAD
(EINDRAPPORT)**

M. Haster (Kohnstamm Instituut UvA BV)

H. Blok (Kohnstamm Instituut UvA BV)

P. A. C. van Lier (Faculteit der Psychologie en Pedagogiek, Vrije Universiteit)

R. J. Oostdam (Kenniscentrum Onderwijs en Opvoeding, Hogeschool van Amsterdam)

C. W. van Overveld (Seminarium voor Orthopedagogiek, Hogeschool Utrecht)

30 September 2015

Kohnstamm Instituut UvA BV Postbus 94208 1090 GE Amsterdam T 020 525 1226 W www.kohnstammstituut.uva.nl	Faculteit der Psychologie en Pedagogiek Vrije Universiteit Van der Boechorststraat 1 1081 BT Amsterdam
Kenniscentrum Onderwijs en Opvoeding Hogeschool van Amsterdam Wibautstraat 2-4 1091 GM Amsterdam	Seminarium voor Orthopedagogiek Hogeschool Utrecht Postbus 14007 3508 SB Utrecht

Motto

Preventie van gedragsproblemen bij leerlingen begint bij de aanpak van de leerkracht.

Financiering

Het onderzoek is gefinancierd door het Ministerie van Onderwijs, Cultuur en Wetenschap in het kader van de prijsvraag Onderwijsbewijs, een programma om het evidence-based werken in het onderwijs te versterken (project ODB10013).

Uitvoering

Het onderzoek is ontworpen en uitgevoerd door een consortium, waarin de volgende personen en instellingen hebben geparticipeerd: M. Haster MSc(Kohnstamm Instituut), Dr. H. Blok (Kohnstamm Instituut), Prof. Dr. P. A. C. van Lier (Vrije Universiteit Amsterdam), Prof. Dr. R. J. Oostdam (Hogeschool van Amsterdam, Universiteit van Amsterdam) en Dr. C. W. van Overveld (Seminarium voor Orthopedagogiek). De implementatie van de programma's Taakspel en PAD is ondersteund door medewerkers van de Hogeschool Windesheim Flevoland en het Seminarium voor Orthopedagogiek.

Inhoudsopgave

Samenvatting.....	2
1 Inleiding en onderzoeksvragen.....	4
2 Onderzoeksopzet.....	7
2.1 Populatie en steekproeftrekking.....	7
2.2 De gedragsinterventieprogramma's.....	9
2.3 De implementatie van de experimentele condities.....	12
2.4 Meetinstrumenten en meetschema.....	13
2.5 Gegevensverzameling.....	17
2.6 Bestandsopbouw en gegevensanalyses.....	18
3 Resultaten.....	20
3.1 Resultaten op de beginmeting M1.....	20
3.2 Resultaten op de tussenmeting M2.....	21
3.3 Resultaten op de eindmeting M3.....	22
3.4 Resultaten op de follow-upmeting M4.....	23
3.5 Resultaten in termen van contrasten tussen de condities.....	24
3.6 Samenvatting van de resultaten per domein.....	29
3.7 Aantallen leerlingen met gedragsproblemen.....	32
4 Conclusies.....	34
Referenties.....	37
Bijlage.....	40
Appendix.....	43

Samenvatting

Leerlingen die al in de eerste jaren van de basisschool gedragsproblemen laten zien, lopen een risico op latere negatieve gevolgen. Dit gegeven is een belangrijk argument voor preventieve maatregelen vroeg in het basisonderwijs. Twee interventies die in eerder onderzoek effectief zijn gebleken ten aanzien van de preventie van gedragsproblemen, zijn Taakspel en PAD (Programma Alternatieve Denkstrategieën). Taakspel is een programma voor klasmanagement, een middel voor het creëren van een gestructureerd en veilig klasklimaat. Taakspel leert leerlingen evenwel niet rechtstreeks sociale vaardigheden aan. Het programma PAD doet dit wel en lijkt voort te bouwen op Taakspel door het aanbieden van sociaal-emotionele inzichten en vaardigheden. Door hun verschillende aanpak maar dankzij een overlap in uiteindelijke doelen zijn de twee programma's complementair: Taakspel creëert een context waarin de leerlingen zich veilig voelen, PAD leert kinderen de vaardigheden die nodig zijn om gedragsproblemen te voorkomen. Deze studie onderzocht de preventieve werking van de combinatie van Taakspel en PAD op het aantal leerlingen met, en de mate van gedragsproblemen.

De eerste verwachting was dat Taakspel – in vergelijking met de controleconditie – effectief zou zijn wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen. Verder werd verwacht dat 'Taakspel + PAD' – in vergelijking met de controleconditie – effectief zou zijn wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen. Tot slot werd de hypothese geformuleerd dat 'Taakspel + PAD' – in vergelijking met de conditie Taakspel – effectiever zou zijn wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.

De studie werd opgezet als een *randomized controlled trial*, met drie condities, waarbij 650 leerlingen uit 42 schoolklassen binnen 30 scholen, longitudinaal zijn gevolgd vanaf de start van leerjaar 3 tot en met eind leerjaar 5. Er waren drie condities: een conditie waarbij in leerjaar 3 Taakspel werd geïmplementeerd, evenals in leerjaar 4 (TenT-conditie), een conditie waarbij in leerjaar 3 Taakspel werd geïmplementeerd en een jaar later in leerjaar 4 daarnaast ook nog PAD (TenP-conditie) en tot slot een conditie waarin in leerjaar 3 noch in leerjaar 4 een gedragsinterventieprogramma in gebruik was. Er zijn vier meetrondes geweest: start leerjaar 3 (beginmeting), eind leerjaar 3 (tussenmeting), eind leerjaar 4 (eindmeting) en eind leerjaar 5 (follow-upmeting). Tijdens deze meetrondes werd een batterij aan instrumenten afgenomen met leerkrachten dan wel leerlingen als informant, met één tot drie schalen op een viertal domeinen: gedragsproblemen, leerkracht-leerlingrelatie, sociaal-emotionele ontwikkeling en zorgelijke ontwikkeling.

Multivariate analyses lieten zien dat er verschillen waren tussen de condities per meetronde op het domein gedragsproblemen. Zoals verwacht, nam de ernst van de gedragsproblemen tussen begin- en eindmeting af in zowel de TenT- als de TenP-conditie ten opzichte van de controlegroep. De follow-upmeting liet zien dat dit effect duurzaam was. Ook tussen de twee interventiecondities werd

een verschil gevonden, ten gunste van de TenP-conditie. Daarnaast bleek, conform de verwachting, dat het aantal leerlingen met gedragsproblemen afnam in de interventiecondities ten opzichte van de controlegroep. Er was echter, tegen de verwachting in, geen verschil in reductie van het aantal leerlingen met gedragsproblemen tussen de twee interventiecondities.

Op het domein leerkracht-leerlingrelatie werd op de eindmeting, zowel in de perceptie van de leerkracht als die van de leerling, meer conflict waargenomen bij leerlingen in de controlegroep dan in beide interventiecondities. Dit effect bleek evenwel niet duurzaam blijkens de follow-upmeting na een jaar.

Op het domein sociaal-emotionele ontwikkeling bleek dat leerlingen in de interventiecondities zich na twee jaar gunstiger ontwikkelen dan leerlingen in de controlegroep. Er werd over deze leerlingen zowel minder fysieke en relationele victimisatie als meer emotionele competentie gerapporteerd. Leerlingen in de TenP-conditie waren hierbij in het voordeel ten opzichte van leerlingen in de TenT-conditie. Dit effect bleek duurzaam en terug te vinden in de follow-upmeting.

Op het domein zorgelijke ontwikkeling bleken er op de eindmeting over leerlingen in de TenP-conditie minder zorgen te worden gerapporteerd dan over leerlingen in de beide andere condities.

Deze *randomized controlled trial* levert aanwijzingen op dat de combinatie van Taakspel en PAD complementair werkt en effectief is wat betreft de reductie van de ernst van gedragsproblemen en het aantal leerlingen met gedragsproblemen. Daarnaast lijkt de combinatie van Taakspel en PAD ook een positieve uitwerking te hebben op het sociaal-emotionele domein met een duurzaam effect dat ook een jaar later terug te vinden is. Ook is er een positief effect op de leerkracht-leerlingrelatie. Dit laatste effect is niet duurzaam. Bij de interpretatie moet in acht worden genomen dat de gevonden verschillen veelal te classificeren zijn als klein.

Analyses op basis van *mixed multilevel* modellen, gerapporteerd in een addendum, laten zien dat de verschillen tussen de condities overwegend klein en niet statistisch significant zijn. Dit noopt tot voorzichtigheid ten aanzien van de generalisatie van de uitkomsten uit dit experiment naar de populatie van alle basisschoolleerlingen. Alleen op de eindmeting zijn significante effecten te zien op variabelen binnen het domein gedragsproblemen. Leerlingen in de TenT- en de TenP-conditie vertonen op deze meting minder gedragsproblemen dan leerlingen in de controleconditie. Een jaar later tijdens de follow-upmeting zijn de verschillen er nog steeds, maar ze zijn niet meer statistisch significant.

Kinderen die reeds in de vroege basisschoolperiode gedragsproblemen laten zien, lopen een groot risico op negatieve gevolgen, zoals de continuering of verergering van hun probleemgedrag, delinquentie en contact met politie en justitie, schooluitval, psychiatrische stoornissen en bredere maatschappelijke uitval (Fergusson, Horwood & Ridder, 2005; Moffitt, Caspi, Harrington & Milne, 2002). Dit gegeven vormt een belangrijk argument voor preventieve maatregelen vroeg in het basisonderwijs. De omvang van de groep leerlingen met ernstige gedragsproblemen bedraagt in deze periode ongeveer vijf tot tien procent van alle leerlingen. Leerlingen ontwikkelen deze problemen, omdat ze een tekort hebben aan emotionele en sociale vaardigheden, hetgeen tot uitdrukking komt in problematische relaties met ouders en – tijdens de basisschool – met leerkrachten en leeftijdsgenoten (Moffitt & Caspi, 2001).

In Nederland zijn verscheidene programma's beschikbaar die zich richten op de preventie van gedragsproblemen bij leerlingen in het basisonderwijs. Van deze programma's heeft alleen *Taakspel* (Dolan, Jaylan, Werthamer & Kellam, 1989; Van der Sar & Goudswaard, 2001) de status 'effectief volgens sterke aanwijzingen' (zie de *Databank effectieve jeugdinterventies* van het Nederlands Jeugdinstituut). Weliswaar is naar de effectiviteit van sommige andere programma's onderzoek gedaan, maar vanwege de kwaliteit daarvan moeten deze programma's vooralsnog genoeg nemen met een lagere rating.

In Nederlands onderzoek naar de effecten van Taakspel middels een *randomized controlled trial* zijn reducties gevonden in externaliserende problemen tijdens de interventie en na een jaar follow-up (Van Lier, Muthén, Van der Sar & Crijnen, 2004; Van Lier, Vuijk & Crijnen, 2005) en reducties in internaliserende problemen en middelengebruik tot vijf jaar na de interventie (Van Lier, Huizink & Crijnen, 2009; Vuijk, Van Lier, Crijnen & Huizink, 2007). Uit Amerikaans onderzoek naar het *Good Behavior Game* – de Amerikaanse versie van Taakspel – blijken gedragsproblemen gedurende het programma te verminderen (Ialongo e.a., 1999; Kellam, Rebok, Ialongo & Mayer, 1994). In een review op basis van 22 afzonderlijke studies zijn medium tot sterke effecten vastgesteld op *challenging behaviors* in de klas, zoals fysieke en verbale agressie, ordeverstoring en niet luisteren naar aanwijzingen (Flower, McKenna, Bunuan, Muething & Vega (2014). Na zes jaar, zo blijkt uit follow-up onderzoek, zijn de gunstige effecten nog steeds aantoonbaar (Ialongo, Poduska, Werthamer & Kellam, 2001). Daarnaast is een reductie in het voorkomen van antisociale persoonlijkheidsstoornissen in de jongvolwassen leeftijd gerapporteerd (Petras e.a., 2008), evenals reducties van roken in de adolescentie en jonge volwassenheid (Furr-Holden, Ialongo, Anthony, Petras & Kellam, 2004; Kellam & Anthony, 1998; Kellam e.a., 2008; Storr, Ialongo, Kellam & Anthony, 2002). Wilcox e.a. (2008) rapporteerden een teruggang in suïcidale gedachten en Poduska e.a. (2008) vonden een verminderde consumptie van gezondheidszorg in verband met psychische klachten of middelengebruik in de jongvolwassenheid onder kinderen die het *Good Behavior Game* hadden gehad

tijdens de basisschool. De uitzonderlijke status van Taakspel berust dus op uitgebreid en overtuigend nationaal en internationaal onderzoek.

Taakspel is in essentie een programma voor klasmanagement. Het geeft leerkrachten een krachtig middel om een gestructureerd en veilig klasklimaat te creëren, onder andere door het scheppen van een heldere en positieve structuur, het bevorderen van leerkracht-leerling interacties en het versterken van positieve relaties tussen leerlingen onderling. Taakspel leert leerlingen evenwel geen directe sociale vaardigheden aan. Het preventieprogramma PAD (Programma Alternatieve Denkstrategieën) doet dat wél (Greenberg, Kusché, Calderon & Gustafson, 1987, 2005). Het leert leerlingen emoties te herkennen en effectieve prosociale en emotionele vaardigheden toe te passen. Ook rond PAD is effectiviteitoronderzoek verricht, zowel in Nederland als in de Verenigde Staten. In Nederland is vastgesteld dat agressie van jongens in het basisonderwijs significant afneemt onder invloed van PAD (Louwe & Van Overveld, 2008). De Amerikaanse variant van PAD, het programma *PATHS (Promoting Alternative Thinking Strategies)* is het onderwerp van meerdere onderzoeken geweest. In de Verenigde Staten is een afname gevonden van 32 procent in agressief gedrag, naast een minder grote maar toch significante afname in angsten en depressies (Greenberg e.a., 1998, 2002). In *het Fast Track Project*, waar PATHS de universele interventie was, zijn reducties in gedragsproblemen en ADHD symptomen waargenomen (Conduct Problems Prevention Research Group, 1999a, 1999b). Drie jaar na de interventie was 37 procent van de aanvankelijk hoogrisico kinderen vrij van ernstige gedragsproblemen. Hetzelfde gold voor 27 procent van de hoogrisico kinderen uit de controleconditie (Conduct Problems Prevention Research Group, 2002).

PATHS is in de VS beschreven als een *model program for the prevention of violence* (Greenberg & Kusché, 2002). Taakspel (Good Behavior Game) en PAD (PATHS) worden in de Verenigde Staten ingezet als complementaire programma's: Taakspel creëert een gestructureerde klasomgeving waarin leerlingen zich veilig voelen, PAD bouwt hierop voort door het aanbieden van de sociaal-emotionele inzichten en vaardigheden die nodig zijn ter voorkoming van gedragsproblemen. Men heeft in de V.S. het zogeheten *PATHS to PAX* programma ontwikkeld, een integratie van enkele bewezen effectieve onderdelen uit zowel Taakspel als PAD. Er loopt daar een *randomized controlled trial* betreffende dit geïntegreerde programma (Domitrovich e.a., 2010). Vanwege de genoemde complementariteit bestaan er hoge verwachtingen over de preventieve werking, indien Taakspel en PAD samen zouden worden gevoegd. De verwachting is dat van de gecombineerde programma's een sterkere preventieve werking uitgaat dan van Taakspel alleen.

