

Open Review-proces - van Loon ea De effectiviteit van een schoolinterventie voor het vergroten van een pro sociaal klasklimaat voor basisschoolkinderen Jeugd in Ontwikkeling

Table of Contents

<i>Open Review-proces - van Loon ea De effectiviteit van een schoolinterventie voor het vergroten van een pro sociaal klasklimaat voor basisschoolkinderen Jeugd in Ontwikkeling</i>	1
<i>Originele manuscript</i>	2
<i>Mail over eerste versie</i>	43
<i>Mail met reviews over de eerste versie</i>	44
<i>Mail met reviews over de tweede versie</i>	56
<i>Mail met reviews over de derde versie</i>	66

Originele manuscript

Wat is er nodig voor het creëren van prosociale klassen? De effectiviteit van een schoolinterventie voor basisschoolkinderen.

Nederlandse samenvatting

Een prosociaal klasklimaat is cruciaal voor het bevorderen van het school- en psychologische welzijn van kinderen, evenals het optimaliseren van een effectieve leeromgeving.

Schoolprogramma's kunnen een prosociaal klasklimaat bevorderen, maar er zijn weinig op-evidentie-gebaseerde opties beschikbaar. De huidige studie onderzocht daarom de effectiviteit van een basisschoolprogramma gericht op het bevorderen van prosocialiteit door de intrinsieke motivatie van leerlingen te verhogen door de basispsychologische behoeften te versterken (autonomie, competentie, en verbondenheid). Gegevens zijn afkomstig van 2237 leerlingen (46% jongen, $M = 9.55$ jaar, $SD = 1.21$) en 225 leerkrachten (76% vrouw, $M = 38.69$ jaar, $SD = 11.09$). Scholen werden gerandomiseerd in een experimentele conditie ($N = 20$ scholen) of een controleconditie ($N = 22$ scholen). Regressieanalyses werden uitgevoerd halverwege en aan het einde van programma-implementatie, gecorrigeerd voor baseline niveaus, waarbij condities werden vergeleken in de volledige steekproef en in een subgroep met ideale programmatrouw. Wanneer het programma volgens plan werd geïmplementeerd, waren leerlingen beter af dan leerlingen in de controleconditie wat betreft prosociaal klasklimaat, basispsychologische behoeften, welbevinden en veiligheid op school, en antisociaal gedrag. Leerkrachten rapporteerden een toegenomen positieve houding ten opzichte van de stimulering van basispsychologische behoeften en burgerschapsvaardigheden van leerlingen. Inzicht in optimale implementatie is essentieel voor programma-effectiviteit.

Trefwoorden: Schoolinterventies, kinderen, prosociaal klasklimaat, welbevinden, zelfdeterminatietheorie.

Abstract

Having a prosocial classroom climate is crucial for promoting children's school and psychological wellbeing, as well as in optimizing an effective learning environment. School-based programs can foster more prosocial classroom climates, but few evidence-based options are available. This study examined the effectiveness of a primary school program aimed at promoting prosociality by enhancing students' intrinsic motivation through the support of basic psychological needs (autonomy, competence, and relatedness). Data stem from 2,237 students (46% boys, $M_{\text{age}} = 9.55$ years, $SD = 1.21$) and 225 teachers (76% female, $M_{\text{age}} = 38.69$ years, $SD = 11.09$) across 42 schools. Schools were randomized to an experimental condition ($N = 20$ schools) or control condition ($N = 22$ schools). Regression analyses were conducted at mid- and end-of-year, adjusted for baseline levels, comparing conditions in the full sample and a subsample with ideal program fidelity. When implemented as intended, students who received the program fared better than those in the control condition in terms of prosocial classroom climate, basic psychological needs, school wellbeing and safety, and antisocial behavior. Teachers reported increased attitudes toward stimulating basic psychological need fulfilment and civic skills of students. Understanding the key to optimal implementation is crucial for program effectiveness.

Keywords: School-based intervention programs, children, prosocial classroom climate, wellbeing, self-determination theory.

Bijsluiter voor de praktijk

- Een pro sociaal klasklimaat is van cruciaal belang om het school- en psychologisch welzijn van leerlingen te bevorderen en een effectieve leeromgeving te optimaliseren.
- Het schoolprogramma [GEMASKEERD VOOR REVIEW] richt zich op het theoretisch kader dat het versterken van basispsychologische behoeften – autonomie, competentie, en verbondenheid – intrinsieke (prosociale) motivatie en het creëren van pro sociale klassen bevordert.
- In de huidige studie onderzochten we met een cluster RCT design de resultaten van [GEMASKEERD VOOR REVIEW] bij implementatie op meer dan 40 basisscholen in Nederland.
- Wanneer het programma wordt uitgevoerd als bedoeld, waren leerlingen op interventiescholen beter af dan leerlingen op controlescholen wat betreft een pro sociaal klasklimaat, basispsychologische behoeften, welbevinden en veiligheid op school en antisociaal gedrag.
- Het is belangrijk om de sleutel tot optimale implementatie te begrijpen, aangezien dit een cruciale voorwaarde is voor een succesvol en effectief programma.

Totaal aantal woorden: 9710 woorden.

1. Inleiding

Om het school- en psychologisch welzijn van leerlingen te bevorderen en een effectieve leeromgeving te optimaliseren, is een prosociaal klasklimaat van cruciaal belang (Evans et al., 2009; Wang et al., 2020). Een positief, zorgzaam en ondersteunend klasklimaat wordt geassocieerd met minder pestgedrag en slachtofferschap van pesten (Saarento & Salmivalli, 2015; Thornberg et al., 2022). Prosociale klassen worden gekenmerkt door prosociaal gedrag, een prosociale norm in de klas, een hoge mate van samenwerking, weinig conflicten en hoogwaardige interacties tussen leerkrachten en klasgenoten (Khalfaoui et al., 2021; Villardón-Gallego et al., 2018; Wang et al., 2020).

Gebaseerd op de theorie van basispsychologische behoeften (Ryan & Deci, 2017) wordt gesteld dat het creëren van een prosociaal klasklimaat strategieën vereist die de intrinsieke motivatie van leerlingen voor positief gedrag stimuleren door te voorzien in hun basispsychologische behoeften (Cheon et al., 2019; Ryan & Deci, 2000, 2017). Deze behoeften hebben betrekking op autonomie (het gevoel van controle over eigen beslissingen), competentie (het vertrouwen in eigen kunnen) en verbondenheid (het gevoel van verbondenheid en veiligheid). Vervulling van deze basispsychologische behoeften wordt geassocieerd met meer empathie voor anderen, betere vaardigheden om met conflicten om te gaan, internalisering van een prosociale klasnorm, een zorgzame klasomgeving en de motivatie voor prosociaal gedrag (Cheon et al., 2019). De *goal-framing* theorie verklaart hoe dergelijk gedrag in klassen kan blijven bestaan: de heersende groepsnorm beïnvloedt welk gedrag wordt ondersteund en versterkt en welk gedrag niet (Lindenberg & Steg, 2007). Prosociaal gedrag moet daarom een sociale norm in de klas worden om te zorgen dat dit gedrag een blijvend onderdeel wordt van het klasklimaat. Het handhaven van een dergelijke norm gebeurt door het waarnemen van prosociale attitudes en gedrag, en door de steun van anderen (Lindenberg & Steg, 2007).

Om dit doel te bereiken, onderzoekt de huidige studie de effectiviteit van een schoolprogramma dat een prosociaal klasklimaat in basisscholen bevordert, als gevolg van het verhogen van intrinsieke prosociale motivatie door het vervullen van de basispsychologische behoeften (Nederlandse aanpassing van *Meaningful Roles*; [GEMASKEERD VOOR REVIEW]). *Meaningful Roles* is een antipestprogramma voor middelbare scholieren dat pesters prosociale alternatieven voor hun gedrag biedt door het vergroten van autonomie en verbondenheid (Ellis et al., 2016). Resultaten laten zien dat het programma agressief en probleemgedrag vermindert (Ellis et al., 2016), wat wijst op positieve effecten van het bevorderen van prosociaal gedrag. Ander onderzoek toonde ook de

effectiviteit van interventieprogramma's aan die prosociaal gedrag bij kinderen en adolescenten bevorderen (Caprara et al., 2015; Mesurado et al., 2018; Shin & Lee, 2021), met name voor programma's die ontwikkelt zijn om de sociale competentie vergroten in plaats van probleemgedrag te voorkomen (Shin & Lee, 2021). Bovendien toonden meta-analyses aan dat programma's voor sociaal en emotioneel leren in de klas kleine tot grote effecten hadden op het verbeteren van sociale en emotionele vaardigheden, positieve attitudes, positief sociaal gedrag en psychologisch welzijn bij kinderen en adolescenten (Durlak et al., 2011; Sklad et al., 2012; Tanner-Smith et al., 2018; Taylor et al., 2017). Daarnaast toonde een meta-analyse onder jongeren en volwassenen aan dat programma's gericht op het versterken van basispsychologische behoeften (i.e., autonomie, competentie en verbondenheid) geassocieerd waren met een toename van basisbehoeften en fysiek en psychologisch welzijn (Ntoumanis et al., 2021). Bovendien leverde eerder onderzoek in basis- en middelbare scholen, waarbij programma's werden geïmplementeerd gericht op motivatie en bevrediging van basispsychologische behoeften, positieve effecten op basisbehoeften, prosociaal gedrag en een positief klasklimaat, evenals een afname van antisociaal gedrag (Cheon et al., 2019; Manzano-Sánchez & Valero-Valenzuela, 2019). Er is echter behoefte aan (cluster) gerandomiseerde gecontroleerde onderzoeken om de effecten van dit soort programma's in diverse, grote steekproeven van basisschoolleerlingen uitgebreid te evalueren.