Onderzoekshypothesen

Het hier beschreven onderzoek heeft tot doel gehad te onderzoeken wat het complementaire effect is in preventieve werking van Taakspel en PAD. Hiervoor wordt de implementatie en dataverzameling volgens het design van een *randomized controlled trial* uitgevoerd. Onderzocht wordt of deze interventies een reductie van het aantal kinderen met gedragsproblemen en van de ernst van deze

problemen tot gevolg heeft. Hierbij zijn vergelijkingen gemaakt tussen drie verschillende condities: Taakspel, Taakspel + PAD en de controleconditie. De hypothesen zijn als volgt geformuleerd:

- a. Taakspel is – in vergelijking met de controleconditie – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.
- b. ‘Taakspel + PAD’ is – in vergelijking met de controleconditie – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.
- c. ‘Taakspel + PAD’ is – in vergelijking met de conditie Taakspel – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.

Het onderzoek is opgezet als een *randomized controlled trial* met drie condities, waarbinnen leerlingen longitudinaal gevolgd zijn vanaf de start van leerjaar 3 tot en met eind leerjaar 5. De condities zijn:

- a. een conditie waarbij in leerjaar 3 Taakspel is geïmplementeerd, evenals een jaar later in leerjaar 4 (de TenT-conditie);
- b. een conditie waarbij in leerjaar 3 Taakspel is geïmplementeerd en een jaar later in leerjaar 4 PAD (de TenP-conditie);
- c. een conditie waarin in leerjaar 3 noch een jaar later in leerjaar 4 een gedragsinterventieprogramma in gebruik is (de controleconditie).

In leerjaar 5 is geen interventie meer gedaan. Scholen hebben derhalve een vrije keuze gehad (zie Tabel 2.1).

Tabel 2.1 De onderzoekscondities

Conditie	Schooljaar		
	2011-2012	2012-2013	2013-2014
TenT	Implementatie Taakspel in leerjaar 3	Continuering Taakspel in leerjaar 4	Vrij naar keuze van de school
TenP	Implementatie Taakspel in leerjaar 3	Continuering Taakspel in leerjaar 4, geïntegreerd met PAD	Vrij naar keuze van de school
Controle	Geen gedrags-interventieprogramma in leerjaar 3	Geen gedrags-interventieprogramma in leerjaar 4	Vrij naar keuze van de school

Er zijn vier meetmomenten geweest: start leerjaar 3 (beginmeting), einde leerjaar 3 (tussenmeting), einde leerjaar 4 (eindmeting), einde leerjaar 5 (follow-up meting).

2.1 Populatie en steekproeftrekking

De populatie wordt gevormd door alle leerlingen in Nederland in het derde en vierde leerjaar. Aan het onderzoek hebben 42 schoolklassen deelgenomen, verdeeld over 30 scholen. De scholen zijn voornamelijk geworven in de provincie Flevoland. Toen het moeilijk bleek om binnen deze provincie voldoende scholen te werven zijn aan de steekproef enkele scholen toegevoegd uit de omgeving van Almelo ($n = 3$ scholen) en uit Amsterdam ($n = 4$ scholen). De toedeling van de scholen aan de drie condities is grotendeels random gedaan. Een uitzondering vormen de Amsterdamse scholen, die toegewezen zijn aan de conditie Taakspel. Voor deze scholen hadden besturen al afspraken gemaakt

met de GGD Amsterdam over de implementatie van Taakspel. De randomisatie heeft plaatsgevonden op schoolniveau, om te voorkomen dat verschillende klassen op dezelfde locatie aan een andere conditie zouden worden toegewezen. Elke locatie van een school werd behandeld als een aparte school. Omdat het moeilijk bleek om scholen te werven voor deelname, is besloten om de interventiecondities wel volledig te vullen (twaalf per conditie), maar werden aan de controleconditie minder scholen toegewezen. Dit resulteerde in het aantal scholen per conditie per meetronde zoals weergegeven in Tabel 2.2. In meetronde 1 namen aan de TenT-conditie uiteindelijk 11 scholen deel, omdat één school al vrij snel na de start van het project aangaf toch prioriteit te geven aan andere projecten binnen de school. In de TenP-conditie zag één school af van deelname tijdens de tweede meetronde vanwege een overdaad aan lopende onderzoeksprojecten. Deze school participeerde wel weer in de derde meetronde. Na het eerste jaar werden van één school in de controlegroep twee locaties samengevoegd tot één. Een aantal scholen besloot voor de vierde meetronde af te zien van deelname, met als redenen schoolverhuizing (n = 1), drukte (n = 2), langdurige ziekte van leerkrachten (n = 2) en een gering belang van deze meetronde voor de school zelf (n = 1).

Tabel 2.2 Aantal deelnemende scholen per conditie en meetmoment

Meetronde/ leerjaar	Schooljaar	Conditie		
		Controle	TenT	TenP
Meetronde 1/ start leerjaar 3	2011-2012	7	11	12
Meetronde 2/ eind leerjaar 3	2011-2012	7	11	11
Meetronde 3/ eind leerjaar 4	2012-2013	5	11	12
Meetronde 4/ eind leerjaar 5	2013-2014	5	8	10

Het aantal deelnemende leerlingen varieerde per meetmoment (zie Tabel 2.3). De gemiddelden en standaarddeviaties zijn berekend op basis van *listwise deletion* bij missende scores, zoals beschreven in paragraaf 2.6. Er zijn geen significante verschillen gevonden op geslacht tussen de leerlingen in de drie condities per meetronde. Wel is er een significant verschil in leeftijd tussen de drie condities op leeftijd in maanden tijdens de eerste en tweede meetronde. Leerlingen in de controleconditie zijn gemiddeld iets jonger dan leerlingen in de interventiecondities. Dat dit verschil in de eerste twee meetrondes wordt gevonden is te verklaren doordat deze twee meetrondes in hetzelfde leerjaar plaatsvinden, en hier dus relatief weinig verloop is. Het gevonden leeftijdsverschil is volgens de normen van Cohen (1988) te classificeren als klein (respectievelijke partiële $\eta^2 = .019$ en $.015$) en daarmee te verwaarlozen in verdere analyses.

Tabel 2.3 Gemiddelden en standaarddeviaties op leeftijd in maanden en sekse per conditie en meetmoment

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)		F	Contrast
	M	SD	M	SD	M	SD		
<i>Begin leerjaar 3 (n = 650 lln.)</i>								
Sekse (1 = jongens)	.46	.50	.53	.50	.53	.50	.89	1=2=3
Leeftijd in maanden	76.91	4.35	78.38	5.07	78.94	5.61	6.33	1<2=3
<i>Eind leerjaar 3 (n = 582 lln.)</i>								
Sekse (1 = jongens)	.45	.50	.51	.50	.53	.50	.95	1=2=3
Leeftijd in maanden	83.18	4.42	84.69	5.11	84.86	5.96	4.32	1<2=3
<i>Eind leerjaar 4 (n = 551 lln.)</i>								
Sekse (1 = jongens)	.47	.50	.50	.50	.53	.50	.49	1=2=3
Leeftijd in maanden	95.64	4.90	96.43	4.99	96.58	5.74	1.17	1=2=3
<i>Eind leerjaar 5 (n = 382 lln.)</i>								
Sekse (1 = jongens)	.57	.50	.50	.50	.53	.50	.63	1=2=3
Leeftijd in maanden	107.78	4.96	109.04	5.26	109.38	6.06	2.30	1=2=3

2.2 De gedragsinterventieprogramma's

Bij de vormgeving van de experimentele condities is gebruik gemaakt van twee gedragsinterventieprogramma's, Taakspel en PAD (Programma Alternatieve Denkstrategieën). We beschrijven ze hier, gebruikmakend van de beschrijvingen die het Nederlands Jeugdinstituut ten behoeve van de *Databank effectieve jeugdinterventies* heeft opgesteld (Ince, 2009; Ince, 2010). Aan het einde van de paragraaf beschrijven we de conditie TenP (de gezamenlijke aanbieding van Taakspel en PAD).

Taakspel

Taakspel is ontwikkeld voor leerlingen in de leerjaren 4 tot en met 8 van het basisonderwijs. Het programma is gericht op a. het vermeerderen van taakgericht gedrag, b. het verminderen van regelovertrekend gedrag bij leerlingen en c. het bevorderen van een positief onderwijsklimaat. Het uiteindelijke doel is beginnend probleemgedrag bij kinderen in een vroeg stadium te verminderen en om te buigen in positiever gedrag. Taakspel is een bewerking van het Amerikaanse *Good Behavior Game*.

Taakspel hanteert een groepsgerichte werkwijze, waarbij leerlingen tijdens de reguliere lessen via een spel leren zich aan klassenregels te houden. De klas wordt na een observatieperiode door de leerkracht in teams verdeeld. Leerlingen die bij elkaar in een team zitten, stimuleren elkaar zich aan de regels te houden, zodat zij als team in aanmerking komen voor een beloning. Taakspel kan aan de

situatie worden aangepast door met meer of minder regels te werken, door het langer of korter te spelen, door de beloning direct of uitgesteld te geven. De leerkracht leert vooral aandacht te besteden aan gewenst gedrag. De leerlingen leren op deze manier dat ze aandacht krijgen als ze gewenst gedrag laten zien.

De implementatie van het Taakspel verloopt in drie fasen. Iedere fase duurt ongeveer drie maanden. In iedere fase ontvangt de leerkracht een training en wordt hem of haar begeleiding geboden in de vorm van klassenconsultaties. De consultaties worden uitgevoerd door een schoolbegeleider die daarvoor gecertificeerd is.

- a. In de invoeringsfase wordt het spel geïntroduceerd bij de leerlingen met behulp van uitgewerkte introductielessen. De leerlingen stellen samen met de leerkracht positieve regels op en verzinnen beloningen die passend zijn voor de klassensituatie. Vervolgens wordt het spel drie keer per week gespeeld. De speeltijd van Taakspel wordt geleidelijk opgebouwd. Het is belangrijk dat het zo gespeeld wordt dat iedereen kan winnen.
- b. In de uitbreidingsfase wordt Taakspel steeds langer en in verschillende onderwijssituaties gespeeld.
- c. In de generalisatiefase wordt vooral aandacht besteed aan het in stand houden van de effecten van Taakspel, ook wanneer het spel niet gespeeld wordt.

Bij het spelen van Taakspel maakt de leerkracht gebruik van een materialendoos met onder andere een docentenhandleiding, pictogrammen en observatieformulieren.

Taakspel wordt gespeeld tijdens gewone lessen, zoals lessen lezen, taal of rekenen. In aanmerking komen lesmomenten van zelfstandig werken en instructie. De klas wordt na een observatieperiode door de leerkracht in teams verdeeld. De leerlingen met veel regelovertrekend gedrag worden verdeeld over de teams, zo ook de leerlingen met matig of geen regelovertrekend gedrag. Voor iedere onderwijssituatie bedenken de leerlingen en leerkrachten positief geformuleerde regels, regels die gewenst gedrag omschrijven. De leerlingen houden zich gedurende de afgesproken tijd aan drie afgesproken regels. De leerkracht geeft daarvoor complimenten. Wanneer een leerling een afgesproken regel overtreedt, pakt de leerkracht een kaart weg. De leerkracht reageert verder niet op het regelovertrekende gedrag. De leerlingen zitten in teams bij elkaar en mogen elkaar helpen zich aan de regels te houden. Wanneer de teams kaarten overhouden, hebben ze gewonnen. Ze krijgen dan een beloning die ze zelf hebben bedacht. De leerlingen kunnen sparen voor een grotere beloning door op een poster kruisjes te zetten of stickers te plakken. Het spel wordt maximaal drie keer per week gespeeld.

Op grond van Nederlands onderzoek naar de effectiviteit van Taakspel heeft het programma van het Nederlands Jeugdinstituut de hoogste status gekregen, namelijk 'effectief volgens sterke aanwijzingen'.

Programma Alternatieve Denkstrategieën (PAD)

PAD is bedoeld voor leerlingen van vier tot en met twaalf jaar in het primair onderwijs of op scholen voor speciaal onderwijs. Het is een klassikaal leerplan, waarmee scholen op systematische en planmatige wijze de socialemotionele ontwikkeling van hun leerlingen kunnen bevorderen. Het leerplan bestaat uit zeven mappen met in totaal 161 lessen voor de leerjaren 1/2, 3, 4, 5, 6, 7 en 8. Het leerplan kent vier hoofdthema's: zelfbeeld, zelfcontrole, emoties en probleemoplossen. PAD is een bewerking van het Amerikaanse programma *PATHS (Promoting Alternative Thinking Strategies)*.

Onder het thema *Zelfbeeld* worden activiteiten ondernomen om elk kind een positief zelfbeeld te laten opbouwen. Bij het thema *Zelfcontrole* leren leerlingen een driestapsprocedure om zichzelf weer onder controle te krijgen in situaties waarin boosheid, verdriet of angst de overhand dreigt te krijgen. In het thema *Emoties* wordt leerlingen geleerd om emoties bij anderen en bij zichzelf te onderkennen. Bij het thema *Probleemoplossen* ligt het accent op het leren oplossen van persoonlijke en sociale moeilijkheden die leerlingen in het dagelijks leven kunnen tegenkomen.

Scholen die PAD willen invoeren, moeten een teamgerichte nascholing volgen bij het Seminarium voor Orthopedagogiek. Pas na het volgen van de cursus kunnen scholen de PAD-materialen aanschaffen. De materialen bestaan uit onder meer het leerplan (geordend in mappen per leerjaar), handpoppen en emotiekaartjes.

Op grond van Nederlands onderzoek naar de effectiviteit van PAD heeft het programma van het Nederlands Jeugdinstituut de op een na hoogste erkenning gekregen, namelijk 'effectief volgens goede aanwijzingen'.

Taakspel en PAD

De gezamenlijke aanbidding van Taakspel en PAD is alleen in het vierde leerjaar gedaan. Om zo dicht mogelijk bij de oorspronkelijke programma's te blijven, hebben we ervoor gekozen de programma's gescheiden aan te bieden. Wat betreft Taakspel is sprake van een invoering in dezelfde drie fasen als hiervoor aangegeven: de invoeringsfase, de uitbreidingsfase en de generalisatiefase. Weliswaar zullen bijna alle leerlingen bekend zijn met Taakspel – zij kregen het programma al in het derde leerjaar aangeboden – maar voor de meeste leerkrachten geldt dat niet. Voor hen is het een nieuw programma, hetgeen een belangrijk argument oplevert om het werken met Taakspel ook in het vierde leerjaar weer zorgvuldig en vanaf het begin op te bouwen. De PAD-lesmap voor het vierde leerjaar bevat 24 uitgewerkte lessen. Omdat de leerlingen in deze conditie geen ervaring hebben met de eerdere PAD-lessen uit het derde leerjaar, is ervoor gekozen om eerst enkele lessen uit het derde leerjaar aan te bieden, namelijk de lessen 3, 4, 7-10, 22 en 23. Vanaf de herfstvakantie komen dan de lessen uit het vierde leerjaar aan bod. Vier daarvan hebben we laten vervallen: de lessen 1, 12, 20 en 24. In totaal beslaat het PAD-deel derhalve geen 24 lessen, maar 28 lessen. De lessen zijn zo gekozen dat alle vier de hoofdthema's (zelfbeeld, zelfcontrole, emoties en probleemoplossen) voldoende aan bod komen. In de instructies aan de leerkrachten is ervan uitgegaan dat de eerste elf PAD-lessen samenvallen met de

invoeringsfase van Taakspel (vanaf de start van het schooljaar tot de kerstvakantie), dat de volgende zeven PAD-lessen samenvallen met de uitbreidingsfase van Taakspel (tussen de kerst- en de voorjaarsvakantie) en dat de laatste tien lessen samenvallen met de generalisatiefase van Taakspel (vanaf de voorjaarsvakantie tot aan de zomervakantie).