Om deze hiaten aan te pakken, beoogt deze studie de kennis over school-gerelateerde welzijnsonderzoek op twee manieren te vergroten door: a) empirische ondersteuning te bieden aan het theoretisch kader dat het versterken van basispsychologische behoeften een prosociaal klasklimaat kan bevorderen, en b) gebruik te maken van een cluster gerandomiseerd gecontroleerd design (RCT) – een gouden standaard in effectiviteitsonderzoek (Hariton & Locascio, 2018). Specifiek onderzoekt de huidige studie de effectiviteit van een schoolprogramma dat beoogt een prosociaal klasklimaat te bevorderen (de Nederlandse versie van *Meaningful Roles*, zie Figuur 1 voor het conceptuele model). We verwachten 1) een verbeterd prosociaal klasklimaat (i.e., prosociaal gedrag, prosociale groepsnorm, meer samenwerking, minder conflicten), 2) een verhoogde intrinsieke prosociale motivatie, 3) vervulling van basispsychologische behoeften (i.e., autonomie, competentie, verbondenheid), 4) een verhoogd schoolwelzijn en veiligheid op school, 5) een afname van pesten en slachtofferschap van pesten, 6) een toegenomen positieve houding van leerkrachten ten opzichte van pesten (perceptie van leerlingen), 7) een toegenomen positieve houding van leerkrachten ten opzichte van het stimuleren van de vervulling van basispsychologische behoeften (gerapporteerd door leerkrachten) en 8) verbeterde

burgerschapsvaardigheden van leerlingen (gerapporteerd door leerkrachten) op scholen die de interventie implementeren in vergelijking met controlescholen.

[hier Figuur 1 plaatsen]

Daarnaast hebben we ook de effecten van programmatrouw bepaald – wanneer een programma wordt geïmplementeerd zoals bedoeld – waardoor onderzoekers accurate conclusies kunnen trekken over de effectiviteit van een programma (Ma et al., 2023; Sanetti et al., 2021). Gebrek aan effectiviteit kan namelijk voortkomen uit een slechte implementatie, en niet omdat het een inherent ineffectief programma is (Lane et al., 2004; Sanetti et al., 2021). Programma's die met een hogere getrouwheid worden geïmplementeerd, leveren over het algemeen positievere resultaten op, wat de rol van implementatiekwaliteit op programmasucces benadrukt (Durlak et al., 2011; Haataja et al., 2014).

2. Methode

2.1. Procedure en design

Vanaf januari 2023 werden meer dan 400 scholen die het KiVa-programma (antipestprogramma) implementeren (KiVa-scholen) benaderd om deel te nemen aan het huidige onderzoek, zowel persoonlijk, per email als telefonisch. We hebben ervoor gekozen om alleen KiVa-scholen te werven omdat [GEMASKEERD VOOR REVIEW] is ontwikkeld door dezelfde organisatie dat het KiVa-programma heeft ontwikkeld. Scholen werden in april 2023 gerandomiseerd op basis van schoolgrootte (aantal deelnemende klassen) met behulp van een online randomisatietool (*Research Randomizer*, 2023) in een verhouding van 1:1. Er werd een cluster RCT uitgevoerd. Na randomisatie was er één controleschool uitgevallen vanwege externe omstandigheden (schoolreorganisatie).

Vanaf mei 2023 verzamelden scholen toestemming voor deelname aan het onderzoek. Scholen informeerde ouders en leerlingen over het project via een brochure en informatiebrief, met papieren en/of digitale toestemmingsformulieren voor ouders, leerlingen (> 11 jaar) en leerkrachten/schoolleiders. Gemiddeld was de oudertoestemming 65%, wat de verwachtingen overtrof op basis van eerdere studies die gemiddelde ouderlijke toestemmingpercentages tussen de 30% en 60% rapporteerden (Tigges, 2003; Wolfenden et al., 2009).

Het design bestond uit drie dataverzamelingsmomenten. Het eerste meetmoment vond plaats in september-oktober 2023 (T1), de tweede in februari-maart 2024 (T2) en het laatste meetmoment in mei-juni 2024 (T3). Na het eerste meetmoment haakte één interventieschool af (zowel met de implementatie als met deelname aan het onderzoek), omdat ze de tijdsinvestering in combinatie met andere ontwikkelingen binnen de school hadden

onderschat. De huidige studie is geregistreerd in ClinicalTrials ([GEMASKEERD VOOR REVIEW]) en de ethische commissie van ([GEMASKEERD VOOR REVIEW]) heeft het design van de studie goedgekeurd ([GEMASKEERD VOOR REVIEW]) (voor het onderzoeksprotocol, zie [GEMASKEERD VOOR REVIEW]).

2.2. [gemaskeerd voor review]

De Nederlandse versie van Meaningful Roles ([GEMASKEERD VOOR REVIEW]) is een klassikaal programma dat wordt uitgevoerd door leerkrachten in de groepen 5-8 van het basisonderwijs (leerlingen tussen de 8-12 jaar oud). Leerkrachten van interventiescholen kregen tussen augustus en oktober 2023 twee trainingssessies, verzorgd door getrainde en ervaren onderwijsprofessionals ($N = 4$, die ieder 3-6 scholen trainde). Het eerste meetmoment vond zo vroeg mogelijk in het schooljaar plaats, vaak na de eerste leerkrachttraining (zodat de trainer de school nog kon herinneren aan de toestemmingsformulieren of andere stappen die genomen moesten worden) en vóór de tweede trainingssessie. Leerkrachten mochten pas beginnen met het implementeren van [GEMASKEERD VOOR REVIEW] nadat de vragenlijsten waren afgenomen.

2.3. Controle conditie

De controlegroep ontving geen materiaal of training met betrekking tot [GEMASKEERD VOOR REVIEW] tijdens de implementatie van het programma in de experimentele conditie (*treatment-as-usual*). Scholen in de controlegroep kregen de mogelijkheid om materiaal en de training te ontvangen nadat de dataverzameling was afgerond.

2.4. Instrumenten

Tabel 1 geeft een overzicht van de instrumenten op elk meetmoment. Informatie over prosociaal klasklimaat (prosociaal gedrag, prosociale groepsnorm, samenwerking en conflict in de klas), basispsychologische behoeften (autonomie, competentie, verbondenheid), welzijn en veiligheid op school, en antisociaal gedrag (pesten, slachtofferschap van pesten, leerkrachthouding ten opzichte van pesten) werd door leerlingen gerapporteerd. Daarnaast verstrekten leerkrachten informatie over hun houding ten opzichte van het stimuleren van de vervulling van basispsychologische behoeften en burgerschapsvaardigheden van leerlingen (zie Appendix A).

Tabel 1. Instrumenten die gebruikt zijn, met uitkomst/naam van variabele, meetmoment, type variabele en bron van variabele.

Uitkomst	Naam van variabele	Instrument	Meetmoment	Type variabele	Bron
Primaire uitkomsten					
Prosociaal gedrag	Prosociaal gedrag (individueel)	Gebaseerd op de Strengths and Difficulties Questionnaire (SDQ, van Widenfelt et al., 2003), subscaal prosociaal gedrag (4 items, $\alpha = 0.63-0.69$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
	Groepsnorm	Ontwikkeld voor deze studie (3 items, $\alpha = 0.71-0.83$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
	Samenwerking in de klas	Classroom Peer Context Questionnaire (CPCQ, Boor-Klip et al., 2014, 2016, 2017), subscaal samenwerking (4 items, $\alpha = 0.74-0.81$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
	Conflict in de klas	Classroom Peer Context Questionnaire (CPCQ, Boor-Klip et al., 2014, 2016, 2017), subscaal conflict (2 items, $\alpha = 0.71-0.76$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
Intrinsieke (prosociale) motivatie	Sociale doelen	Ontwikkeld voor deze studie (4 items, $\alpha = 0.65-0.71$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
Autonomie	Sociale autonomie en verantwoordelijkheid	Ontwikkeld voor deze studie, gebaseerd op onderzoek naar de Vreedzame School (Battistich et al., 1997; Lickona & Davidson, 2003; Pauw, 2013) en de Perceived Choice Awareness Scale (Sheldon et al., 1996; 5 items, $\alpha = 0.67-0.74$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
Competentie	Zelfvertrouwen	Ontwikkeld voor deze studie, gebaseerd op de Rosenberg Self-Esteem Scale (RSES, Franck et al., 2008; Rosenberg, 1965; 5 items, $\alpha = 0.77-0.81$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
Verbondenheid	Comfort en cohesie	Classroom Peer Context Questionnaire (CPCQ, Boor-	T1, T2, T3	Uitkomst	Leerling