2.3 De implementatie van de experimentele condities

In het eerste onderzoeksjaar – schooljaar 2011-2012 – bevonden de leerlingen zich in leerjaar 3. In twee condities is Taakspel ingevoerd. De implementatie is ondersteund door gecertificeerde medewerkers van Windesheim Flevoland en het Seminarium voor Orthopedagogiek¹. Deze hebben, verspreid over het schooljaar, maximaal tien coachingsbezoeken per schoolklas afgelegd. Leraren hebben zelf geregistreerd hoe vaak ze met hun klas Taakspel hebben gespeeld. Het streven is geweest om dat drie maal per week te doen. Uit de logboekgegevens van de leraren blijkt dit streven ten dele gerealiseerd te zijn (Tabel 2.3), óók als we verdisconteren dat de invoering pas na de herfstvakantie gestart kon worden. Vanaf de herfstvakantie tot aan de zomervakantie hebben leraren 35 effectieve schoolweken gehad om Taakspel te spelen, met een theoretisch maximum van 105 speelmomenten (namelijk 35 weken met drie speelmomenten). Het totaal aantal daadwerkelijke speelmomenten bedraagt gemiddeld 62, zo blijkt uit de laatste kolom van Tabel 2.4. Tussen de klassen in beide condities zijn de verschillen klein en niet statistisch significant.

Tabel 2.4 Implementatie Taakspel in leerjaar 3; het gemiddelde aantal weken waarin Taakspel drie keer gespeeld is (of twee keer of één keer) en het totaal aantal keren dat Taakspel in het schooljaar is gespeeld; standaarddeviaties tussen haakjes

Conditie	Aantal klassen	Aantal malen gespeeld per week			Totaal aantal malen gespeeld
		Drie keer	Twee maal	Eén maal	
Taakspel	17	14.5 (5.9)	9.2 (4.8)	0.4 (1.2)	62.3 (16.0)
Taakspel en PAD	12	17.1 (5.6)	5.1 (3.4)	0.3 (0.9)	61.8 (12.7)
Totaal	29	15.6 (5.9)	7.5 (4.7)	0.4 (0.4)	62.1 (14.5)

De implementatie van Taakspel een jaar later in leerjaar 4 laat ongeveer hetzelfde beeld zien (Tabel 2.5). Ook in dit schooljaar is de invoering pas na de herfstvakantie gestart. Dat heeft tot gevolg gehad dat voor de implementatie effectief 32 schoolweken beschikbaar zijn geweest, resulterend in een

¹ De medewerkers van Windesheim Flevoland hebben hun certificaat verworven aan het eind van het eerste onderzoeksjaar, als onderdeel van het door hen gevolgde opleidingstraject vanuit het Centrum voor Educatieve Dienstverlening uit Rotterdam.

theoretisch maximum van 96 spelmomenten (32 schoolweken met elk drie spelmomenten). Het totaal aantal daadwerkelijke spelmomenten volgens opgave van de leerkrachten bedraagt gemiddeld omstreeks 58. Tussen de klassen in de twee condities zijn de verschillen klein en statistisch niet significant.

Tabel 2.5 Implementatie Taakspel in leerjaar 4; het gemiddelde aantal weken waarin Taakspel drie keer gespeeld is (of twee keer of één keer) en het totaal aantal keren dat Taakspel in het schooljaar is gespeeld; standaarddeviaties tussen haakjes

Conditie	Aantal klassen	Aantal malen gespeeld per week			Totaal aantal malen gespeeld
		Drie keer	Twee maal	Eén maal	
Taakspel	17	14.6 (4.3)	7.3 (3.4)	0.6 (0.9)	58.9 (11.2)
Taakspel en PAD	12	13.5 (4.0)	7.7 (3.0)	0.8 (0.8)	56.8 (11.9)
Totaal	29	14.1 (4.2)	7.5 (3.2)	0.7 (0.8)	58.1 (11.3)

Over de implementatie van de PAD-lessen in leerjaar 4 zijn geen gegevens verzameld, hoewel zulks wel de bedoeling is geweest. We kunnen bijgevolg alleen afgaan op algemene indrukken van de schoolbegeleiders die de leraren begeleiding hebben geboden. Deze indrukken wijzen uit dat het merendeel van de leerkrachten nauwgezet de implementatie van de PAD-lessen heeft doorgevoerd: zij hebben conform de opdrachten in de handleiding de aangewezen PAD-lessen gegeven. Van twee leerkrachten is bekend dat zij alleen enkele elementen uit het PAD-leerplan (PAD-kind van de dag, het concept ‘padden’) hebben overgenomen, omdat zij de sociaal-emotionele lessen te tijdsintensief vonden.

2.4 Meetinstrumenten en meetschema

We onderscheiden twee typen meetinstrumenten: meetinstrumenten waarbij leerkrachten vragen over hun leerlingen beantwoorden en meetinstrumenten waarbij leerlingen zelf antwoorden geven (Tabel 2.6).

Tabel 2.6 Meetschema met meetinstrumenten per meetronde en informant

Meetinstrument	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
<i>Leerkracht</i>				
SDQ ^a Totaal	x	x	x	
PBSI ^b Totaal				x
LLRV ^c Conflict	x	x	x	x
SEQ ^d Relatoneel	x	x	x	x
SEQ Fysiek	x	x	x	x
HSCS ^e	x	x	x	x
Gut feeling	x	x	x	
<i>Leerling</i>				
TAK ^f Zinsbegrip	x	x		
TAK Zinsvorming	x	x		
YCATS ^g Warmte	x	x	x	x
YCATS Conflict	x	x	x	x

Noot a: Strengths and Difficulties Questionnaire; b: Problem Behavior At School Interview; c: Leerkracht Leerling Relatie Vragenlijst; d: Social Experiences Questionnaire; e: Head Start Emotional Competence Scale; f: Taaltoets Alle Kinderen; g: Young Children's Appraisal of Teacher Support.

SDQ (Strengths and Difficulties Questionnaire Leerkrachtversie, totaalscore)

De Strengths and Difficulties Questionnaire is een screeningsinstrument waarmee de perceptie van de leerkracht betreffende de psychische problematiek en vaardigheden van kinderen gemeten wordt (Goodman, 1997). We maken gebruik van de totaalscore die is opgebouwd uit vier subschalen (emotionele symptomen, gedragsproblemen, hyperactiviteit en problemen met leeftijdsgenoten). De totaalscore bestaat uit 20 vragen, die beantwoord worden op een 5-puntsschaal (van 'helemaal onjuist' tot 'helemaal juist'). Een voorbeeldvraag is "Deze leerlingen heeft veel zorgen, lijkt vaak over dingen in te zitten". Scores lopen van 20 tot en met 100, waarbij een hogere score ernstiger gedragsproblematiek indiceert. De interne consistentie van de totaalscore bedraagt .70 (Cronbach's alpha), vastgesteld op basis van de eerste meetronde (n = 650 leerlingen).

PBSI (Problem Behavior at Schools Interview)

Het Problem Behavior at Schools Interview is een instrument waarmee de perceptie van de leerkracht betreffende gedragsproblematiek van leerlingen wordt gemeten (Erasmus Medical Center, 2000). We maken gebruik van de totaalscore, die is opgebouwd uit vijf subschalen (angst, ADHD, ODD, depressiviteit en gedragsstoornis). De vragenlijst bestaat uit 30 items die door de leerkracht worden beantwoord op een vijf-puntsschaal ('nooit', 'zelden', 'soms', 'vrij vaak', 'vaak'). Voorbeelditems zijn: "Deze leerling is opstandig" en "Deze leerling huult of is verdrietig op school". De totaalscore loopt van 30 tot en met 150; een hogere waarde duidt op ernstiger probleemgedrag. De interne consistentie

van de totaalscore bedraagt .95 (Cronbach's alpha) vastgesteld op basis van vierde meetronde (n = 382). Dit instrument is alleen gebruikt in de vierde meetronde, namelijk als vervanging van de SDQ.

LLRV- conflict (Leerkracht Leerling Relatie Vragenlijst, conflictschaal)

De LLRV is een instrument waarmee de perceptie van leerkrachten inzake hun relatie met afzonderlijke leerlingen vastgesteld kan worden (Koomen, Verschueren & Pianta, 2011). De vragenlijst resulteert in een indicatie van de algemene kwaliteit van de relatie en van specifieke relatiepatronen. We hebben alleen gebruik gemaakt van de schaal Conflict. Deze bestaat uit vijf vragen (negativiteit, conflicten, onvoorspelbaarheid, boosheid, dwingend gedrag, gebrek aan effectiviteit). Een voorbeeldvraag is "Dit kind en ik lijken voortdurend strijd met elkaar te leveren". Schaalscores lopen van vijf tot en met 25, waarbij een hogere score meer leerkracht-leerling conflict representeert. De homogeniteit van de schaal Conflict bedraagt .95 (Cronbach's alpha), vastgesteld op basis van de eerste meetronde (n = 650 leerlingen).

SEQ (Social Experiences Questionnaire, subschalen fysiek en relationeel)

De Social Experiences Questionnaire (SEQ) is opgesteld om de perceptie van de leerkracht inzake de victimisatie van individuele leerlingen vast te stellen (Crick & Grotpeter, 1996). Wij gebruiken een verkorte versie van deze vragenlijst, bestaande uit zes vragen. De vragenlijst bestaat uit twee subschalen: relationele victimisatie (drie items, bijvoorbeeld 'Kinderen dreigen de vriendschap met dit kind te verbreken als die niet doet wat ze willen') en fysieke victimisatie (drie items, bijvoorbeeld 'Wordt geschopt of geslagen door klasgenoten'). Voor beide schalen lopen de scores van drie tot en met vijftien, waarbij een hogere score meer relationele dan wel fysieke victimisatie weergeeft. De homogeniteit van de subschaal relationele victimisatie bedraagt .92 (Cronbach's alpha, vastgesteld op basis van de eerste meetronde (n = 650 leerlingen)). De subschaal fysieke victimisatie heeft een homogeniteit van .88 (Cronbach's alpha), eveneens vastgesteld op basis van de eerste meetronde (n = 650 leerlingen).

HSCS (Head Start Competence Scale)

De Head Start Competence Scale (HSCS) is een instrument waarmee de perceptie van leerkrachten inzake de sociale en emotionele vaardigheden van leerlingen in kaart gebracht wordt (Domitrovich, Cortes & Greenberg, 2001). Wij maken gebruik van een verkorte versie, bestaande uit acht vragen, te beantwoorden op een vijfpuntsschaal ('helemaal niet waar' tot 'zeker waar'). Voorbeeldvragen zijn: 'Deze leerling neemt gevoelens op de juiste manier waar', 'Deze leerling kan omgaan met woede of frustratie', 'Deze leerling kan accepteren dat dingen anders lopen dan hij/zij wil'. Scores op dit instrument lopen van acht tot en met 40, waarbij een hogere score aangeeft dat leerkrachten deze leerling als meer emotioneel competent inschatten. De homogeniteit van de HSCS is .94 (Cronbach's alpha) op basis van de eerste meetronde (n = 650 leerlingen).

Gut Feeling

De Gut Feeling is een zelf samengestelde lijst met vragen die een gevoelsindicatie van de leerkracht over de mate van problematiek bij individuele leerlingen moet weergeven. De vragenlijst bestaat uit vier vragen waarvan de leerkracht aangeeft in hoeverre hij/zij zich hier zorgen om maakt (gedrag, sociale ontwikkeling, emotionele ontwikkeling en schoolvorderingen). Antwoorden worden gegeven op een vijfpuntsschaal (van ‘helemaal niet waar’ tot ‘zeker waar’). De totaalscore op de Gut Feeling loopt van vier tot en met 20 en indiceert een algemene zorg om de leerling. De homogeniteit van de Gut Feeling is .88 (Cronbach’s alpha) op basis van de eerste meetronde (n= 650 leerlingen). De Gut Feeling is afgenomen in de eerste, tweede en derde meetronde, maar in de vierde meetronde niet meer opgenomen.

Taaltoets Alle Kinderen (TAK, subschaal zinsbegrip en zinsvorming)

De Taaltoets Alle Kinderen (TAK) is een Nederlandse, gestandaardiseerde toets, ontwikkeld om receptieve en productieve taalvaardigheid in kaart te brengen (Verhoeven & Vermeer, 2011). Wij hebben alleen gebruik gemaakt van de subschalen Zinsbegrip en Zinsvorming. De TAK is alleen afgenomen in de eerste twee meetronden, omdat deze toets ontwikkeld is voor leerlingen in de onderbouw van de basisschool. Vanaf de derde meetronde zaten de deelnemende leerlingen in leerjaar 4 en zou de toets mogelijk niet meer onderscheidend werken.

De subschaal Zinsbegrip bestaat uit 42 items, met elk een ander focusgebied (hoeveelheidsaanduidende woorden, ruimtelijke woorden, persoonsaanduidende woorden en voegwoorden). Er worden drie verschillende plaatsjes getoond, gevolgd door een gesproken zin. Leerlingen kiezen het juiste plaatje bij de aangeboden zin. Bijvoorbeeld, het voegwoord ‘hoewel’ in de zin ‘Hoewel het regende, ging hij fietsen’ wordt aangeboden samen met drie plaatjes die respectievelijk een man fietsend in de regen, fietsend in de zon en niet fietsend in de regen weergaven. Antwoorden worden gecodeerd als goed dan wel fout, resulterend in een testscore die loopt van nul tot en met 42, waarbij een hogere score meer taalvaardigheid representeert. De schaal heeft een homogeniteit van .57 (Cronbach’s alpha) op basis van de eerste meetronde (n= 650 leerlingen)

De subschaal Zinsvorming bestaat uit 20 zinnen met een zinspatroon en een functiewoord. Leerlingen horen een gesproken zin en worden gevraagd deze zin te herhalen. Bijvoorbeeld ‘Terwijl de kinderen spelen, wast vader de auto’. De onjuiste zinsreproducties worden genoteerd, en afgetrokken van het totaal aantal items, resulterend in een somscore die loopt van nul tot 40, waarbij een hogere score meer taalvaardigheid weergeeft. De homogeniteit van de subschaal Zinsvorming is .86 (Cronbach’s alpha) op basis van de eerste meetronde (n = 650 leerlingen).

Y-CATS (Young Children's Appraisal of Teacher Support, subschalen conflict en warmte)

De Young Children's Appraisal of Teacher Support is een speelse vragenlijst waarbij de perceptie van leerlingen inzake hun relatie met de groepsleerkracht in kaart gebracht wordt (Mantzicopoulos & Neuharth-Pritchett, 2001). De vragenlijst resulteert in een indicatie van de perceptie van de mate conflict en warmte die het kind ervaart bij de leerkracht, weergegeven in twee subschalen. De subschalen Conflict en Warmte bestaan respectievelijk uit tien en elf stellingen, waarbij leerlingen de keuze hebben tussen 'eens' en 'oneens'. Leerlingen krijgen een stelling te horen en maken hun keuze kenbaar door deze stelling dan wel in de vuilnisbak (oneens) of de kluis (eens) te stoppen. Voorbeeldvragen zijn "Mijn juf/meester vindt mijn familie aardig" en "Mijn juf/meester zegt vaak dat ik iets fout doe". De subschalen resulteren in een somscore, die voor Warmte loopt van elf tot en met 22, waarbij een hogere score meer leerling-leerkracht warmte representeert. Voor Conflict loopt de schaal van tien tot en met 20, met een hogere score als weergave van meer leerling-leerkracht conflict. Dit alles resulteert in een homogeniteit van .67 voor conflict (n= 650 leerlingen) en .57 voor warmte (Cronbach's alpha) op basis van de eerste meetronde (n = 650 leerlingen).

Zie tabel 2.7 voor het totaaloverzicht van de gehanteerde afhankelijke variabelen.

Tabel 2.7 Range en richting van uitkomstmaten

Variabelen	Min.	Max.	Hogere score
SDQ Totaal	20	100	Ernstiger gedragsproblematiek
PBSI Totaal	30	150	Ernstiger gedragsproblematiek
LLRV Conflict	5	25	Meer leerkracht-leerling conflict
SEQ Relationeel	3	15	Meer relationele victimisatie
SEQ Fysiek	3	15	Meer fysieke victimisatie
HSCS	8	40	Meer emotioneel competent
Gut Feeling	4	20	Meer zorgen
TAK Zinsbegrip	0	42	Meer taalvaardig
TAK Zinsvorming	0	40	Meer taalvaardig
Y-CATS Warmte	11	22	Meer leerling-leerkracht warmte
Y-CATS Conflict	10	20	Meer leerling-leerkracht conflict

2.5 Gegevensverzameling

Voorafgaand aan de meting ontvingen alle leerkrachten een informatiebrief, met bijgevoegde informatiebrochures voor alle ouders. Hierin werd beschreven waar het onderzoek toe diende, hoe het verloop van de testdag zou zijn en werd er de mogelijkheid gegeven om bezwaar voor deelname aan te tekenen. Leerkrachten verspreidden deze brieven over de betrokken ouders.

Alle vier de meetrondes zijn volgens dezelfde procedure verlopen. Voorafgaand aan de meting werden testleiders geworven, studenten sociale wetenschappen (pedagogiek, psychologie,

onderwijskunde). Deze testleiders werden getraind door de projectleiders en werkten volgens een uitgebreide testhandleiding. Elke testdag vond plaats onder supervisie van een dagcoördinator.

Bij de tweede meetronde werd actief toestemming gevraagd aan ouders voor deelname van de leerling aan het project. Scholen werden benaderd of ze hiermee akkoord gingen. Hierbij ontvingen ouders een brief met uitleg over het onderzoek, waarbij gevraagd werd om in te stemmen dan wel bezwaar aan te tekenen tegen deelname. Leerkrachten verzamelden per klas deze toestemmingsformulieren en zonden deze naar de projectleiders. Scholen werd gevraagd telefoonnummers van ouders aan te leveren. Wanneer scholen hiermee akkoord gingen, werden ouders van leerlingen waarvan de toestemmingsbrieven niet terugkwamen gebeld om alsnog een beslissing kenbaar te maken.

De testdag zelf verliep volgens een gestandaardiseerde procedure met een op de school afgestemde dagplanning. Hierin werd per tijdspad aangegeven welke leerlingen op dat moment getest zouden worden. Gedurende de schooldag werden alle leerlingen individueel getest. Wanneer een leerling ziek of om een andere reden afwezig was, werd deze leerling voor die meetronde niet getest. Wanneer er vanwege de schoolgrootte een extra testdag op deze school plaatsvond, werden afwezige kinderen zo mogelijk op de tweede testdag alsnog getest. Aan het eind van de testdag kregen alle leerlingen een klein cadeautje als dank voor hun deelname.

Leerkrachtinterviews vonden in de middagpauze en na schooltijd plaats. Wanneer dit door omstandigheden op de school niet werkbaar was, werd leerkrachten toegestaan de vragenlijsten in eigen tijd in te vullen en binnen een week op te sturen naar de projectleider. Als dank voor hun deelname ontvingen leerkrachten een VVV-bon en een doos chocolaatjes.

Na elke meetronde ontving de school een terugrapportage over die meetronde. Hierin werd kort beschreven welke instrumenten er waren afgenomen, en werd per school voor een selectie van de meetinstrumenten kort beschreven hoe de resultaten te beschrijven waren ten opzichte van andere scholen in de regio en in de steekproef.

2.6 Bestandsopbouw en gegevensanalyses

Voorafgaand aan de analyses zijn via *listwise deletion* alle deelnemende leerlingen geselecteerd van wie per meetronde alle uitkomstmaten (inclusief geslacht en leeftijd) aanwezig waren. Dit resulteerde in een bestand van 650 leerlingen in de eerste meetronde, van 582 leerlingen in de tweede meetronde, van 551 leerlingen in de derde meetronde en van 382 leerlingen in de vierde meetronde. Een overzicht van het aantal leerlingen per meetronde per conditie is weergegeven in Tabel 2.8.

Allereerst zijn er analyses of variance (ANOVA's) uitgevoerd op de demografische gegevens, geslacht en leeftijd in maanden. Vervolgens zijn er per meetronde multivariate ANOVA's (MANOVA's) uitgevoerd op alle uitkomstmaten met conditie als onafhankelijke variabele. De

effectgroottes (geschat als partiële η^2) van gevonden significante verschillen zijn geëclassificeerd volgens de normen van Cohen (1988), waarbij partiële η^2 van .02 als kleine, .13 als medium en .26 als grote effecten zijn geëclassificeerd. In de appendix rapporteren we ook nog uitkomsten op basis van meerniveau-analyses, waarbij rekening wordt gehouden met de hiërarchische structuur van de data.

Tabel 2.8 Aantal leerlingen met complete data per conditie per meetronde

Conditie	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle	124	121	102	85
TenT	296	245	238	158
TenP	230	216	211	139
Totaal	650	582	551	382

3.1 Resultaten op de beginmeting M1

Via MANOVA werden de verschillen tussen de condities op alle uitkomstmaten onderzocht. Er werd een significant multivariaat effect gevonden van conditie op de tien uitkomstmaten, $F(20, 1278) = 4.57, p < .001, \text{partiële } \eta^2 = .07$. Univariate toetsen, gerapporteerd in Tabel 3.1, lieten zien dat het effect van conditie alleen significant was voor SEQ Relationeel ($F(2, 647) = 11.39, p < .001, \text{partiële } \eta^2 = .03$), SEQ Fysiek ($F(2, 647) = 15.93, p < .001, \text{partiële } \eta^2 = .05$) en de HSCS ($F(2, 647) = 4.96, p < .05, \text{partiële } \eta^2 = .01$). Uit contrastanalyses bleek dat leerlingen in de controleconditie, vergeleken met de interventiecondities, lager scoorden op zowel fysieke als relationele victimisatie. Dit betekent dat door leerkrachten over deze leerlingen minder relationele en fysieke victimisatie werd gerapporteerd dan over de leerlingen de interventiecondities. Deze effectgroottes kunnen worden geclassificeerd als klein tot medium, volgens de normen van Cohen (1988). Leerlingen in de controleconditie scoorden hoger op de HSCS dan de leerlingen in de interventiecondities. Dit houdt in dat leerkrachten deze leerlingen als meer emotioneel competent inschatten dan de leerlingen in de interventiecondities. Dit effect is te classificeren als klein (Cohen, 1988).

Tabel 3.1 Gemiddelden, standaarddeviaties en verschiltoetsen op een aantal variabelen per conditie op meetmoment 1 ($n=650$)

Variabelen	Conditie						Verskil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SD	M	SD	M	SD	F	Contrast
SDQ Totaal	44.10	10.18	44.97	9.03	45.95	7.52	1.90	1=2=3
LLRV Conflict	8.62	4.58	8.90	3.88	9.07	3.17	.56	1=2=3
SEQ Relationeel	4.87	2.03	5.82	1.86	5.50	1.79	11.39	1<2=3
SEQ Fysiek	4.54	1.63	5.52	1.63	5.25	1.60	15.93	1<2=3
HSCS	30.85	6.10	29.37	4.55	29.41	4.28	4.69	1>2=3
Gut Feeling	8.59	3.64	8.61	3.47	8.56	3.21	.02	1=2=3
TAK Zinsbegrip	36.65	2.81	36.74	2.56	36.73	2.65	.05	1=2=3
TAK Zinsvorming	30.39	6.64	30.97	5.73	30.13	6.14	1.29	1=2=3
Y-CATS Warmte	20.18	1.77	20.32	1.51	20.43	1.66	1.04	1=2=3
Y-CATS Conflict	12.92	2.31	12.54	2.06	12.44	1.94	2.22	1=2=3

3.2 Resultaten op de tussenmeting M2

Met een MANOVA werd het effect van conditie op de tien verschillende uitkomstmaten tijdens de tweede meetronde onderzocht. Er werd een significant multivariaat effect gevonden van conditie, $F(20, 1142) = 3.11, p < .001, \text{partiële } \eta^2 = .05$. De univariate toetsing, gerapporteerd in Tabel 3.2, liet zien dat het effect van conditie significant was voor SDQ Totaal ($F(2, 579) = 11.60, p < .001, \text{partiële } \eta^2 = .04$), SEQ Relationeel ($F(2, 579) = 5.34, p < .01, \text{partiële } \eta^2 = .02$) en SEQ fysiek ($F(2, 579) = 12.58, p < .001, \text{partiële } \eta^2 = .04$). Uit de contrastanalyses bleek dat over leerlingen in de TenP-conditie, vergeleken met leerlingen in de controle en TenT-conditie, ernstiger gedragsproblemen werden gerapporteerd door de leerkracht. Dit effect is te classificeren als een klein tot medium effect, volgens de normen van Cohen (1988). Leerlingen in de controleconditie, vergeleken met de interventiecondities, scoorden lager op relationele en fysieke victimisatie. Dit betekent dat door leerkrachten minder relationele en fysieke victimisatie werd gerapporteerd over deze leerlingen dan over de leerlingen in de interventiecondities. Op SEQ Fysiek bleek daarnaast ook dat de leerlingen in de TenT-conditie lager scoorden dan leerlingen in de TenP-conditie, wat betekent dat over leerlingen in de TenT-conditie minder fysieke victimisatie werd gerapporteerd dan over leerlingen in de TenP-conditie. De effectgroottes van deze effecten kunnen worden geclassificeerd als klein tot medium (Cohen, 1988).

Tabel 3.2 Gemiddelden, standaarddeviaties en verschiltoetsen op een aantal variabelen per conditie op meetmoment 2 ($n = 582$)

Variabelen	Condities						Verskil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SD	M	SD	M	SD	F	Contrast
SDQ Totaal	41.53	8.64	41.42	8.99	44.92	7.52	11.60	1=2<3
LLRV Conflict	8.17	3.64	8.37	3.52	8.92	3.13	2.38	1=2=3
SEQ Relationeel	5.00	1.78	5.44	1.79	5.64	1.60	5.34	1<2=3
SEQ Fysiek	4.71	1.54	5.07	1.56	5.54	1.43	12.58	1<2<3
HSCS	30.27	3.59	29.90	4.44	30.00	3.68	.35	1=2=3
Gut Feeling	8.17	2.76	8.30	3.22	8.63	3.48	.98	1=2=3
TAK Zinsbegrip	37.74	2.23	37.91	2.22	37.93	2.20	.30	1=2=3
TAK Zinsvorming	31.48	5.94	31.94	5.46	31.70	5.37	.30	1=2=3
Y-CATS Warmte	20.07	1.96	20.37	1.52	20.19	1.89	1.34	1=2=3
Y-CATS Conflict	12.81	2.21	12.43	1.99	12.67	2.23	1.49	1=2=3

3.3 Resultaten op de eindmeting M3

Een MANOVA werd uitgevoerd om de verschillen tussen de condities op alle uitkomstmaten tijdens de derde meetronde te toetsen. Er werd een significant multivariaat effect gevonden van conditie op de acht uitkomstmaten, $F(16, 1084) = 4.57, p < .001, \text{partiële } \eta^2 = .06$. Univariate toetsen lieten zien dat het effect van conditie, behalve voor Y-CATS Warmte, voor alle uitkomstmaten significant was. Gemiddelden, standaarddeviaties en F-waarden voor deze meetronde zijn te vinden in Tabel 3.3. Uit de contrastanalyses bleek over leerlingen in de controleconditie ernstiger gedragsproblematiek te worden gerapporteerd dan over leerlingen in de TenT-conditie. Leerlingen in de TenT-conditie scoorden weer hoger dan leerlingen in de TenP-conditie ($F(2, 548) = 12.02, p < .001, \text{partiële } \eta^2 = .04$). Dit effect kan worden geclassificeerd als klein tot medium (Cohen, 1988). Verder lieten de contrastanalyses zien dat leerlingen in de TenP-conditie hoger scoren op de HSCS dan leerlingen in de controle en TenT-conditie, $F(2, 548) = 3.22, p < .05, \text{partiële } \eta^2 = .01$. Dit betekent dat leerkrachten de leerlingen in de TenP-conditie als meer emotioneel competent inschatten dan leerlingen in de andere twee condities. Dit effect is te classificeren als klein (Cohen, 1988). Op de Gut Feeling blijken leerlingen in de TenP-conditie lager te scoren dan leerlingen in de controle en TenT-conditie ($F(2, 548) = 4.30, p < .05, \text{partiële } \eta^2 = .02$). Dit houdt in dat leerkrachten over leerlingen in de TenP-conditie gemiddeld minder zorgen rapporteren dan over leerlingen in de controle en TenT-conditie. Dit effect kan worden geclassificeerd als klein volgens de normen van Cohen (1988). Tot slot bleek uit de contrastanalyses dat leerlingen in de controleconditie lager hoger scoorden dan leerlingen in de interventiecondities op de uitkomstmaten LLRV Conflict ($F(2, 548) = 9.88, p < .001, \text{partiële } \eta^2 = .04$), SEQ Relationeel ($F(2, 548) = 4.16, p < .05, \text{partiële } \eta^2 = .02$), SEQ Fysiek ($F(2, 548) = 8.00, p < .001, \text{partiële } \eta^2 = .03$) en Y-CATS Conflict ($F(2, 548) = 4.40, p < .05, \text{partiële } \eta^2 = .02$). Dit betekent dat over leerlingen in de controleconditie meer leerkracht-leerling conflict, relationele victimisatie, fysieke victimisatie wordt gerapporteerd dan over leerlingen in de interventiecondities. Daarnaast houdt dit in dat leerlingen in de controleconditie zelf meer leerling-leerkrachtconflict rapporteren dan leerlingen in de interventiecondities. De effectgroottes zijn te classificeren als klein tot medium (Cohen, 1988).