Klip, 2014, 2016, 2017),
subschaal comfort en cohesie (6
items, $\alpha = 0.87-0.89$ for T1-T3)

Secundaire uitkomsten					
School welzijn	School verbondenheid en veiligheid	Gebaseerd op eerder onderzoek naar KiVa (Kärnä et al., 2011; Kaufman et al., 2021; 7 items, $\alpha = 0.90-0.91$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
School veiligheid	Veiligheid op school	Gebaseerd op eerder onderzoek naar KiVa (1 item)	T1, T2, T2	Uitkomst	Leerling
Antisociaal gedrag	Pesten en slachtofferschap van pesten	Olweus Bully-Victim Questionnaire (OBVQ, Olweus, 1996; 9 items, $\alpha = 0.89-0.88$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
	Leerkrachthouding over pesten	Gebaseerd op eerder onderzoek naar KiVa (Veenstra et al., 2014; 3 items, $\alpha = 0.35-0.42$ for T1-T3)	T1, T2, T3	Uitkomst	Leerling
Leerkrachthouding	Leerkrachthouding over stimulering van basispsychologische behoeften	Ontwikkeld voor deze studie (23 items, $\alpha = 0.92-0.93$ for T1-T3)	T1, T2, T3	Uitkomst	Leerkracht
Burgerschapsvaardigheden van leerlingen	Burgerschapsvaardigheden	Gebaseerd op de Vragenlijst Burgerschap Meten (10 items, $\alpha = 0.81-0.85$ for T1-T3)	T1, T2, T3	Uitkomst	Leerkracht
Werkzame elementen					
	Betekenisvolle rollen	Ontwikkeld voor deze studie (12 and 9 items)	T2, T3	Uitkomst / moderator	Leerling, leerkracht
	Klassenvergaderingen	Ontwikkeld voor deze studie (4 and 5 items)	T2, T3	Uitkomst / moderator	Leerling, leerkracht
	Complimenten	Ontwikkeld voor deze studie (7 and 6 items)	T2, T3	Uitkomst / moderator	Leerling, leerkracht
Moderators					
Leerling karakteristieken	Demografische gegevens	Ontwikkeld voor deze studie (9 items)	T1	Moderator / covariaat	Leerling
Leerkracht karakteristieken	Demografische gegevens	Ontwikkeld voor deze studie (7 items)	T1	Moderator / covariaat	Leerkracht
Programma karakteristieken	Programmatrouw	Ontwikkeld voor deze studie (25 items)	T2, T3	Moderator / covariaat	Leerkracht

2.5. Statistische analyses

Statistische analyses werden uitgevoerd met SPSS en Mplus. We hanteerden een *intention-to-treat* aanpak en namen alle deelnemers op in de analyses, inclusief deelnemers uit klassen die niet met het programma waren begonnen of het programma niet (goed) hadden geïmplementeerd. We volgden deze aanpak om motivatie als mogelijk *confounding* effect te verminderen en de effectiviteit van de toegewezen interventie te onderzoeken (Montori & Guyatt, 2001). Om te controleren of de randomisatie succesvol was, werden voor de analyses mogelijke baseline (T1) verschillen in demografische gegevens en uitkomstvariabelen (baseline meetmoment) tussen leerlingen in de interventie- en controleconditie gecontroleerd met onafhankelijke t-toetsen (voor continue variabelen) en χ^2 analyses (voor categorische variabelen).

Om de effectiviteit van het programma te onderzoeken, werden regressieanalyses uitgevoerd in Mplus. De afhankelijke variabelen waren de uitkomsten halverwege het programma (T2) of aan het einde van het schooljaar (T3), conditie was de voorspellende variabele (1 = interventie, 2 = controle), en baseline (T1), potentiële demografische verschillen tussen de experimentele- en controleconditie op T1, en/of T2 metingen (wanneer de afhankelijke variabelen de uitkomstmaten op T3 waren) waren de covariaten. De covariantie van conditie en de specifieke variabele(n) werden ook in de modellen opgenomen. De regressieanalyses werden geclusterd op schoolniveau en *Robust Maximum Likelihood* (MLR) werd gebruikt als *estimator*.

We hebben sensitiviteitsanalyses uitgevoerd om de impact van programmatrouw op de effectiviteit van het programma te bepalen. We hebben de regressieanalyses herhaald in en subgroep van leerlingen met een ideale programmatrouw. Om dit te bepalen, gebruikten we zelfrapportagegegevens van leerlingen over de drie belangrijkste programmacomponenten, namelijk betekenisvolle rollen, klassenbijeenkomsten en complimenten. Deze gegevens werden gecombineerd om één variabele te creëren voor elk meetmoment (T2 en T3). Voorwaarden waren dat leerlingen in de afgelopen maanden minstens één rol hadden gehad en minstens één keer per maand bezig waren met de rol(len) (T2 $N = 751$, 68%; T3 $N = 679$, 61%), minstens één klassenbijeenkomst per week hadden (T2 $N = 560$, 51%; T3 $N = 473$, 43%) en minstens één compliment per week ontvingen (T2 $N = 657$, 59%; T3 $N = 655$, 59%), zie Tabel 2 voor de beschrijvende statistieken. De subgroep omvatte alleen leerlingen die aan alle drie de voorwaarden voldeden (T2 $N = 253$, 23%; T3 $N = 202$, 18%).

Tabel 2. Beschrijvende statistieken van vragen over programma-integriteit voor leerlingen (alleen experimentele groep).

	T2		T3
	<i>N (%)</i>	<i>M (SD)</i>	<i>N (%)</i>
			<i>M (SD)</i>
Heb je de afgelopen maanden een rol gehad?		2.36 (1.12)	2.56 (1.23)
Ik had geen rol	271 (24.4)		250 (22.4)
Ik had één rol	393 (35.4)		335 (30.1)
Ik had twee rollen	288 (25.9)		290 (26.0)
Ik had drie rollen	92 (8.3)		134 (12.0)
Ik had vier of meer rollen	66 (5.9)		105 (9.4)
Hoe vaak was je bezig met de rol(len)?		2.60 (1.31)	2.61 (1.29)
Meerdere keren op een dag	180 (21.3)		162 (18.7)
Een keer per dag	264 (31.3)		297 (34.3)
Een keer per week	241 (28.6)		255 (29.4)
Een keer per maand	70 (8.3)		68 (7.8)
Minder dan een keer per maand	53 (6.3)		40 (4.6)
Nooit	36 (4.3)		44 (5.1)
Heb je met de klas vergaderingen gehad de afgelopen maanden?		3.69 (1.56)	3.54 (1.53)
Geen vergadering	179 (16.2)		178 (16.0)
Een vergadering	135 (12.2)		161 (14.5)
Een vergadering per maand	92 (8.3)		133 (12.0)
Twee vergaderingen per maand	142 (12.8)		166 (14.9)
Een vergadering per week	560 (50.5)		473 (42.6)
Ik heb complimenten gekregen van mijn klasgenoten		3.37 (1.56)	3.33 (1.52)
Meerdere keren op een dag	145 (13.1)		148 (13.3)
Een keer per dag	190 (17.2)		191 (17.2)
Een keer per week	322 (29.1)		316 (28.4)
Een keer per maand	166 (15.0)		173 (15.6)
Minder dan een keer per maand	128 (11.6)		166 (14.9)
Nooit	155 (14.0)		117 (10.5)

Regressieanalyses werden herhaald in een subgroep van leerkrachten met een ideale programmatrouw, met de volgende voorwaarden: leerkrachten gaven aan dat leerlingen de rol(len) (zeer) zelfstandig konden uitvoeren (T2 $N = 56$, 74%; T3 $N = 53$, 77%), minstens één klassenbijeenkomst per week hadden (T2 $N = 33$, 44%; T3 $N = 25$, 36%) en dat leerkrachten aangaven dat ze aandacht besteedde aan complimenten en ervoor zorgde dat complimenten minstens één keer per week werden gegeven (T2 $N = 67$, 92%; T3 $N = 63$, 93%), zie Tabel 3 voor de beschrijvende statistieken. De subgroep bestond alleen uit docenten die aan alle drie de voorwaarden voldeden (T2 $N = 25$, 34%; T3 $N = 18$, 26%).

Tabel 3. Beschrijvende statistieken van vragen over programma-integriteit voor leerkrachten (alleen experimentele groep).