Tabel 3.3 Gemiddelden, standaarddeviaties en verschiltoetsen op een aantal variabelen per conditie op meetmoment 3 ($n = 551$)

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)		F	Contrast
	M	SD	M	SD	M	SD		
SDQ Totaal	46.71	13.12	42.27	11.75	39.66	11.68	12.02	1>2>3
LLRV Conflict	10.50	4.42	8.42	4.13	8.57	3.99	9.87	1>2=3
SEQ Relationeel	6.34	2.58	5.65	2.36	5.55	2.28	4.16	1>2=3
SEQ Fysiek	5.96	2.17	5.16	1.96	5.06	1.82	8.00	1>2=3

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SD	M	SD	M	SD	F	Contrast
HSCS	29.15	6.95	29.33	5.81	30.61	5.83	3.22	1=2<3
Gut Feeling	9.56	3.89	8.78	3.92	8.22	3.74	4.30	1=2>3
Y-CATS Warmte	20.11	1.59	20.45	1.52	20.52	1.48	2.73	1=2=3
Y-CATS Conflict	12.80	2.18	12.15	1.94	12.20	1.85	4.40	1>2=3

3.4 Resultaten op de follow-upmeting M4

Ook voor de vierde meetronde werd een MANOVA uitgevoerd om te onderzoeken wat het effect van conditie is op de verzameling van uitkomstmaten. Er werd een significant multivariaat effect gevonden van conditie op de zeven uitkomstmaten, $F(14, 748) = 4.01, p < .001, \text{partiële } \eta^2 = .07$. Univariate toetsen, gerapporteerd in Tabel 3.4, lieten zien dat het effect van conditie significant was voor PBSI Totaal ($F(2, 379) = 10.74, p < .001, \text{partiële } \eta^2 = .05$), SEQ Relationeel ($F(2, 379) = 10.47, p < .001, \text{partiële } \eta^2 = .05$), SEQ Fysiek ($F(2, 379) = 9.47, p < .001, \text{partiële } \eta^2 = .05$) en de HSCS ($F(2, 379) = 4.86, p < .01, \text{partiële } \eta^2 = .03$). Uit contrastanalyses bleek dat leerkrachten over leerlingen in de controleconditie ernstiger gedragsproblemen rapporteerden dan over leerlingen in de TenT-conditie. Over leerlingen in de TenT-conditie werd minder ernstiger gedragsproblematiek gerapporteerd dan over leerlingen in de TenP-conditie. Dit effect is te classificeren als klein tot medium (Cohen, 1988). Verder bleek er over leerlingen in de TenP-conditie significant minder relationele en fysieke victimisatie te worden gerapporteerd dan over leerlingen in de controle en TenT-conditie. Deze effecten kunnen eveneens worden geclassificeerd als klein tot medium volgens de normen van Cohen (1988). Tot slot scoorden leerlingen in de TenP-conditie significant hoger op de HSCS dan leerlingen in de controle en TenT-conditie. Dit houdt in dat leerkrachten leerlingen in de TenP-conditie als meer emotioneel competent inschatten dan leerlingen in de controle en TenT-conditie. Dit effect is te classificeren als klein tot medium (Cohen, 1988).

Tabel 3.4 Gemiddelden, standaarddeviaties en verschiltoetsen op een aantal variabelen per conditie op meetmoment 4 ($n = 382$)

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SD	M	SD	M	SD	F	Contrast
PBSI Totaal	55.80	16.94	50.18	17.34	45.33	15.04	10.74	1>2>3
LLRV Conflict	7.65	4.22	7.06	3.25	6.89	3.24	1.31	1=2=3
SEQ Relationeel	5.25	2.46	4.74	2.23	3.95	1.86	10.47	1=2>3
SEQ Fysiek	4.12	1.72	4.28	1.79	3.50	1.21	9.47	1=2>3
HSCS	29.54	4.95	30.85	5.54	31.78	5.02	4.86	1=2<3

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SD	M	SD	M	SD	F	Contrast
Y-CATS Warmte	19.99	1.90	20.23	2.08	20.29	1.98	.62	1=2=3
Y-CATS Conflict	12.56	2.18	12.40	2.23	12.46	2.27	.14	1=2=3

3.5 Resultaten in termen van contrasten tussen de condities

In deze paragraaf richten we ons op de verschillen tussen de condities die er mogelijkwijs in de loop van de meetmomenten zijn ontstaan. Deze beschouwingwijze is relevant gezien de verwachtingen ten aanzien van de uitkomsten die aan het eind van hoofdstuk 1 zijn geformuleerd. Deze verwachtingen hebben betrekking op nader aangeduide verschillen in gemiddelden die naar verwachting tussen condities ontstaan. We behandelen steeds drie contrasten, respectievelijk tussen de controleconditie en de TenT-conditie, tussen de controleconditie en de TenP-conditie, en als laatste tussen de TenT-conditie en de TenP-conditie². De contrasten zijn geschat op basis van effectgroottes, het gestandaardiseerde verschil in gemiddelden. Standardisatie is steeds gedaan op basis van de standaarddeviatie in de controlegroep. Een positieve effectgrootte geeft aan dat het gemiddelde in de als eerste aangeduide conditie hoger is dan in de als tweede aangeduide conditie. Voor een negatieve effectgrootte geldt het omgekeerde. Effectgroottes worden alleen vermeld als deze statistisch significant van nul verschillen (bepaald op basis van de uitkomsten van de MANOVA's, zoals vermeld in de paragrafen 3.1 tot en met 3.4). Wat betreft de interpretatie van de effectgroottes baseren we ons op de suggesties van Cohen (1988). Bijgevolg spreken we bij een effectgrootte tot 0.20 van een klein effect, bij een effectgrootte van 0.50 van een medium effect, bij een effectgrootte van 0.80 van een groot effect.

Als eerste bespreken we contrasten die op de SDQ- of PBSI-totaalschaal aan het licht getreden zijn (Tabel 3.5). Leerlingen in de TenT-conditie hebben op de eindmeting en op de follow-upmeting minder gedragsproblemen dan de leerlingen in de controlegroep, blijkend uit de positieve effectgroottes. De omvang van de verschillen is tussen klein en medium. Evenwel op de tussenmeting is er geen verschil ontstaan. Leerlingen in de TenP-conditie hebben in vergelijking met de leerlingen in de controleconditie op de tussenmeting meer gedragsproblemen gekregen (een klein tot medium effect), maar op de beide laatste metingen daarentegen minder gedragsproblemen (een medium tot groot effect). Vergelijken we tot slot de beide interventiecondities, dan is vast te stellen dat de

² Het derde contrast is in principe redundant ten opzichten van de eerste twee contrasten (het derde contrast is namelijk het verschil tussen de beide eerdere contrasten). Maar omdat wij alleen significante effectgroottes bespreken, kan de lezer toch niet in alle gevallen zelf de grootte van het derde contrast afleiden. Om deze reden voegen wij het derde contrast aan de tabellen toe.

leerlingen in de TenP-conditie op de tussenmeting – op dat moment was de PAD-component nog niet aangeboden! – meer gedragsproblemen laten zien (zoals af te leiden valt uit een negatieve effectgrootte), maar op de latere metingen juist minder gedragsproblemen laten zien dan de leerlingen in de TenT-conditie.

Tabel 3.5 *SDQ Totaal; effectgroottes tussen de condities*

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	n.s.	n.s.	0.34	0.33
Controle (1) vs TenP (3)	n.s.	- 0.39	0.54	0.62
TenT (2) vs TenP (3)	n.s.	- 0.41	0.20	0.29

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder. De effectgroottes op M4 (Follow-up) hebben betrekking op de PBSI totaal, een instrument dat grotendeels dezelfde betekenis heeft als de SDQ totaal.

De contrasten die op de LLRV Conflictschaal zijn gevonden zijn weergegeven in Tabel 3.6. Op de beginmeting, tussenmeting en follow-upmeting zijn geen verschillen gevonden in de mate van leerkracht-leerling conflict naar rapportage van de leerkracht. Leerlingen in de TenT-conditie hebben op de eindmeting minder leerkracht-leerlingconflict dan leerlingen in de controlegroep, blijkend uit de positieve effectgrootte. Leerlingen in de TenP-conditie hebben in vergelijking met de leerlingen in de controleconditie op de eindmeting eveneens minder leerkracht-leerlingconflict. De omvang van beide verschillen is klein tot medium.

Tabel 3.6 *LLRV Conflict; effectgroottes tussen de condities*

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	n.s.	n.s.	.47	n.s.
Controle (1) vs TenP (3)	n.s.	n.s.	.44	n.s.
TenT (2) vs TenP (3)	n.s.	n.s.	n.s.	n.s.

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

De contrasten die aan het licht zijn getreden op de SEQ Relationeel zijn opgenomen in Tabel 3.7. Op alle meetmomenten worden verschillen gevonden tussen de controleconditie en een of meer interventiecondities. Op de beginmeting wordt over leerlingen in de TenT-conditie en TenP-conditie meer relationele victimisatie gerapporteerd dan over leerlingen in de controlegroep. Dit is eveneens

het geval op de tussenmeting. Deze effecten zijn te classificeren als klein tot medium. Op de eindmeting draaien de verschillen om en worden over leerlingen in de TenT-conditie en TenP-conditie minder relationele victimisatie gerapporteerd dan over leerlingen in de controlegroep. Deze verschillen zijn eveneens te classificeren als klein tot medium. Tijdens de eerste drie meetmomenten (begin-, tussen- en eindmeting) zijn geen verschillen tussen de leerlingen in de interventiecondities. Op de follow-upmeting wordt er echter wel een verschil gevonden, waarbij over de leerlingen in de TenP-conditie minder relationele victimisatie wordt gerapporteerd dan over leerlingen in de TenT-conditie. Tot slot wordt er meer relationele victimisatie gerapporteerd over leerlingen in de TenP-conditie in vergelijking met leerlingen in de controleconditie. Beide verschillen zijn te classificeren als klein tot medium.

Tabel 3.7 SEQ Relationeel; effectgroottes tussen de condities

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	-.47	-.25	.27	n.s.
Controle (1) vs TenP (3)	-.31	-.36	.31	.53
TenT (2) vs TenP (3)	n.s.	n.s.	n.s.	.32

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

Voor de SEQ Fysiek worden verschillen gevonden op alle meetmomenten tussen leerlingen in de controlegroep en een of meer interventiecondities (Tabel 3.8). Op de beginmeting wordt er over leerlingen in zowel de TenT-conditie en TenP-conditie meer fysieke victimisatie gerapporteerd dan over leerlingen in de controlegroep. Hetzelfde is het geval op de tussenmeting. De omvang van deze effecten is klein tot medium. Over leerlingen in de TenP-conditie wordt meer fysieke victimisatie gerapporteerd dan over leerlingen in de TenT-conditie. Dit effect kan worden geclassificeerd als klein tot medium. Bij de eindmeting draait de richting van de gevonden effecten om en wordt er over leerlingen in zowel de TenT-conditie als de TenP-conditie minder fysieke victimisatie gerapporteerd dan over leerlingen in de controlegroep. Deze effecten zijn klein tot medium. Tijdens de follow-upmeting blijkt er geen verschil te zijn tussen de leerlingen in de TenT-conditie en de controlegroep. Wel wordt over leerlingen in de TenP-conditie minder fysieke victimisatie gerapporteerd dan over leerlingen in de controlegroep (klein tot medium effect). Tot slot wordt er een verschil gevonden tussen de interventiecondities tijdens de follow-upmeting. Leerlingen in de TenP-conditie laten minder fysieke victimisatie zien dan leerlingen in de TenT-conditie (klein tot medium effect).

Tabel 3.8 SEQ Fysiek; effectgroottes tussen de condities

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	-.60	-.23	.37	n.s.
Controle (1) vs TenP (3)	-.44	-.54	.41	.36
TenT (2) vs TenP (3)	n.s.	-.31	n.s.	.45

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

Voor de HSCS zijn de contrastanalyses opgenomen in Tabel 3.9. Op de beginmeting worden er verschillen gevonden tussen de leerlingen in de controlegroep en leerlingen in de beide interventiecondities. Leerlingen in de controleconditie worden als meer emotioneel competent ingeschat dan leerlingen in de interventiecondities (klein tot medium effect). Op de tussenmeting worden geen verschillen tussen de condities gevonden. Op de eind- en follow-upmeting worden geen verschillen gevonden tussen de TenT-conditie en de controlegroep. Hier worden echter wel verschillen gevonden tussen de controlegroep en de TenP-conditie, waarbij leerlingen in de TenP-conditie als meer emotioneel competent worden ingeschat dan leerlingen in de controleconditie (klein tot medium effect). Wanneer de interventiecondities worden vergeleken, is vast te stellen dat op zowel de eind- als follow-upmeting leerlingen in de TenTconditie als minder emotioneel competent worden ingeschat dan leerlingen in de TenP-conditie (negatieve effectgrootte). Deze verschillen kunnen worden geclassificeerd als klein.

Tabel 3.9 HSCS; effectgroottes tussen de condities

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	.24	n.s.	n.s.	n.s.
Controle (1) vs TenP (3)	.24	n.s.	-.21	-.45
TenT (2) vs TenP (3)	n.s.	n.s.	-.18	-.19

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

De Gut Feeling is alleen in de eerste drie meetrondes afgenomen. Contrastanalyses laten zien dat er op begin- en tussenmeting geen verschillen zijn tussen de verschillende condities (zie Tabel 3.10). Op de eindmeting is er geen verschil tussen de controleconditie en de TenT-conditie. Wel worden er over de leerlingen in de TenP-conditie minder zorgen gerapporteerd dan over leerlingen in de controlegroep (klein tot medium effect). Ook tussen de interventiecondities wordt een verschil gevonden, over

leerlingen in de TenT-conditie worden meer zorgen gerapporteerd dan over leerlingen in de TenP-conditie. De omvang van dit effect is klein.

Tabel 3.10 Gut Feeling; effectgroottes tussen de condities

Contrast	Meetronde		
	M1 (Begin)	M2 (Tussen)	M3 (Eind)
Controle (1) vs TenT (2)	n.s.	n.s.	n.s.
Controle (1) vs TenP (3)	n.s.	n.s.	.34
TenT (2) vs TenP (3)	n.s.	n.s.	.14

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

Op de Y-CATS Warmte is geen enkel significant verschil gevonden tussen de verschillende condities op de verschillende meetmomenten. Om compleet te zijn, zijn deze bevindingen toch opgenomen in Tabel 3.11.

Tabel 3.11 Y-CATS Warmte; effectgroottes tussen de condities

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	n.s.	n.s.	n.s.	n.s.
Controle (1) vs TenP (3)	n.s.	n.s.	n.s.	n.s.
TenT (2) vs TenP (3)	n.s.	n.s.	n.s.	n.s.

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

Tot slot worden ook de contrasten die aan het licht zijn getreden op de Y-CATS Conflict besproken (zie Tabel 3.12). Op begin-, tussen- en follow-upmeting zijn geen verschillen ontstaan tussen de verschillende condities. Ook wordt op de eindmeting geen verschil gevonden tussen de interventiecondities. Wel blijkt de controlegroep af te wijken van de interventiecondities op de eindmeting. Leerlingen in de controlegroep rapporteren op de eindmeting meer conflict met hun leerkracht dan leerlingen in zowel de TenT- conditie als de TenP-conditie. Deze effecten worden ingeschat als klein tot medium.

Tabel 3.12 Y-CATS Conflict; effectgroottes tussen de condities

Contrast	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
Controle (1) vs TenT (2)	n.s.	n.s.	.30	n.s.
Controle (1) vs TenP (3)	n.s.	n.s.	.28	n.s.
TenT (2) vs TenP (3)	n.s.	n.s.	n.s.	n.s.

Noot Alle vermelde effectgroottes zijn significant op $p < .05$ of minder.

3.6 Samenvatting van de resultaten per domein

De afgenomen instrumenten zijn in te delen in vier domeinen: gedragsproblemen, leerkracht-leerlingrelatie, sociaal-emotionele ontwikkeling en zorgelijke ontwikkeling. Elk domein bestaat uit minstens één en maximaal drie schalen. Het domein gedragsproblemen omvat de SDQ/PBSI Totaal. Het domein leerkracht-leerlingrelatie wordt gerepresenteerd door de LLRV Conflict, als maat voor de perceptie van de leerkracht, en de Y-CATS Warmte en Conflict, als maten voor de positieve en negatieve perceptie van de leerling. Het domein sociaal-emotionele ontwikkeling omvat drie instrumenten: de SEQ Relationeel, SEQ Fysiek en HSCS. Deze instrumenten beschrijven elk de perceptie van de leerkracht, op het gebied van relationele dan wel fysieke victimisatie en emotionele competentie. Tot slot omvat het domein zorgelijke ontwikkeling enkel de Gut Feeling. Deze maat is alleen in de eerste drie meetrondes afgenomen. Per domein beschrijven we wat de trends in de verschillen tussen de condities over tijd zijn. Een overzicht hiervan is opgenomen in Tabel 3.13.

Op het domein gedragsproblemen ontstaan in de loop van de interventie meerdere verschillen. Op de beginmeting zijn nog geen verschillen te zien. Bij de tussenmeting ontstaat een verschil tussen de TenP-conditie en zowel de controle- als TenT-conditie. Leerlingen in de TenP-conditie hebben ernstiger gedragsproblemen vergeleken met de leerlingen in de andere condities. Dit verschil is onverwacht, aangezien de leerlingen in beide interventiecondities op dat moment één jaar Taakspel achter de rug hadden, terwijl hun startpositie op de beginmeting niet verschillend was. Vanaf de eind- naar de follow-upmeting is een stabilisatie van effecten te zien. Leerlingen in de TenT-conditie hebben minder ernstige gedragsproblemen dan in de controlegroep, en leerlingen in de TenP-conditie hebben nog minder ernstige gedragsproblemen dan in beide andere condities. Deze bevinding is in lijn met de verwachtingen, zoals geformuleerd in hoofdstuk 1. Het laat zien dat leerlingen die twee jaar Taakspel achter de rug hebben (TenT-conditie) minder ernstige gedragsproblemen vertonen dan leerlingen in klassen waar geen specifieke sociaal-emotionele interventie heeft plaatsgevonden (controlegroep). Leerlingen die een jaar Taakspel en een jaar Taakspel én PAD (TenP-conditie) hebben gekregen vertonen nog minder ernstige gedragsproblemen dan de leerlingen die alleen Taakspel hebben

gekregen. Dit geldt voor de eindmeting, maar ook voor de follow-upmeting, waarbij in het tussenliggende jaar geen interventie werd toegepast.

Op het tweede domein, de leerkracht-leerlingrelatie, zijn op de begin-, tussen- en follow-upmeting geen verschillen tussen de condities. Opvallend is dat, tijdens de eindmeting, zowel in de perceptie van leerkracht als die van de leerling meer conflict wordt waargenomen bij de leerlingen in de controlegroep dan bij de leerlingen in de interventiecondities. Er is geen sprake van een verschil in warmte tussen de drie groepen op deze meting. Deze resultaten lijken er op te wijzen dat na het spelen van twee jaar Taakspel dan wel Taakspel en PAD zowel leerkrachten als leerlingen minder conflict ervaren in hun relatie. Dit is echter geen blijvend effect, bij de follow-upmeting zijn verschillen tussen de condities weer verdwenen.

Op het derde domein, sociaal-emotionele ontwikkeling, is er bij de beginmeting een verschil tussen leerlingen in de controlegroep en leerlingen interventiegroepen, ten gunste van de controlegroep. Dit is te zien op alle drie de instrumenten binnen dit domein. Na twee jaar, op de eindmeting, is het effect omgedraaid. Leerlingen in de interventiegroepen laten nu minder relationele en fysieke victimisatie zien dan leerlingen in de controlegroep. Dit effect is voor de TenT-conditie niet duurzaam, op de follow-upmeting laten alleen de leerlingen in de TenP-conditie minder victimisatie zien. De ontwikkeling in emotionele competentie valt grotendeels samen met het beeld op de victimisatie-schalen, met dat verschil dat nu enkel leerlingen in de TenP-conditie in het voordeel zijn ten opzichte van de controlegroep. Ook dit effect is duurzaam en terug te zien in de follow-upmeting.

Tot slot is er het domein zorgelijke ontwikkeling, bestaande uit de Gut Feeling. Op dit domein zijn tijdens begin- en tussenmeting geen verschillen tussen de verschillende condities. Op de eindmeting worden over de leerlingen in de TenP-conditie minder zorgen gerapporteerd dan over leerlingen in de controlegroep en TenT-conditie.

Over het geheel gezien lijkt er, vooral in de domeinen gedragsproblemen en sociaal-emotionele ontwikkeling, een lichte tendens ten gunste van de interventiecondities en meest specifiek ten gunste van de TenP-conditie plaats te vinden. Er zijn echter wel enkele afwijkende en lastig verklaarbare verschillen door het verloop van de metingen heen.

Tabel 3.13 Verschillen tussen de drie condities per gedragsdomein en per meetronde.

Instrumenten per domein	Meetronde			
	M1 (Begin)	M2 (Tussen)	M3 (Eind)	M4 (Follow-up)
<i>Gedragsproblemen</i>				
SDQ/PBSI Totaal	1=2=3 Geen verschillen	1=2<3 Ernstige gedragsproblemen in TenP dan in beide andere condities	1>2>3 Minder ernstige gedragsproblemen in TenT, en nog minder in TenP	1>2>3 Minder ernstige gedragsproblemen in TenT, en nog minder in TenP
<i>Leerkracht-leerlingrelatie</i>				
LLRV Conflict	1=2=3 Geen verschillen	1=2=3 Geen verschillen	1>2=3 Meer conflict in CON dan in beide andere condities	1=2=3 Geen verschillen
YCATS Conflict	1=2=3 Geen verschillen	1=2=3 Geen verschillen	1>2=3 Meer conflict in CON dan in beide andere condities	1=2=3 Geen verschillen
YCATS Warmte	1=2=3 Geen verschillen	1=2=3 Geen verschillen	1=2=3 Geen verschillen	1=2=3 Geen verschillen
<i>Sociaal-emotionele ontwikkeling</i>				
SEQ Relatieve	1<2=3 Minder victimisatie in CON dan in beide andere condities	1<2=3 Minder victimisatie in CON dan in beide andere condities	1>2=3 Meer victimisatie in CON dan in beide andere condities	1=2>3 Minder victimisatie in TenP dan in beide andere condities
SEQ Fysiek	1<2=3 Minder victimisatie in CON dan in beide andere condities	1<2<3 Minder victimisatie in TenT en nog minder in CON	1>2=3 Meer victimisatie in CON dan in beide andere condities	1=2>3 Minder victimisatie in TenP dan in beide andere condities
HSCS	1>2=3 Gunstiger ontwikkeling in CON dan in beide andere condities	1=2=3 Geen verschillen	1=2<3 Gunstiger ontwikkeling in TenP dan in beide andere condities	1=2<3 Gunstiger ontwikkeling in TenP dan in beide andere condities
<i>Zorgelijke ontwikkeling</i>				
Gut Feeling	1=2=3 Geen verschillen	1=2=3 Geen verschillen	1=2>3 Minder zorgelijke ontwikkeling in TenP dan in beide andere condities	Niet afgenomen

Noot 1 = Controle (CON), 2 = TenT, 3 = TenP.

3.7 Aantallen leerlingen met gedragsproblemen

In deze paragraaf richten we ons op de verschillen in de aantallen kinderen met gedragsproblemen per meetmoment tussen de drie verschillende condities. Voor de SDQ zijn normscores beschikbaar om leerlingen in te delen in vier categorieën: geen gedragsproblemen, licht verhoogd, hoog en erg hoog (Goodman & Goodman, 2011). Omdat de aantallen leerlingen met gedragsproblemen in de categorieën verhoogd, hoog en erg hoog elk niet groot zijn, is ervoor gekozen om een dummy variabele te creëren, waarbij onderscheid wordt gemaakt tussen leerlingen zonder gedragsproblemen en leerlingen met een indicatie van gedragsproblemen (van licht verhoogd tot en met erg hoog). De follow-upmeting wordt buiten beschouwing gelaten, aangezien hier als maat voor gedragsproblematiek een ander instrument (PBSI) is afgenomen, met andere normscores.

Om te onderzoeken of er verschillen zijn in de verdeling van leerlingen met en zonder gedragsproblemen per conditie is er gebruik gemaakt van chi-kwadraattoetsing. Omdat er per meetronde verschillende aantallen leerlingen met complete data waren (zoals beschreven in paragraaf 2.3 en 2.6), is ervoor gekozen om de analyses per meetronde te doen. Wat betreft de interpretatie van de effectgrootte (Cohen's w) baseren we ons op de suggesties van Cohen (1988). Bijgevolg spreken we bij een effectgrootte tot 0.10 van een klein effect, bij een effectgrootte van 0.30 van een medium effect, bij een effectgrootte van 0.50 van een groot effect. Contrastanalyses zijn gedaan door de condities tegen elkaar af te zetten middels de chi-kwadraattoets. Interpretatie van deze verschillen is gebaseerd op kansverhoudingen (*odds ratios*).

De aantallen en percentages leerlingen met gedragsproblemen zijn weergegeven in Tabel 3.14. Tijdens de begin- en tussenmeting zijn er geen significante verschillen vastgesteld in het aantal leerlingen met gedragsproblemen, met respectievelijk $\chi^2(df \text{ is } 2) = 3.90, p = .14$ en $\chi^2(df \text{ is } 2) = .82, p = .66$. Op de eindmeting is echter wel een verschil gevonden, met $\chi^2(df \text{ is } 2) = 16.40, p < .01$. Dit betekent dat de condities verschillen in de verdeling van leerlingen met en zonder gedragsproblematiek tijdens de derde meetronde, de eindmeting. De effectgrootte bedraagt 0.17 (Cohen's w), hetgeen kan worden geïnterpreteerd als een klein tot medium effect.

Contrastanalyses hebben twee significante contrasten laten zien. Op de eindmeting bestaat er een significant verschil tussen het aantal leerlingen met gedragsproblemen in de controlegroep en het aantal leerlingen met gedragsproblemen in de TenT-conditie, $\chi^2(df \text{ is } 1) = 6.70, p < .01$. Dit verschil valt te interpreteren op basis van *odds ratios*; het laat zien dat een leerling in de controlegroep 1.86 keer zoveel kans had op gedragsproblemen als een leerling in de TenT-conditie. Ook tussen leerlingen in de controlegroep en leerlingen in de TenP-conditie blijkt er een verschil te bestaan, $\chi^2(df \text{ is } 1) = 16.44, p < .001$. Dit betekent, op basis van *odds ratios*, dat een leerling in de controlegroep 2.76 keer zoveel kans had op gedragsproblemen als een leerling in de TenP-conditie. Tussen de TenT-conditie en de TenP-conditie werd geen significant verschil gevonden in het aantal leerlingen met gedragsproblematiek, $\chi^2(df \text{ is } 1) = 3.51, p = .06$.

Tabel 3.14 Aantal leerlingen met gedragsproblemen per conditie en per meetronde; tussen haakjes de percentages ten opzichte van het totaal aantal leerlingen per conditie en meetronde

Conditie	M1 (Begin) n = 650 leerlingen	M2 (Tussen) n = 582 leerlingen	M3 (Eind) n = 551 leerlingen
Controle	38 (30.6)	31 (25.6)	49 (48.0)
TenT	111 (37.5)	59 (24.1)	79 (33.2)
TenP	95 (41.3)	60 (27.8)	53 (25.1)
Totaal	244 (37.5)	150 (25.8)	181 (32.8)

Het onderzoek heeft tot doel gehad te onderzoeken wat het complementaire effect is in preventieve werking van de gecombineerde aanbieder van Taakspel en PAD. Onderzocht is of deze combinatie een reductie van het aantal kinderen met gedragsproblemen en van de ernst van deze problemen tot gevolg heeft. Hierbij zijn vergelijkingen gemaakt tussen drie verschillende condities: Taakspel (TenT-conditie), Taakspel + PAD (TenP-conditie) en de controleconditie (geen interventie). De verwachtingen zijn als volgt geformuleerd:

- a. Taakspel is – in vergelijking met de controleconditie – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.
- b. ‘Taakspel + PAD’ is – in vergelijking met de controleconditie – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.
- c. ‘Taakspel + PAD’ is – in vergelijking met de conditie Taakspel – effectief wat betreft de reductie van de ernst van gedragsproblemen en van het aantal leerlingen met gedragsproblemen.

Hypothese a: contrast tussen controlegroep en TenT-conditie

De eerste hypothese betreft het verschil tussen de leerlingen in de controlegroepen en leerlingen in de Taakspelconditie. Aan de hand van de hiervoor beschreven bevindingen kunnen we vijf conclusies formuleren.

1. Zoals voorspeld is, is ten opzichte van de controlegroep de ernst van de gedragsproblemen in de TenT-conditie afgenomen. Blijkens de follow-upmeting na één jaar is dit effect stabiel en blijft het bestaan.
2. Zoals voorspeld is, is ten opzichte van de controlegroep het aantal leerlingen met gedragsproblemen in de TenT-conditie afgenomen.
3. Bovendien rapporteren leerkrachten en leerlingen in de TenT-conditie minder conflict in de leerkracht-leerlingrelatie dan leerkrachten en leerlingen in de controlegroep. Dit effect is evenwel niet duurzaam en valt blijkens de follow-upmeting na een jaar weg.
4. Bovendien laten leerlingen in de TenT-conditie na twee jaar een gunstiger sociaal-emotionele ontwikkeling zien dan leerlingen in de controlegroep. Dit effect is niet duurzaam en valt weg in de follow-upmeting na een jaar.
5. Er zijn geen verschillen in zorgelijke ontwikkeling zoals gerapporteerd door de leerkracht over leerlingen in de TenT-conditie en leerlingen in de controlegroep.

Hypothese b: contrast tussen controlegroep en TenP-conditie

De tweede hypothese betreft het verschil tussen leerlingen in de controlegroep en leerlingen in de

conditie 'Taakspel + PAD'. Aan de hand van de hiervoor beschreven bevindingen kunnen we vijf conclusies formuleren.

1. Zoals voorspeld is, is ten opzichte van de controlegroep de ernst van de gedragsproblemen in de TenP-conditie afgenomen. Blijkens de follow-upmeting blijft dit effect bestaan.
2. Zoals voorspeld is, is ten opzichte van de controlegroep het aantal leerlingen met gedragsproblemen in de TenP-conditie afgenomen.
3. Bovendien rapporteren leerkrachten en leerlingen in de TenP-conditie minder conflict in de leerkracht-leerlingrelatie dan leerkrachten en leerlingen in de controlegroep. Dit effect is evenwel niet duurzaam, en valt blijkens de follow-upmeting na een jaar weg.
4. Bovendien laten leerlingen in de TenP-conditie na twee jaar een betere sociaal-emotionele ontwikkeling zien dan leerlingen in de controlegroep. Dit effect is duurzaam en blijvend, zoals blijkt uit de follow-upmeting een jaar na de eindmeting.
5. Daarnaast wordt over leerlingen in de TenP-conditie na de eindmeting een minder zorgelijke ontwikkeling gerapporteerd door de leerkracht dan over leerlingen in de controlegroep.

Hypothese c: contrast tussen TenT-conditie en TenP-conditie

De derde hypothese betreft het verschil tussen leerlingen in de Taakspelconditie en leerlingen in de Taakspel + PAD-conditie. Hiermee wordt het complementaire effect van PAD op Taakspel gezien. Aan de hand van de hiervoor beschreven bevindingen kunnen vijf conclusies worden geformuleerd.