	T2		T3	
	<i>N (%)</i>	<i>M (SD)</i>	<i>N (%)</i>	<i>M (SD)</i>
In hoeverre konden de leerlingen de rollen zelfstandig uitvoeren?		3.92 (0.76)		4.06 (0.73)
Helemaal niet zelfstandig	1 (1.3)		0 (0.0)	
Niet zelfstandig	0 (0.0)		0 (0.0)	
Enigszins zelfstandig	19 (25.0)		16 (23.2)	
Zelfstandig	40 (52.6)		33 (47.8)	
Zeer zelfstandig	16 (21.1)		20 (29.0)	
Hoe vaak is er een vergadering geweest?		3.97 (1.24)		3.97 (1.07)
Nooit	6 (8.0)		2 (2.9)	
Een enkele keer	5 (6.7)		5 (7.2)	
Een keer per maand	7 (9.3)		12 (17.4)	
Twee keer per maand	24 (32.0)		25 (36.2)	
Elke week	33 (44.0)		24 (34.8)	
Meerdere keren per week	0 (0.0)		1 (1.4)	
Hoe vaak per week is er aandacht geweest voor de complimenten?		2.72 (1.00)		2.57 (0.88)
Nooit	5 (6.8)		3 (4.3)	
Een keer per week	31 (41.9)		36 (52.2)	
Meerdere keren per week	26 (35.1)		21 (30.4)	
Een keer per dag	4 (5.4)		6 (8.7)	
Meerdere keren per dag	8 (10.8)		3 (4.3)	
Hoeveel complimenten werden er gegeven in de klas (ongeveer)?		2.97 (1.11)		2.97 (1.11)
Geen	3 (4.2)		3 (4.5)	
Een keer per week	24 (33.3)		21 (31.3)	
Meerdere keren per week	29 (40.3)		30 (44.8)	
Een keer per dag	4 (5.6)		1 (1.5)	
Meerdere keren per dag	12 (16.7)		12 (17.9)	

3. Resultaten

3.1. Participanten

In totaal waren 46 scholen geïnteresseerd in deelname aan het onderzoek, waarvan drie scholen vóór de randomisatie afhaakten. Scholen werden gerandomiseerd in een experimentele conditie ($N = 22$ scholen, $N = 76$ klassen) of een controleconditie ($N = 20$ scholen, $N = 78$ klassen). Van de totale steekproef haakte één controleschool af na de randomisatie en één interventieschool na het eerste meetmoment ($N = 4$ klassen, T1). Er waren geen andere uitvallers op schoolniveau op T2 of T3.

Tabel 4 toont de demografische gegevens van de leerlingen ($N = 2337$ leerlingen van 6-13 jaar, $M_{\text{leeftijd}} = 9.55$, $SD = 1.21$; sekse jongen 48%, gender 46%; sekse meisje 48%, gender 43%; andere sekse 4%, gender 12%). De meeste leerlingen voelden zich verbonden met Westerse landen ($N = 1932$, 88%; bijvoorbeeld Nederland, Indonesië, Italië of Frankrijk), anderen met niet-Westerse landen ($N = 264$, 12%; bijvoorbeeld Turkije, Marokko of Suriname). De deelname over leerjaren was ongeveer gelijk (met een iets groter aandeel in de hogere leerjaren).

Tabel 5 toont de demografische gegevens van leerkrachten ($N = 225$ leerkrachten van 21-65 jaar; $M_{\text{leeftijd}} = 38.7$, $SD = 11.1$; sekse vrouwelijk 77%, gender 76%, 1% wil het liever niet zeggen). De meeste leerkrachten voelden zich verbonden met een Westers land (97%, voornamelijk Nederland). De deelname over schooljaren was ongeveer gelijk, vergelijkbaar met de leerling rapportages. Deze studie volgde de *Consolidated Standards of Reporting Trials Guidelines* (CONSORT; Begg et al., 1996), zoals weergegeven in Figuur 2.

[Hier Figuur 2 plaatsen]

Tabel 4. Demografische gegevens op T1 en verschillen op T1 tussen de experimentele- en controleconditie voor leerlingen.

	Experimentele conditie	Controle conditie	Verschillen op T1
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>
Leeftijd (jaren) ^a	9.59 (1.20)	9.51 (1.22)	1.643 (.101)
	<i>N (%)</i>	<i>N (%)</i>	$\chi^2 (p)$
Sekse			2.258 (.323)
Jongen	547 (47.6)	534 (49.3)	
Meisje	553 (48.1)	516 (47.6)	
Anders of weet ik niet	49 (4.3)	34 (3.1)	
Gender			1.360 (.507)
Jongen	524 (45.6)	503 (46.4)	
Meisje	500 (43.5)	449 (41.5)	
Anders ^b	125 (10.9)	131 (12.1)	
Etnische identiteit			0.014 (.905)
Westers	995 (87.9)	937 (88.1)	
Mix Westers-niet	137 (12.1)	127 (11.9)	
Westers – Niet-Westers			
Groep/leerjaar ^c			5.197 (.158)
5	210 (18.3)	237 (21.9)	
6	287 (25.0)	276 (25.5)	
7	324 (28.3)	286 (26.4)	
8	325 (28.4)	284 (26.2)	

Noot: ^a: $N = 2229$ leerlingen rapporteerden leeftijd ($N = 1147$ experimentele conditie, $N = 1082$ controle conditie). ^b: Categorieën zijn anders, jongen en meisje, geen meisje of jongen, weet ik niet of wil ik niet zeggen. ^c: $N = 3$ leerlingen rapporteerde dat ze in groep/leerjaar 4 zitten, allen afkomstig uit de experimentele conditie. Deze leerlingen werden niet in de analyses opgenomen.

Tabel 5. Demografische en descriptieve gegevens op T1, en verschillen op T1 tussen de experimentele- en controleconditie voor leerkrachten.

	Experimentele conditie	Controle conditie	Verschillen op T1
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>
Leeftijd (jaren)	39.38 (10.75)	37.95 (11.45)	0.905 (.366)
	<i>N (%)</i>	<i>N (%)</i>	$\chi^2 (p)$
Sekse			11.605 (.003)
Man	11 (12.5)	33 (32.7)	
Vrouw	76 (86.4)	68 (67.3)	
Wil ik liever niet zeggen	1 (1.1)	0 (0.0)	
Gender			10.972 (.004)
Man	11 (12.5)	32 (31.7)	
Vrouw	77 (87.5)	68 (67.3)	
Wil ik liever niet zeggen	0 (0.0)	1 (1.0)	
Etnische identiteit			2.626 (.105)
Westers	87 (100.0)	98 (97.0)	
Mix Westers-niet Westers	0 (0.0)	3 (3.0)	
Groep/leerjaar			4.125 (.127)
5-6	51 (44.3)	48 (43.6)	
7-8	58 (50.4)	48 (43.6)	
Anders (e.g., 5-8, 6-8, 6-7)	6 (5.2)	14 (12.7)	

Noot: Vetgedrukte statistieken geven significante resultaten weer.

3.2. Randomisatie en dropouts

Wat betreft de zelfrapportage van leerlingen waren er op T1 geen verschillen tussen de experimentele- en de controleconditie voor demografische gegevens (zie Tabel 4). Voor de studievariabelen waren er op T1 verschillen tussen de experimentele- en controleconditie wat betreft prosociaal gedrag (individueel en samenwerking in de klas), veiligheid op school, leerkrachthouding ten opzichte van pesten, en slachtofferschap van pesten (Tabel 6). De controleconditie scoorde hoger op prosociaal gedrag en slachtofferschap van pesten, terwijl de experimentele conditie hoger scoorde op veiligheid op school en leerkrachthouding ten opzichte van pesten. Er waren enkele verschillen op T1 tussen deelnemers en dropouts op T2 en T3. Dropouts hadden vaker een niet-Westerse etnische identiteit, kwamen vaker uit lagere leerjaren en rapporteerden lagere niveaus van samenwerking in de klas (Tabel 7 en 8). Er waren geen andere verschillen op T1 tussen de experimentele- en controleconditie en tussen deelnemers en dropouts met betrekking tot demografische gegevens en studievariabelen.

Wat betreft de zelfrapportage van leerkrachten waren er op T1 verschillen tussen de experimentele- en controleconditie voor geslacht en gender (zie Tabel 5). Er waren minder vrouwen in de controleconditie (zowel voor geslacht als gender). Er waren op T1 geen verschillen tussen deelnemers en dropouts voor de studievariabelen op T2 en T3, alleen wat betreft geslacht en conditie. Dropouts waren vaker mannen en kwamen uit de experimentele conditie (Tabel 9).

Tabel 6. Descriptieve statistieken, verschillen op T1 tussen de experimentele en controlegroep, en regressieanalyses voor leerlingen.