1. Zoals voorspeld is, is ten opzichte van de TenT-conditie de ernst van de gedragsproblemen bij leerlingen in de TenP-conditie afgenomen. Blijkens de follow-upmeting is dit effect duurzaam.
2. Tegen de voorspelling in is, ten opzichte van de TenT-conditie het aantal leerlingen met gedragsproblemen in de TenP-conditie niet geringer.
3. Daarnaast zijn er geen verschillen gevonden tussen de TenT-conditie en de TenP-conditie in de mate van conflict in de leerkracht-leerlingrelatie.
4. Wel laten leerlingen in de TenP-conditie na twee jaar een gunstiger sociaal-emotionele ontwikkeling zien dan leerlingen in de TenT-conditie. Blijkens de follow-upmeting na een jaar is dit effect duurzaam.
5. Bovendien wordt over leerlingen in de TenP-conditie op de eindmeting een minder zorgelijke ontwikkeling gerapporteerd door de leerkracht dan over leerlingen in de TenT-conditie.

Meerniveau-analyses

Tot slot, in een addendum rapporteren we analyses op basis van multilevel modellen, waarin de hiërarchische structuur van de data tot uitdrukking wordt gebracht. Deze analyses leiden wat betreft de grootte van de verschillen tussen de condities tot dezelfde conclusies: leerlingen in de TenT- en de TenP-conditie zijn in een aantal opzichten – hiervoor genoemd -- in het voordeel. Echter, in veel

gevallen zijn de verschillen niet statistisch significant. Implicatie daarvan is dat de generaliseerbaarheid van de experimentele uitkomsten naar de populatie van alle basisschoolleerlingen onzeker is. Wel bleek uit de meerniveau-analyses dat leerlingen in de TenT- en de TenP-conditie op de eindmeting minder gedragsproblemen vertonen dan leerlingen in de controleconditie. Een jaar later tijdens de follow-upmeting zijn de verschillen er nog steeds, maar ze zijn niet meer statistisch significant.

Referenties

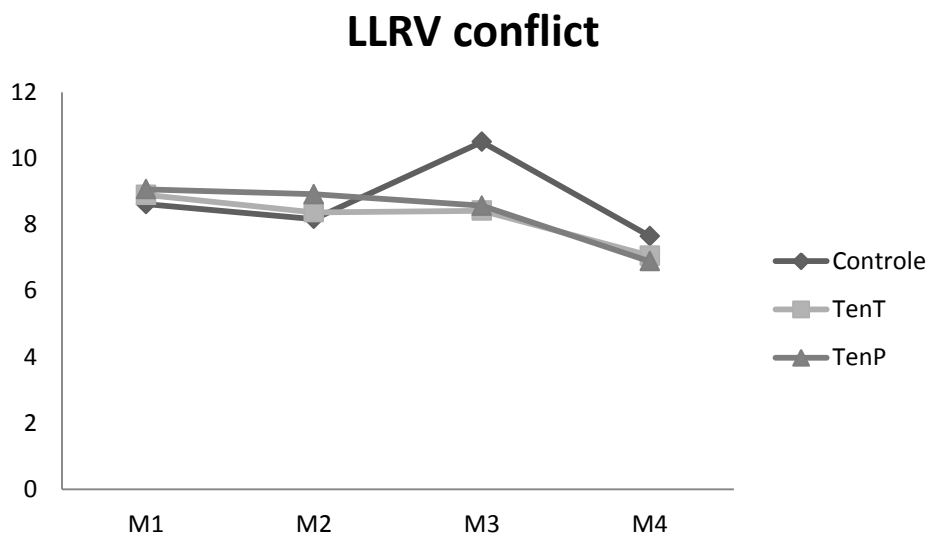
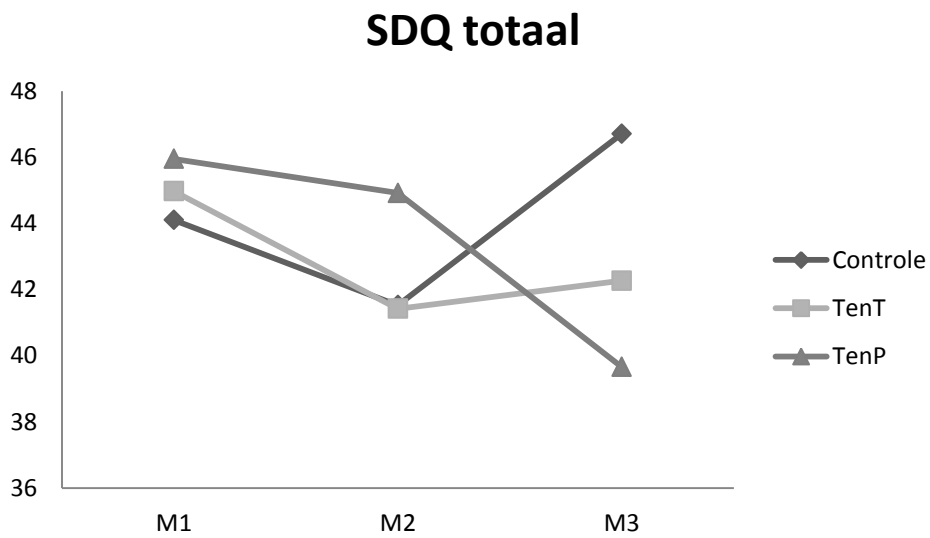
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (second ed.)*. Lawrence Erlbaum Associates.
- Conduct Problems Prevention Research Group (1999a). Initial impact of the fast track prevention trial for conduct problems: I. The high-risk sample. *Journal of Consulting and Clinical Psychology, 67*(5), 631-647.
- Conduct Problems Prevention Research Group (1999b). Initial impact of the fast track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology, 67*(5), 648-657.
- Conduct Problems Prevention Research Group (2002). Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology, 30*(1), 19-35.
- Crick, N. R., & Grotpeter, J. K. (1996). Children's treatment by peers: victims of relational and overt aggression. *Developmental Psychopathology, 8*, 367-380.
- Dolan, L. J., Jaylan, T., Werthamer, L., & Kellam, S. (1989). *The Good Behavior Game Manual*. Baltimore: The John Hopkins Prevention Research Center.
- Domitrovich, C., Bradshaw, C. P., Greenberg, M. T., Embry, D., Poduska, J., M., & Ialongo, N. S. (2010). Integrated models of school-based prevention: logic and theory. *Psychology in the Schools, 41*(1), 71-88.
- Domitrovich, C., Cortes, R. C., & Greenberg, M. T. (2001). *Head Start Competence Scale Technical Report* (Unpublished manuscript). University Park: Pennsylvania State University.
- Erasmus Medical Center (2000). *Problem behavior at school interview*. Rotterdam: Erasmus Medical Center.
- Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2005). Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry, 46*(8), 837-849.
- Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on challenging behaviors in school settings. *Review of Educational Research, 84*, 546-571.
- Furr-Holden, C. D., Ialongo, N. S., Anthony, J. C., Petras, H., & Kellam, S. G. (2004). Developmentally inspired drug prevention: Middle school outcomes in a school-based randomized prevention trial. *Drug and Alcohol Dependence, 73*(2), 149-158.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.
- Goodman, A., & Goodman, R. (2011). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry, 52*, 100-108.
- Greenberg, M. T., & Kusché, C. A. (2002). *Promoting alternative thinking strategies. Blueprint for violence prevention*. Boulder: Institute of Behavioral Sciences, University of Colorado.

- Greenberg, M. T., Kusché, C. A., & Mihalic, S. F. (1998, 2005). *Promoting Alternative Thinking Strategies (PATHS). 1st and 4th ed.* Boulder, CO: Center for the Study and Prevention of Violence.
- Ialongo, N. S., Poduska, J., Werthamer, L., & Kellam, S. (2001). The distal impact of two first-grade preventive interventions on conduct problems and disorder in early adolescence. *Journal of Emotional and Behavioral Disorders, 9*(3), 146-160.
- Ince, D. (2009). *Databank effectieve jeugdinterventies: beschrijving 'Taakspel'*. Utrecht: Nederlands Jeugdinstituut.
- Ince, D. (2010). *Databank effectieve jeugdinterventies: beschrijving 'Programma Alternatieve Denkstrategieën (PAD)'*. Utrecht: Nederlands Jeugdinstituut.
- Ialongo, N. S., Werthamer, L., Kellam, S. G., Brown, C. H., Wang, S., & Lin, Y. (1999). Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. *American Journal of Community Psychology, 27*(5), 599-641.
- Kellam, S. G., & Anthony, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trial. *American Journal of Public Health, 88*(10), 1490-1495.
- Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*(1), 5-28.
- Kellam, S. G., Rebok, G. W., Ialongo, N., & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry, 35*(2), 259-281.
- Koomen, H., Verschueren, K., & Pianta, R. (2011). *Leerkracht leerling relatie vragenlijst*. Houten: Bohn Stafleu van Loghum.
- Louwe, J. J., & Van Overveld, C. W. (2008). *Een PAD naar minder agressie* (Dissertatie). Utrecht: Uitgeverij Agiel.
- Mantzicopoulos P., & Neuharth-Pritchett, S. (2001) Development and validation of a measure to assess Head Start children's appraisals of teacher support. *Journal of School Psychology, 41*(6), 431-451.
- Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology, 13*(2), 355-375.
- Moffitt, T. E., Caspi, A., Harrington, H., & Milne, B. J. (2002). Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years. *Development and Psychopathology, 14*(1), 179-207.
- Petras, H., Kellam, S. G., Brown, C. H., Muthen, B. O., Ialongo, N. S., & Poduska, J. M. (2008). Developmental epidemiological courses leading to antisocial personality disorder and violent and criminal behavior: effects by young adulthood of a universal preventive intervention in first- and second-grade classrooms. *Drug and Alcohol Dependence, 95*(1), 45-59.
- Poduska, J. M., Kellam, S. G., Wang, W., Brown, C. H., Ialongo, N. S., & Toyinbo, P. (2008). Impact of the Good Behavior Game, a universal classroom-based behavior intervention, on young

- adult service use for problems with emotions, behavior, or drugs or alcohol. *Drug and Alcohol Dependence*, 95(1), 29-44.
- Storr, C. L., Ialongo, N. S., Kellam, S. G., & Anthony, J. C. (2002). A randomized controlled trial of two primary school intervention strategies to prevent early onset tobacco smoking. *Drug and Alcohol Dependence*, 66(1), 51-60.
- Van der Sar, A. M., & Goudswaard, M. (2001). *Docenthandleiding Taakspel voor basisonderwijs*. Rotterdam: Pedologisch Instituut.
- Van Lier, P. A. C., Huizink, A. C., & Crijnen, A. A. (2009). Impact of a preventive intervention targeting childhood disruptive behavior problems on tobacco and alcohol initiation from age 10 to 13 years. *Drug and Alcohol Dependence*, 100(3), 228-233.
- Van Lier, P. A. C., Muthén, B. O., van der Sar, R. M., & Crijnen, A. A. (2004). Preventing disruptive behavior in elementary school children: Impact of a universal, classroom based intervention. *Journal of Consulting and Clinical Psychology*, 72(3), 467-478.
- Van Lier, P. A. C., Vuijk, P., & Crijnen, A. A. (2005). Understanding mechanisms of change in the development of antisocial behavior: The impact of a universal intervention. *Journal of Abnormal Child Psychology*, 33(5), 521-535.
- Verhoeven, L., & Vermeer, A. (2001). *Taaltoets Alle Kinderen. Diagnostische toets voor de mondelinge vaardigheid Nederlands bij kinderen van groep 1 tot en met 4*. Arnhem: Cito.
- Vuijk, P., Van Lier, P. A. C., Crijnen, A. A., & Huizink, A. C. (2007). Testing sex-specific pathways from peer victimization to anxiety and depression in early adolescents through a randomized intervention trial. *Journal of Affective Disorders*, 100(1-3), 221-226.
- Wilcox, H. C., Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., et al. (2008). The impact of two universal randomized first- and second-grade classroom interventions on young adult suicide ideation and attempts. *Drug and Alcohol Dependence*, 9(1), 60-73.

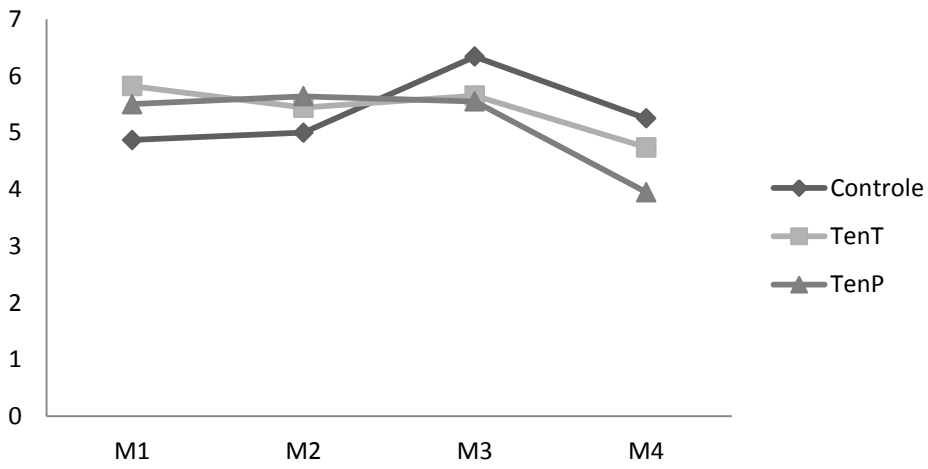
Bijlage

Grafieken met gemiddelden³ per meetinstrument, conditie en meetronde

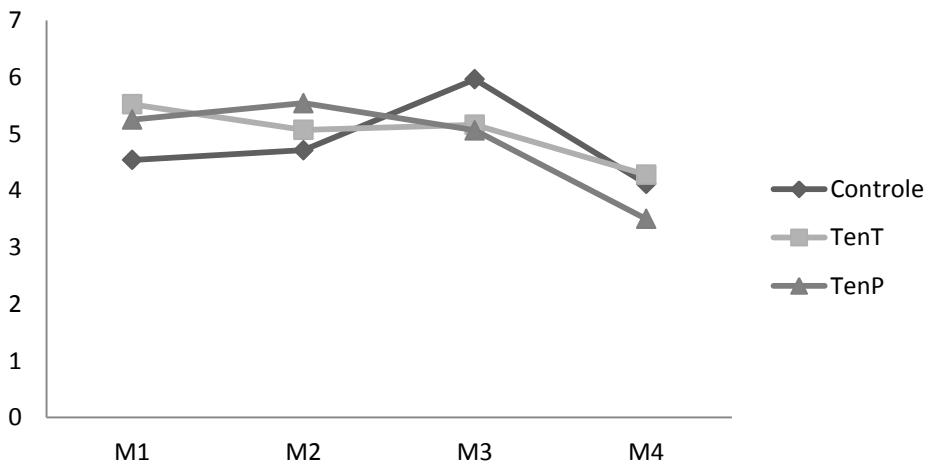


³ Zie voor standaarddeviaties en aantallen leerlingen de tabellen 3.1 tot en met 3.4.