	Experimentale conditie ^a			Controle conditie ^a			Verschillen op T1		Regressie		Regressie	
	T1	T2	T3	T1	T2	T3			Uitkomst op T2		Uitkomst op T3	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>
Prosociaal gedrag (individueel)	2.80 (0.51)	2.85 (0.50)	2.84 (0.52)	2.88 (0.51)	2.86 (0.52)	2.88 (0.53)	-3.453 (<.001)	-0.15	-0.025 (.357)	0.14	0.017 (.479)	0.08
Prosociaal gedrag: injunctieve norm	2.48 (0.49)	2.41 (0.53)	2.36 (0.57)	2.50 (0.51)	2.40 (0.53)	2.37 (0.58)	-0.879 (.379)	-0.04	-0.019 (.562)	0.06	0.010 (.682)	0.02
Prosociaal gedrag: samenwerking	3.92 (0.59)	3.88 (0.59)	3.87 (0.65)	3.98 (0.62)	3.92 (0.64)	3.95 (0.63)	-2.422 (.016)	-0.10	0.007 (.859)	0.03	0.027 (.193)	-0.03
Prosociaal gedrag: conflict	2.11 (0.85)	2.17 (0.86)	2.18 (0.84)	2.17 (0.87)	2.16 (0.86)	2.13 (0.86)	-1.787 (.074)	-0.08	-0.023 (.550)	0.08	-0.032 (.243)	0.13
Intrinsieke (prosociale) motivatie	2.65 (0.38)	2.65 (0.38)	2.64 (0.38)	2.66 (0.38)	2.64 (0.39)	2.63 (0.42)	-0.795 (.427)	-0.03	-0.027 (.348)	0.05	-0.012 (.636)	0.05
Sociale autonomie	2.96 (0.52)	2.94 (0.53)	2.93 (0.55)	2.94 (0.54)	2.92 (0.54)	2.94 (0.55)	0.997 (.319)	0.04	-0.005 (.871)	0.00	0.021 (.396)	-0.06
Sociale competentie	2.59 (0.42)	2.59 (0.43)	2.59 (0.43)	2.57 (0.42)	2.58 (0.43)	2.57 (0.45)	1.112 (.266)	0.05	0.010 (.680)	-0.02	-0.012 (.513)	0.00
Sociale verbondenheid	4.14 (0.72)	4.05 (0.75)	4.04 (0.76)	4.13 (0.75)	4.06 (0.79)	4.08 (0.78)	0.315 (.753)	0.01	0.015 (.582)	-0.03	0.015 (.308)	-0.07
Schoolwelzijn	3.18 (0.60)	3.12 (0.62)	3.11 (0.63)	3.15 (0.65)	3.10 (0.65)	3.11 (0.64)	1.235 (.217)	0.05	0.002 (.916)	-0.02	0.012 (.546)	-0.05
Veiligheid op school	4.06 (0.89)	4.08 (0.84)	4.08 (0.84)	3.98 (0.97)	4.01 (0.92)	4.05 (0.88)	2.024 (.043)	0.09	-0.021 (.380)	-0.01	0.005 (.883)	-0.05
Pestgedrag	1.19 (0.61)	1.18 (0.58)	1.18 (0.59)	1.24 (0.68)	1.20 (0.65)	1.18 (0.59)	-1.778 (.075)	-0.08	0.011 (.553)	0.05	-0.017 (.430)	0.08
Slachtofferschap van pesten	1.46 (0.99)	1.47 (0.95)	1.39 (0.84)	1.62 (1.10)	1.56 (1.06)	1.47 (0.95)	-3.622 (<.001)	-0.15	0.007 (.748)	0.07	0.006 (.737)	0.08
Slachtofferschap van pesten (max)	1.78 (1.19)	1.77 (1.17)	1.62 (1.06)	1.96 (1.28)	1.87 (1.26)	1.78 (1.19)	-3.382 (<.001)	-0.14	0.002 (.932)	0.07	0.038 (.077)	0.02
Leerkrachthouding over pesten	3.54 (0.68)	3.56 (0.65)	3.48 (0.69)	3.45 (0.73)	3.54 (0.68)	3.42 (0.74)	3.093 (.002)	0.13	0.010 (.803)	-0.10	-0.024 (.503)	-0.04

Noot: ^aConditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdstipmoment (experimentele groep *N* = 1113-1160; controle groep *N* = 1086-1105).

Tabel 7. Verschillen op T1 voor demografische gegevens tussen deelnemers en dropouts T2 en T3 voor leerlingen.

	Deelnemers	Dropouts T2 en T3	Verschillen op T1
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>
Leeftijd (jaren)	9.56 (1.20)	9.28 (1.49)	-1.567 (.121)
	<i>N (%)</i>	<i>N (%)</i>	$\chi^2 (p)$
Sekse			1.574 (.455)
Jongen	1051 (48.6)	30 (42.3)	
Meisje	1032 (47.7)	37 (52.1)	
Anders of weet ik niet	79 (3.7)	4 (5.6)	
Gender			0.867 (.648)
Jongen	995 (46.0)	32 (45.1)	
Meisje	916 (42.4)	33 (46.5)	
Anders ^a	250 (11.6)	6 (8.5)	
Etnische identiteit			66.648 (< .001)
Westers	1893 (89.0)	39 (56.5)	
Mix Westers-niet	234 (11.0)	30 (43.5)	
Westers – non-Western			
Groep/leerjaar			9.438 (.024)
5	425 (19.7)	22 (32.4)	
6	545 (24.5)	18 (26.5)	
7	592 (27.4)	18 (26.5)	
8	599 (27.7)	10 (14.7)	

Noot. ^a: Categorieën zijn anders, jongen en meisje, geen meisje of jongen, weet ik niet of wil ik niet zeggen. Vetgedrukte statistieken geven significante resultaten weer.

Tabel 8. Verschillen op T1 voor studieve variabelen tussen deelnemers en dropouts op T2 en T3 voor leerlingen.

	Deelnemers	Dropouts op T2 en T3	Verschillen op T1
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>
Prosociaal gedrag (individueel)	2.84 (0.51)	2.73 (0.58)	-1.910 (.056)
Prosociaal gedrag: injunctieve norm	2.49 (0.50)	2.51 (0.43)	0.291 (.771)
Prosociaal gedrag: samenwerking	3.96 (0.61)	3.80 (0.59)	-2.189 (.029)
Prosociaal gedrag: conflict	2.14 (0.86)	2.11 (0.84)	-0.274 (.392)
Intrinsieke (prosociale) motivatie	2.66 (0.38)	2.65 (0.41)	-0.263 (.792)
Sociale autonomie	2.95 (0.53)	2.85 (0.49)	-1.714 (.115)
Sociale competentie	2.59 (0.42)	2.51 (0.51)	-1.349 (.181)
Sociale verbondenheid	4.13 (0.74)	4.10 (0.73)	-0.390 (.697)
Schoolwelzijn	3.16 (0.63)	3.30 (0.59)	1.934 (.053)
Veiligheid op school	4.02 (0.93)	3.99 (0.95)	-0.304 (.761)
Pestgedrag	1.21 (0.64)	1.37 (0.83)	1.613 (.111)
Slachtofferschap van pesten	1.53 (1.04)	1.61 (1.20)	0.653 (.514)
Slachtofferschap van pesten (max)	1.86 (1.24)	2.04 (1.40)	1.209 (.227)
Leerkrachthouding over pesten	3.50 (0.71)	3.55 (0.77)	0.628 (.530)

Noot. Vetgedrukte statistieken geven significante resultaten weer.

Tabel 9. Verschillen op T1 tussen deelnemers en dropouts op T2 en T3 voor leerkrachten.

	Deelnemers	Dropouts op T2 en T3	Verschillen op T1
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>
Leeftijd (jaren)	38.87 (10.88)	37.95 (12.02)	-0.461 (.645)
Leerkrachthouding	8.05 (0.69)	8.04 (0.63)	-0.039 (.969)
Burgerschapsvaardigheden	3.03 (0.37)	2.96 (0.35)	-0.776 (.439)
	<i>N (%)</i>	<i>N (%)</i>	$\chi^2 (p)$
Sekse			12.617 (.002)
Man	34 (20.6)	10 (41.7)	
Vrouw	131 (79.4)	13 (54.2)	
Wil ik liever niet zeggen	0 (0.0)	1 (4.2)	
Gender			5.680 (.058)
Man	33 (20.0)	10 (41.7)	
Vrouw	131 (79.4)	14 (58.3)	
Wil ik liever niet zeggen	1 (0.6)	0 (0.0)	
Etnische identiteit			1.264 (.261)
Westers	163 (98.8)	22 (95.7)	
Mix Western-non-Western	2 (1.2)	1 (33.3)	
Groep/leerjaar			3.198 (.202)
5-6	86 (46.7)	13 (31.7)	
7-8	83 (45.1)	23 (56.1)	
Anders (e.g., 5-8, 6-8, 6-7)	15 (8.2)	5 (12.2)	
Conditie			7.724 (.005)
Experimentele conditie	86 (46.7)	29 (70.7)	
Controle conditie	98 (53.3)	12 (29.3)	

Noot: Vetgedrukte statistieken geven significante resultaten weer.

3.3. Programma effectiviteit in de volledige steekproef

Ten eerste werd onderzocht in hoeverre er effecten waren in de volledige steekproef, ongeacht de mate van programmatrouw. Er waren geen significante effecten op de uitkomstmaten (zie Tabel 6).

Tabel 10 presenteert de resultaten van de regressieanalyses voor leerkrachten. De resultaten toonden aan dat er geen significante effecten waren voor de leerkrachthouding ten opzichte van de stimulering van het vervullen van basispsychologische behoeften en burgerschapsvaardigheden van leerlingen.

3.4. Programma effectiviteit voor ideale programmatrouw

Wanneer de analyses werden uitgevoerd met een subgroep van leerlingen die ervaarden dat het programma werd geïmplementeerd zoals bedoeld, toonde het programma significante positieve effecten op T2 voor prosociaal gedrag (individueel en samenwerking in de klas), intrinsieke (prosociale) motivatie, autonomie, verbondenheid, schoolwelzijn en veiligheid op school (Tabel 11). Significante positieve effecten werden gevonden op T3 voor competentie, slachtofferschap van pesten en leerkrachthouding over pesten (Tabel 12). De effecten voor alle uitkomsten waren klein (Cohen, 1988).

Wat betreft de zelfrapportage van leerkrachten, lieten leerkrachten die aangaven dat het programma werd geïmplementeerd zoals bedoeld positieve effecten zien op de houding ten opzichte van stimulering van het vervullen van basispsychologische behoeften op T2 (Tabel 13) en burgerschapsvaardigheden van leerlingen op T3 (Tabel 14). De effecten voor beide uitkomsten waren klein (Cohen, 1988).

Tabel 10. Verschillen op T1 tussen de experimentele- en controleconditie, en regressieanalyses op T2 en T3 voor leerkrachten.

	Experimentele conditie ^a			Controle conditie ^a			Verschillen op T1		Regressie		Regressie	
	T1	T2	T3	T1	T2	T3			Uitkomst op T2		Uitkomst op T3	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>
Leerkrachthouding	7.92 (0.62)	8.01 (0.70)	8.26 (0.64)	8.19 (0.71)	8.14 (0.60)	8.23 (0.63)	-2.681 (.008)	0.41	-0.021 (.782)	0.21	-0.102 (.078)	0.45
Burgerschapsvaardigheden	3.03 (0.39)	3.06 (0.41)	3.09 (0.39)	3.01 (0.35)	3.05 (0.33)	3.03 (0.36)	0.319 (.750)	-0.05	-0.009 (.895)	-0.03	-0.093 (.101)	0.11

Noot. Conditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Gender is geïncludeerd als covariaat in alle analyses. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdsmoment (experimentele groep *N* = 71-97, controle groep *N* = 77-86).

Tabel 11. Sensitiviteitsanalyses met ideale programmatrouw: Descriptieve statistieken, verschillen op T1 tussen de experimentele- en controleconditie, en regressieanalyses op T2 voor leerlingen.

	Experimentele conditie ^a		Controle conditie ^a		Verschillen op T1	Regressie		
	T1	T2	T1	T2		Uitkomst op T2		
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>	<i>d</i>	β (<i>p</i>)	<i>d</i>
Prosociaal gedrag (individueel)	2.96 (0.46)	3.02 (0.44)	2.88 (0.51)	2.86 (0.52)	2.626 (.009)	0.17	-0.084 (.005)	0.16
Prosociaal gedrag: injunctieve norm	2.58 (0.44)	2.51 (0.53)	2.50 (0.51)	2.40 (0.53)	2.501 (.013)	0.16	-0.051 (.148)	0.06
Prosociaal gedrag: samenwerking	4.07 (0.53)	4.08 (0.52)	3.98 (0.62)	3.92 (0.64)	2.221 (.027)	0.14	-0.074 (.005)	0.12
Prosociaal gedrag: conflict	1.98 (0.79)	1.94 (0.78)	2.17 (0.87)	2.16 (0.86)	-3.388 (< .001)	-0.22	0.059 (.081)	-0.04
Intrinsieke (prosociale) motivatie	2.71 (0.33)	2.76 (0.31)	2.66 (0.38)	2.64 (0.39)	2.042 (.042)	0.13	-0.092 (< .000)	0.19
Sociale autonomie	3.09 (0.47)	3.14 (0.47)	2.94 (0.54)	2.92 (0.54)	4.614 (< .001)	0.30	-0.104 (< .000)	0.16
Sociale competentie	2.67 (0.36)	2.68 (0.34)	2.57 (0.42)	2.58 (0.43)	3.781 (< .001)	0.24	-0.039 (.090)	0.00
Sociale verbondenheid	4.32 (0.58)	4.30 (0.59)	4.13 (0.75)	4.06 (0.79)	4.586 (< .001)	0.27	-0.055 (.019)	0.07
Schoolwelzijn	3.40 (0.47)	3.34 (0.52)	3.15 (0.65)	3.10 (0.65)	7.130 (< .001)	0.41	-0.045 (.017)	-0.02
Veiligheid op school	4.16 (0.87)	4.26 (0.71)	3.98 (0.97)	4.01 (0.92)	2.722 (.007)	0.19	-0.075 (.005)	0.07
Pestgedrag	1.05 (0.24)	1.13 (0.51)	1.24 (0.68)	1.20 (0.65)	-7.479 (< .001)	-0.31	0.004 (.813)	0.19
Slachtofferschap van pesten	1.23 (0.79)	1.31 (0.80)	1.62 (1.10)	1.56 (1.06)	-6.437 (< .001)	-0.37	0.024 (.246)	0.13
Slachtofferschap van pesten (max)	1.52 (0.98)	1.55 (0.97)	1.96 (1.28)	1.87 (1.26)	-6.033 (< .001)	-0.36	0.033 (.256)	0.10
Leerkrachthouding over pesten	3.61 (0.61)	3.65 (0.57)	3.45 (0.73)	3.54 (0.68)	3.708 (< .001)	0.23	-0.042 (.433)	-0.07

Noot: ^aConditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdsmoment (experimentele groep $N = 252-253$; controle groep $N = 1096-1104$).

Tabel 12. Sensitiviteitsanalyses met ideale programmatrouw: Descriptieve statistieken, verschillen op T1 tussen de experimentele- en controleconditie, en regressieanalyses op T3 voor leerlingen.

	Experimentele conditie ^a			Controle conditie ^a			Verschillen op T1		Regressie	
	T1	T2	T3	T1	T2	T3			Uitkomst op T3	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>t (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>
Prosociaal gedrag (individueel)	2.91 (0.49)	2.98 (0.49)	2.98 (0.47)	2.88 (0.51)	2.86 (0.52)	2.88 (0.53)	0.721 (.471)	0.06	-0.017 (.571)	0.14
Prosociaal gedrag: injunctieve norm	2.58 (0.43)	2.51 (0.51)	2.49 (0.52)	2.50 (0.51)	2.40 (0.53)	2.37 (0.58)	2.154 (.032)	0.15	-0.028 (.184)	0.08
Prosociaal gedrag: samenwerking	4.08 (0.53)	4.06 (0.52)	4.05 (0.56)	3.98 (0.62)	3.92 (0.64)	3.95 (0.63)	2.289 (.023)	0.16	-0.007 (.798)	0.00
Prosociaal gedrag: conflict	2.00 (0.84)	2.03 (0.85)	2.04 (0.86)	2.17 (0.87)	2.16 (0.86)	2.13 (0.86)	-2.529 (0.012)	- 0.20	0.002 (.961)	0.09
Intrinsieke (prosociale) motivatie	2.70 (0.31)	2.71 (0.34)	2.69 (0.33)	2.66 (0.38)	2.64 (0.39)	2.63 (0.42)	1.411 (.159)	0.09	-0.016 (.599)	0.05
Sociale autonomie	3.05 (0.49)	3.07 (0.51)	3.10 (0.49)	2.94 (0.54)	2.92 (0.54)	2.94 (0.55)	2.814 (.005)	0.22	-0.047 (.129)	0.09
Sociale competentie	2.64 (0.37)	2.65 (0.35)	2.68 (0.36)	2.57 (0.42)	2.58 (0.43)	2.57 (0.45)	2.166 (.031)	0.17	-0.045 (.012)	0.10
Sociale verbondenheid	4.31 (0.58)	4.23 (0.66)	4.26 (0.63)	4.13 (0.75)	4.06 (0.79)	4.08 (0.78)	3.959 (< .001)	0.26	-0.026 (.073)	0.00

Schoolwelzijn	3.34 (0.51)	3.27 (0.60)	3.27 (0.56)	3.15 (0.65)	3.10 (0.65)	3.11 (0.64)	4.477 (< .001)	0.30	-0.018 (.408)	-0.05
Veiligheid op school	4.18 (0.88)	4.18 (0.72)	4.18 (0.76)	3.98 (0.97)	4.01 (0.92)	4.05 (0.87)	2.718 (0.007)	0.21	-0.016 (.576)	-0.07
Pestgedrag	1.17 (0.58)	1.14 (0.52)	1.14 (0.52)	1.24 (0.68)	1.20 (0.65)	1.18 (0.59)	-1.581 (.115)	- 0.11	0.004 (.857)	0.05
Slachtofferschap van pesten	1.27 (0.71)	1.37 (0.83)	1.31 (0.74)	1.62 (1.10)	1.56 (1.06)	1.47 (0.95)	-5.778 (< .001)	- 0.34	0.008 (.644)	0.18
Slachtofferschap van pesten (max)	1.61 (1.02)	1.66 (1.05)	1.48 (0.85)	1.96 (1.28)	1.87 (1.26)	1.78 (1.19)	-3.452 (< .001)	- 0.28	0.041 (.020)	0.04
Leerkrachthouding over pesten	3.63 (0.67)	3.64 (0.62)	3.64 (0.64)	3.45 (0.73)	3.54 (0.68)	3.42 (0.75)	3.452 (< .001)	0.25	-0.072 (.013)	0.06

Noot: ^aConditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdsmoment (experimentele groep $N = 193-202$; controle groep $N = 1086-1105$).