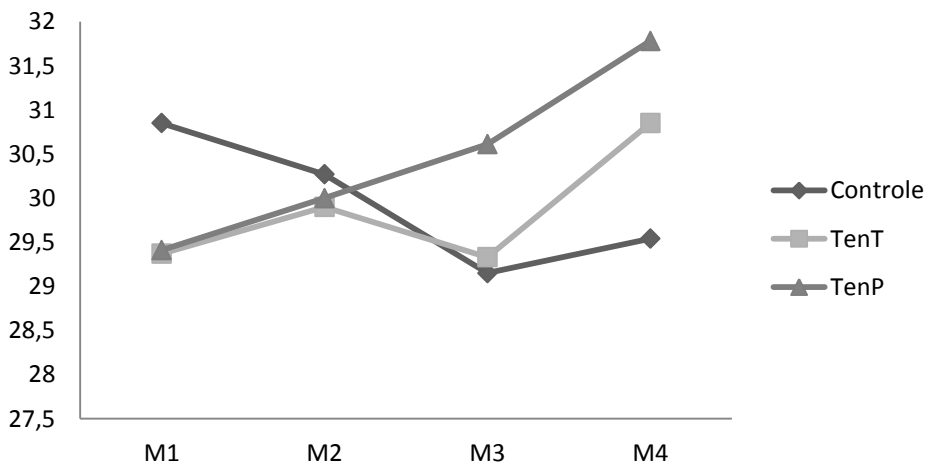
SEQ Relatieveel



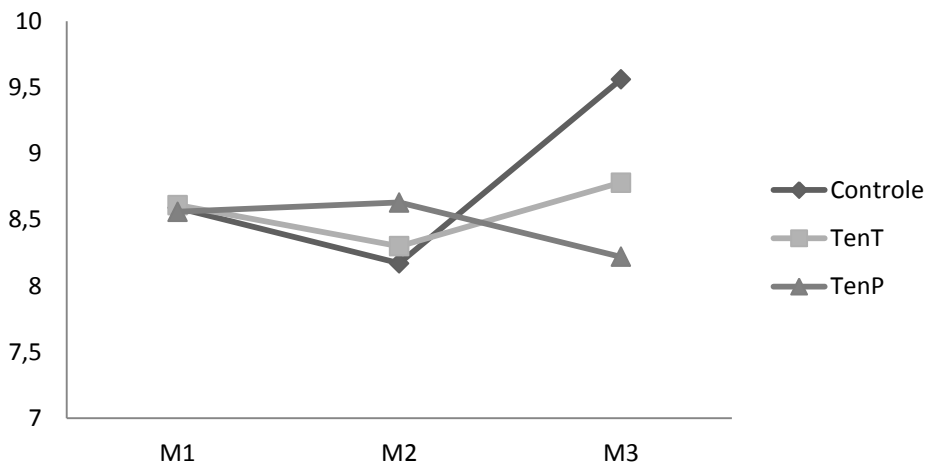
SEQ Fysiek



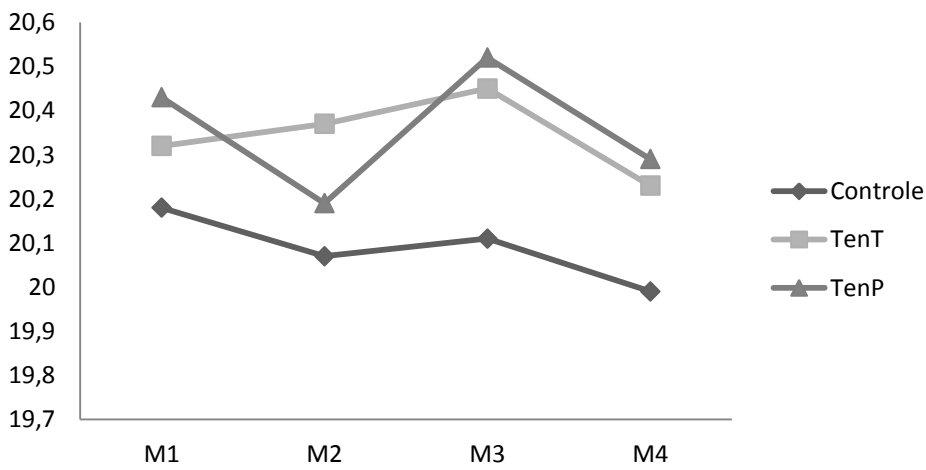
HSCS



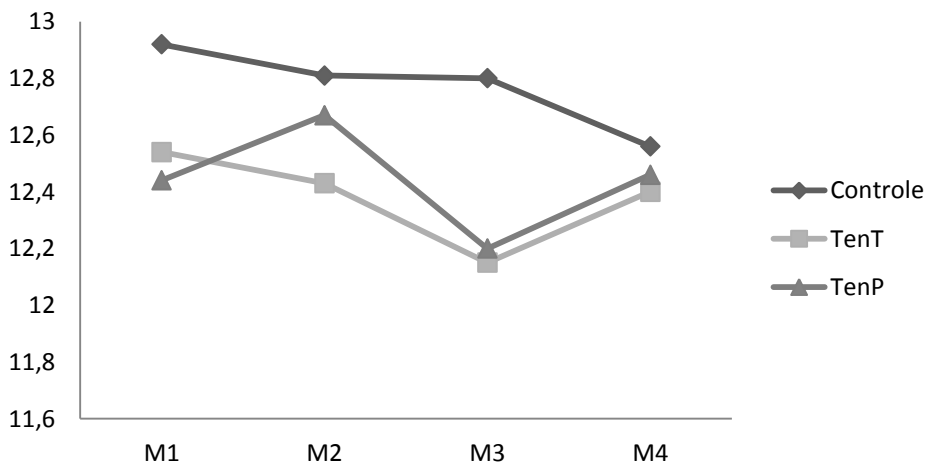
Gut Feeling



Y-CATS Warmte



Y-CATS Conflict



Meerniveau-analyses

Bij de in hoofdstuk 3 gerapporteerde resultaten is er geen rekening mee gehouden dat de data een hiërarchische structuur vertonen. Maar de werkelijkheid is, dat de leerlingen genest zijn in schoolklassen en schoolklassen genest zijn in scholen. Als gevolg daarvan zijn de waarnemingen strikt genomen niet onafhankelijk van elkaar. Leerlingen die in dezelfde klas of op dezelfde school zitten, lijken meer op elkaar dan leerlingen voor wie dat niet geldt. Naarmate deze gelijkensis groter is, zijn de uitkomsten van statistische significantietoetsen onzuiverder, wanneer men de hiërarchische structuur niet zou modelleren. Dat komt doordat de storingstermen in de vergelijkingen met elkaar gecorreleerd zijn.

Om het effect van de gecorreleerde waarnemingen op de uitkomsten van de significantietoetsingen te beoordelen zijn voor de in de paragrafen 3.1 tot en met 3.4 gerapporteerde MANOVA's alternatieve analyses gedaan, rekening houdende met de hiërarchische van de data. Geschat zijn zogenoemde *mixed model* variantie-analyses met drie niveaus (leerlingen, schoolklassen, scholen), met *random* intercepten op de beide hogere niveaus, en met de drie condities (controle, TenT en TenP) als onafhankelijke variabele. Niet te voorkomen is dat er nog enkele andere verschillen optreden tussen de genoemde MANOVA's en de hier gerapporteerde gemengde modellen. De belangrijkste is dat bij de schatting van de gemengde modellen geen multivariate toets gedaan kon worden. Een ander verschil is dat de schattingen van de gemengde modellen berusten op *maximum likelihood estimation*, en niet op de kleinste kwadratenmethode zoals bij de MANOVA's. We rapporteren hier vier tabellen, één voor elk meetmoment M1 (beginmeting) tot en met M4 (follow-up meting), te vergelijken met de tabellen 3.1 tot en met 3.4 in hoofdstuk 3.

Tabel 1 Gemiddelden (*M*), standaardfouten van het gemiddelde (*SEM*) en *F*-waarden behorende bij de verschiltoetsingen op een aantal variabelen per conditie op meetmoment 1 (*n*=650); gemengd hiërarchisch model

Variabelen	Condities						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
SDQ Totaal	44.92	1.92	45.34	1.31	45.46	1.42	0.03	1=2=3
LLRV Conflict	8.91	0.76	9.00	0.52	8.89	0.57	0.01	1=2=3
SEQ Relationeel	5.14	0.49	5.96	0.34	5.47	0.36	1.07	1=2=3
SEQ Fysiek	4.86	0.43	5.49	0.29	5.26	0.32	0.71	1=2=3
HSCS	30.63	0.87	28.93	0.62	29.83	0.65	1.34	1=2=3
Gut Feeling	8.81	0.60	8.66	0.40	8.50	0.44	0.09	1=2=3

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
TAK Zinsbegrip	36.71	0.30	36.67	0.21	36.71	0.23	0.10	1=2=3
TAK Zinsvorming	30.42	0.94	30.51	0.68	29.74	0.69	0.35	1=2=3
Y-CATS Warmte	20.18	0.15	20.32	0.09	20.44	0.11	1.04	1=2=3
Y-CATS Conflict	12.85	0.26	12.61	0.18	12.42	0.20	0.87	1=2=3

Tabel 2 Gemiddelden (M), standaardfouten van het gemiddelde (SEM) en F-waarden behorende bij de verschiltoetsingen op een aantal variabelen per conditie op meetmoment 2 (n=582); gemengd hiërarchisch model

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
SDQ Totaal	42.64	1.57	41.59	1.15	44.47	1.22	1.49	1=2=3
LLRV Conflict	8.81	0.70	8.43	0.52	8.83	0.55	0.17	1=2=3
SEQ Relationeel	5.35	0.40	5.50	0.30	5.62	0.31	0.14	1=2=3
SEQ Fysiek	5.11	0.40	5.06	0.30	5.36	0.31	0.26	1=2=3
HSCS	30.04	0.51	29.77	0.37	29.96	0.39	0.12	1=2=3
Gut Feeling	8.55	0.53	8.38	0.38	8.53	0.41	0.05	1=2=3
TAK Zinsbegrip	37.79	0.27	37.90	0.20	37.80	0.21	0.07	1=2=3
TAK Zinsvorming	31.79	0.88	31.76	0.67	31.43	0.70	0.07	1=2=3
Y-CATS Warmte	20.13	0.23	20.34	0.17	20.28	0.18	0.28	1=2=3
Y-CATS Conflict	12.83	0.33	12.50	0.25	12.65	0.26	0.33	1=2=3

Tabel 3 Gemiddelden (M), standaardfouten van het gemiddelde (SEM) en F-waarden behorende bij de verschiltoetsingen op een aantal variabelen per conditie op meetmoment 3 (n=551); gemengd hiërarchisch model

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
SDQ Totaal	47.41	2.38	42.61	1.55	39.75	1.67	3.51	1>2=3
LLRV Conflict	10.60	0.71	8.59	0.46	8.38	0.50	3.71	1>2=3
SEQ Relationeel	6.50	0.49	5.62	0.32	5.64	0.34	1.25	1=2=3
SEQ Fysiek	6.08	0.47	5.36	0.33	4.94	0.34	1.93	1=2=3
HSCS	28.99	1.01	29.27	0.67	30.54	0.72	1.18	1=2=3
Gut Feeling	9.69	0.63	8.79	0.41	8.22	0.44	1.85	1=2=3
Y-CATS Warmte	20.09	0.19	20.44	0.13	20.51	0.14	1.73	1=2=3

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
Y-CATS Conflict	12.77	0.25	12.13	0.17	12.24	0.18	2.30	1=2=3

Tabel.4 Gemiddelden (M), standaardfouten van het gemiddelde (SEM) en F-waarden behorende bij de verschiltoetsingen op een aantal variabelen per conditie op meetmoment 4 (n=382); gemengd hiërarchisch model

Variabelen	Conditie						Verschil	
	Controle (1)		TenT (2)		TenP (3)			
	M	SEM	M	SEM	M	SEM	F	Contrast
PBSI Totaal	55.67	3.18	49.79	2.50	48.76	2.73	1.53	1=2=3
LLRV Conflict	7.60	0.58	7.07	0.46	7.34	0.50	0.26	1=2=3
SEQ Relationeel	5.24	0.37	4.70	0.28	4.10	0.32	2.85	1=2=3
SEQ Fysiek	4.11	0.31	4.17	0.23	3.66	0.26	1.32	1=2=3
HSCS	29.38	0.98	30.91	0.76	31.16	0.84	1.08	1=2=3
Y-CATS Warmte	20.12	0.39	20.07	0.31	20.03	0.33	0.18	1=2=3
Y-CATS Conflict	12.48	0.44	12.37	0.35	12.87	0.38	0.51	1=2=3

In vergelijking met de tabellen 3.1 tot en met 3.4 zijn de gemiddelden van de afzonderlijke condities nauwelijks veranderd. Dat is ook niet te verwachten, omdat gemiddelden niet gevoelig zijn voor een eventuele hiërarchische datastructuur. De kleine veranderingen die er toch zijn, hangen samen met de andere schattingswijze, geen kleinste-kwadratenmethode maar *maximum likelihood*. Verder zij opgemerkt dat de tabellen 1 tot en met 4 niet de standaarddeviaties bevatten, maar de standaardfouten van het gemiddelde. Dat komt omdat de gebruikte statistische programmatuur geen schattingen produceert voor de standaarddeviaties.

Veranderingen zijn er wel ten aanzien van de statistische toetsingen van de verschillen tussen de condities. Waar op grond van MANOVA's geconcludeerd werd dat er in 17 van de 35 gevallen sprake is van een statistisch significant contrast, luidt de conclusie op basis van de gemengde hiërarchische modellen dat zulks slechts in twee van de 35 gevallen het geval is. Op het derde meetmoment, de eindmeting, wordt over leerlingen in de controleconditie ernstiger gedragsproblematiek gerapporteerd dan over leerlingen in de TenT- en in de TenP-conditie, zowel op de schaal SDQ Totaal als op de schaal LLRV Conflict. Tussen leerlingen in deze beide laatste condities treedt geen verschil op.

De vraag is nu natuurlijk hoe men met deze veranderde uitkomsten om zou kunnen gaan. Daar willen we, ter afsluiting twee opmerkingen over maken. De eerste is dat de analyses op basis van het

gemengde hiërarchische model *niet* tot veranderde inzichten leiden wat betreft de grootte van de verschillen tussen de drie condities. Deze effectgroottes blijven gelijk, onafhankelijk van de uitkomst van statistische toetsingen. De algemene uitkomst blijft daarom dat ten opzichte van de controlegroep de ernst van de gedragsproblemen in de TenT en TenP -condities is afgenomen. In de tweede plaats stellen we vast dat er twijfel is over de generaliseerbaarheid van de uitkomsten in onze steekproef naar de populatie van alle basisschoolleerlingen. Die twijfel wordt niet alleen gevoed doordat er op basis van een meerniveau-analyse weinig statistisch significante verschillen tussen de condities zijn ontstaan, al is op te merken dat de verschillen wat betreft gedragsproblemen van leerlingen ook volgens meerniveau-analyses statistisch significant in het voordeel uitvallen van de leerlingen in de TenT en TenP-condities. Van belang is ook dat er twijfel is aan de representativiteit van de steekproef. Scholen zijn geworven op basis van vrijwilligheid, als gevolg waarvan niet is uit te sluiten dat scholen met relatief veel gedragsproblematische leerlingen oververtegenwoordigd zijn. Ook het hoge uitvalpercentage – we begonnen met 650 leerlingen op de beginmeting maar eindigden met 382 leerlingen op de follow-up meting – draagt aan deze twijfel bij. Om grotere zekerheid te verkrijgen over de werkzaamheid van de programma Taakspel, al dan niet in combinatie met PAD, is meer onderzoek nodig.