Tabel 13. Sensitiviteitsanalyses met ideale programmatrouw: Verschillen op T1 tussen de experimentele- en controleconditie, en regressieanalyses op T2 voor leerkrachten.

	Experimentele conditie ^a			Controle conditie ^a			Verschillen op T1		Regressie	
	T1	T2	T3	T1	T2	T3			Uitkomst T2	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>
Leerkrachthouding	7.84 (0.71)	8.13 (0.62)	8.33 (0.67)	8.19 (0.71)	8.14 (0.60)	8.23 (0.63)	-2.028 (.045)	0.49	-0.110 (.039)	0.48
Burgerschapsvaardigheden	3.10 (0.36)	3.18 (0.37)	3.19 (0.41)	3.01 (0.35)	3.05 (0.33)	3.03 (0.36)	1.089 (.279)	- 0.26	-0.091 (.329)	0.11

Noot. ^aConditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Gender is geïncludeerd als covariaat in alle analyses. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdstip (experimentele groep *N* = 22-25, controle groep *N* = 77-86).

Tabel 14. Sensitiviteitsanalyses met ideale programmatrouw: Verschillen op T1 tussen de experimentele- en controleconditie, en regressieanalyses op T3 voor leerkrachten.

	Experimentele conditie ^a			Controle conditie ^a			Verschillen op T1		Regressie	
	T1	T2	T3	T1	T2	T3			Uitkomsten op T3	
	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>F (p)</i>	<i>d</i>	<i>β (p)</i>	<i>d</i>
Leerkrachthouding	7.85 (0.47)	8.21 (0.60)	8.36 (0.55)	8.19 (0.71)	8.14 (0.60)	8.23 (0.63)	-1.814 (.073)	0.50	-0.063 (.310)	0.6

Burgerschapsvaardighede	3.02 (0.40)	3.09 (0.47)	3.15 (0.43)	3.01 (0.35)	3.05 (0.33)	3.03 (0.36)	0.063 (.950)	-	-0.105 (.035)	0.3
n								0.03		1

Noot. ^aConditie is gecodeerd als experimentele groep = 1 en controle groep = 2. Gender is geïncludeerd als covariaat in alle analyses. Vetgedrukte statistieken geven significante resultaten weer. Aantal deelnemers varieert per variabele en tijdsmoment (experimentele groep $N = 16-18$, controle groep $N = 77-86$).

4. Discussie

Een positief, prosociaal klasklimaat is van cruciaal belang om een effectieve leeromgeving te optimaliseren en het school- en psychologische welzijn van kinderen te bevorderen (Evans et al., 2009; Wang et al., 2020). Schoolinterventies kunnen een ideale aanpak zijn om dit te bereiken, maar er is een gebrek aan kennis over de effectiviteit van programma's gebaseerd op de theorie van de vervulling van basispsychologische behoeften en (cluster) RCT designs om dergelijke effecten te evalueren. Om deze hiaten aan te pakken, onderzocht de huidige studie de effectiviteit van [GEMASKEERD VOOR REVIEW] – een schoolinterventie gericht om een prosociaal klasklimaat in basisscholen te bevorderen door de intrinsieke prosociale motivatie van leerlingen te versterken door het vervullen van de drie basispsychologische behoeften (autonomie, competentie, en verbondenheid). Onze bevindingen toonden aan dat leerlingen die het programma volgden, en het programma werd geïmplementeerd zoals bedoeld, beter af waren dan leerlingen in controlescholen met betrekking tot een prosociaal klasklimaat, basispsychologische behoeften, welzijn en veiligheid op school, en antisociaal gedrag. Bovendien rapporteerden leerkrachten een positievere houding ten opzichte van stimulering van het vervullen van basisbehoeften en betere burgerschapsvaardigheden van leerlingen.

Centraal in deze bevindingen is het empirisch bewijs voor theoretische suggesties dat het ondersteunen van de autonomie, competentie, en verbondenheid van leerlingen bijdraagt aan een prosociaal klasklimaat. Eerdere studies hebben al empirische ondersteuning voor deze aanname geleverd (Cheon et al., 2019; Ntoumanis et al., 2021), al werden deze studies voornamelijk uitgevoerd onder volwassenen of middelbare scholieren en waren het interventies gericht op leerkrachten lichamelijke opvoeding. De huidige studie breidt deze kennis uit door aan te tonen dat het vervullen van basispsychologische behoeften ter bevordering van prosociale klassen ook effectief kan zijn bij jongere leeftijdsgroepen en binnen een breder scala aan interventies. Deze bijdrage wordt verder versterkt door het gebruik van een cluster RCT design en een grote steekproef van leerlingen en leerkrachten.

Daarnaast benadrukken onze bevindingen de cruciale rol van programmatrouw bij het interpreteren van programma effectiviteit. Dit komt overeen met eerder onderzoek naar programma's die gericht zijn op het bevorderen van sociaal en emotioneel leren en het verminderen van antisociaal gedrag (Durlak et al., 2011; Haataja et al., 2014). Onze resultaten

wijzen erop dat dit ook geldt voor programma's die zich richten op basispsychologische behoeften; om prosociale klassen te bevorderen, moeten alle componenten frequent en consistent worden uitgevoerd. In het geval van [GEMASKEERD VOOR REVIEW] betekent dit dat leerlingen verantwoordelijkheid en autonomie krijgen door middel van zinvolle opdrachten (klassenrollen), dat klassenbijeenkomsten worden gefaciliteerd om de groepsdynamiek- en sfeer te bespreken, en dat leerlingen worden aangemoedigd elkaars sterke punten en positief gedrag te herkennen en waarderen door middel van gestructureerde complimenten.

Wanneer leerkrachten zich aan het programma hielden als bedoeld, rapporteerden ze een toename in burgerschapsvaardigheden bij leerlingen. Burgerschapsvaardigheden zijn sociale en maatschappelijke vaardigheden die leerlingen helpen actief bij te dragen aan en deel te nemen aan de maatschappij (Kirlin, 2005). Het gaat hierbij om het vergaren van kennis en vaardigheden en het toepassen daarvan in het dagelijks leven, bijvoorbeeld door te leren over democratie (vrijheid, gelijkheid, diversiteit, solidariteit) of door te leren hoe je op respectvolle wijze omgaat met verschillen tussen mensen. Burgerschapsvaardigheden worden geassocieerd met een hogere mate van maatschappelijke verantwoordelijkheid en betrokkenheid (Lenzi et al., 2014; Lin, 2015). Onze bevindingen benadrukken dat schoolprogramma's gericht op het vervullen van basispsychologische behoeften en het creëren van prosociale klassen ook positieve effecten kunnen hebben op de burgerschapsvaardigheden van leerlingen, waardoor een democratisch schoolklimaat ontstaat.

Opvallend werden de meeste effecten op korte termijn gevonden, halverwege het programma. Dit kan te wijten zijn aan motivatieproblemen tijdens de implementatie, moeite om aandacht te blijven geven aan de componenten, of druk van andere taken die in de tweede helft van het schooljaar verplicht zijn, zoals toetsen en vakanties – zelfs in de subgroep met hoge programmatrouw. In de eerste helft van het jaar waren leerkrachten en leerlingen waarschijnlijk het meest gemotiveerd en besteedden ze de meeste aandacht aan de implementatie, omdat ze recent de training hadden gevolgd. Dit komt overeen met de gerapporteerde programmatrouw, waarbij minder leerlingen aangaven rollen te hebben gehad en bezig te zijn geweest met de rol(len) (T2 68%, T3 61%) en minder leerkrachten en leerlingen aangaven wekelijkse klassenvergaderingen te hebben gehad (leerlingen: T2 51%, T3 45%, leerkrachten: T2 44%, T3 36%) op T3 vergeleken met T2. Aan de andere kant hadden de meeste effecten op T3 betrekking op pesten (slachtofferschap, leerkrachthouding), wat niet de primaire focus van het programma

was. Effecten op negatief, hardnekkig gedrag zoals pesten kunnen langer duren om te ontwikkelen, aangezien deze processen zich voornamelijk ontwikkelen via complexe groepsprocessen (Thornberg, 2011). Het zou kunnen dat langere blootstelling aan het programma voornamelijk verband houdt met effecten op secundaire uitkomsten. Ten slotte is het ook mogelijk dat er minder effecten op T3 werden gevonden vanwege een lagere power ($N = 202$ op T3 vergeleken met $N = 253$ op T2).

4.1. Beperkingen en sterke punten

Er zijn verschillende beperkingen van de huidige studie waarmee rekening gehouden moet worden. Hoewel we een hogere responsepercentage hadden dan in eerder onderzoek (Tigges, 2003; Wolfenden et al., 2009), nam ongeveer een derde van de in aanmerking komende steekproef niet deel aan het onderzoek. Eerdere studies lieten *biased* steekproeven zien bij het gebruik van actieve toestemming in vergelijking met passieve toestemming, zoals meer vrouwelijke en jongere deelnemers, minder etnisch diverse deelnemers en minder probleemgedrag (Courser et al., 2009; Liu et al., 2017; Pokorny et al., 2001). Daarnaast waren er verschillen op baseline tussen de experimentele- en controleconditie en tussen deelnemers en dropouts, zowel onder leerlingen als onder leerkrachten. Hoewel we in de analyses rekening hielden met de verschillen op baseline tussen de experimentele- en controleconditie, suggereren deze verschillen dat de randomisatie niet geheel succesvol was. Ondanks de mogelijk bias nam nog steeds een groot aantal leerlingen deel aan het onderzoek (>2000), wat voldoende power biedt om de effecten van het programma te toetsen. Bovendien waren de deelnemende interventie- en controlescholen verspreid over heel Nederland, zowel in kleinere dorpen als steden, wat suggereert dat de steekproef representatief is voor een groot deel van de Nederlandse bevolking.

Methodologische beperkingen waren ons instrument voor competentie en de relatief lage power voor zowel leerling- als leerkrachtanalyses voor programmatrouw. De competentievragenlijst richt zich vooral op (algemene) zelfwaardering en zelfvertrouwen en minder op klasprocessen, wat de andere instrumenten voor basispsychologische behoeften wel beoordeelden. Dit verklaart mogelijk waarom we pas aan het einde van het jaar (T3) positieve effecten op competentie zagen, en (nog) niet halverwege het schooljaar (op T2). De lage power geeft aan dat slechts een klein deel van de leerlingen en leerkrachten ervoeren dat het

programma was geïmplementeerd zoals bedoeld, ongeveer een kwart van alle deelnemers. Daarnaast kunnen we minder uitspraken doen over de effectiviteit, met name voor leerkrachten.

4.2. Praktische implicaties en suggesties voor vervolgonderzoek

Onze bevindingen hebben belangrijke implicaties voor leerkrachten en andere belanghebbenden. Ten eerste lieten onze resultaten zien dat strategieën om autonomie, competentie, en verbondenheid te stimuleren effectief kunnen zijn bij het creëren van prosociale klassen. Daarom wordt het aanbevolen dat leerkrachten op een basisbehoeften-stimulerende manier werken die de basisbehoeften van leerlingen versterkt, bijvoorbeeld door leerlingen meer autonomie en verantwoordelijkheid te geven bij taken of opdrachten (i.e., betekenisvolle rollen). Eerder onderzoek heeft aangetoond dat het toepassen van een meer autonomie-ondersteunende motiverende stijl in het onderwijs van leerkrachten positieve effecten heeft voor zowel leerkrachten als leerlingen, waaronder meer werkplezier en meer betrokkenheid van leerlingen in de klas (Cheon et al., 2019).

Ten tweede lijkt programmatrouw belangrijk voor optimale programma-effectiviteit. Onze resultaten toonden aan dat het programma alleen effectief was in een subgroep van leerlingen en leerkrachten die een ideale programmatrouw rapporteren. Programmatrouw wordt door weinig studies gerapporteerd (Bruhn et al., 2015; Harn et al., 2013; Loon et al., 2020), terwijl het essentieel is om correcte conclusies te trekken over de effectiviteit van een programma. Ontwikkelaars van interventieprogramma's en onderzoekers die de effectiviteit van dergelijke programma's onderzoeken moeten zich hiervan bewust zijn en worden geadviseerd zich te richten op correcte implementatie van het protocol. Om de implementatie van schoolprogramma's te verbeteren, adviseerde eerder onderzoek om positieve relaties tussen externen en schoolpersoneel te creëren en te onderhouden, het programma aantrekkelijk te maken voor gebruikers, voldoende zichtbaar te zijn en voldoende trainingsmogelijkheden te bieden aan de uitvoerders van het programma (Gee et al., 2021; Kaufman & Huitsing, 2019; Loon et al., 2024; Wagner et al., 2004). Daarnaast werden de meeste programma-effecten waargenomen op de korte termijn, halverwege het programma. Dit geeft aan dat het belangrijk is dat aanbieders betrokken zijn bij de implementatie gedurende het gehele implementatieproces (of langer), niet alleen aan het begin.

Tot slot is verder onderzoek naar [GEMASKEERD VOOR REVIEW] nodig om beter te begrijpen hoe het programma optimaal kan worden geïmplementeerd en gewaarborgd. Een opvallende constatering was dat de drie hoofdcomponenten afzonderlijk redelijk goed werden geïmplementeerd (de gerapporteerde ideale programmatrouw was respectievelijk 43-68% en 36-93% voor leerlingen en leerkrachten), maar de combinatie van alle drie niet (18-23% voor leerlingen, 26-34% voor leerkrachten). Het is daarom belangrijk om te onderzoeken hoe de programmacomponenten worden ingezet (afgezien van de frequentie, zoals onderzocht is in deze studie), wat de werkingsmechanismen zijn, welke ondersteuningsbronnen scholen nodig hebben en welke scholen het meeste profiteren van deze ondersteuning (Gee et al., 2021; Kaufman & Huitsing, 2019). Het is ook waardevol om te onderzoeken of componenten van [GEMASKEERD VOOR REVIEW] – en het programma als geheel – ook effectief zijn op scholen die werken met andere sociale veiligheidsprogramma's anders dan KiVa.

4.3. Conclusie

De huidige studie toont de effecten aan van een schoolprogramma, [GEMASKEERD VOOR REVIEW], gericht op het bevorderen van een prosociaal klasklimaat door het versterken van basispsychologische behoeften – autonomie, competentie, en verbondenheid. Wanneer geïmplementeerd zoals bedoeld, waren leerlingen op interventiescholen beter af dan leerlingen op controlescholen wat betreft prosociaal gedrag, intrinsieke (prosociale) motivatie, bevrediging van basispsychologische behoeften, welbevinden en veiligheid op school, en antisociaal gedrag. Leerkrachten rapporteerden een positievere houding ten opzichte van stimulering van het vervullen van basispsychologische behoeften en betere burgerschapsvaardigheden bij hun leerlingen. Het is belangrijk om de sleutel tot optimale implementatie te begrijpen, aangezien dit een cruciale voorwaarde is voor een succesvol en effectief programma.

Verklaringen

Beschikbaarheid van data

De data die in de huidige studie zijn verkregen, zullen openbaar beschikbaar worden gemaakt na publicatie van de belangrijkste onderzoeksresultaten. De data zullen worden gearhiveerd in de *DANS Easy Environment* van de KNAW, zoals overeengekomen met NWO (*open access policy*). Voor meer informatie kan er contact worden opgenomen met de corresponderende auteur.

Bijdragen van de auteurs

Beide auteurs hebben bijgedragen aan de opzet en het ontwerp van de studie. [GEMASKEERD VOOR REVIEW] coördineerde en voerde de dataverzameling tijdens de studie uit, [GEMASKEERD VOOR REVIEW] begeleidde de dataverzameling. [GEMASKEERD VOOR REVIEW] schreef de conceptversie van het manuscript, [GEMASKEERD VOOR REVIEW] gaf kritische feedback. Beide auteurs hebben het definitieve manuscript gelezen en goedgekeurd.

References

- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K. F., Simel, D., & Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *Journal of the American Medical Association*, 276(8), 637–639. <https://doi.org/10.1001/jama.276.8.637>
- Bruhn, A. L., Hirsch, S. E., & Lloyd, J. W. (2015). Treatment Integrity in School-Wide Programs: A Review of the Literature (1993–2012). *Journal of Primary Prevention*, 36(5), 335–349. <https://doi.org/10.1007/s10935-015-0400-9>
- Caprara, G. V., Luengo Kanacri, B. P., Zuffianò, A., Gerbino, M., & Pastorelli, C. (2015). Why and How to Promote Adolescents' Prosocial Behaviors: Direct, Mediated and Moderated Effects of the CEPIDEA School-Based Program. *Journal of Youth and Adolescence*, 44(12), 2211–2229. <https://doi.org/10.1007/s10964-015-0293-1>
- Cheon, S. H., Reeve, J., & Ntoumanis, N. (2019). An intervention to help teachers establish a prosocial peer climate in physical education. *Learning and Instruction*, 64(June 2018), 101223. <https://doi.org/10.1016/j.learninstruc.2019.101223>
- Cohen, J. (1988). Statistical power analysis for the social sciences. *Hillsdale, NJ: Erlbaum*.
- Courser, M. W., Shamblen, S. R., Lavrakas, P. J., Collins, D., & Ditterline, P. (2009). The impact of active consent procedures on nonresponse and nonresponse error in youth survey data: Evidence from a new experiment. *Evaluation Review*, 33(4), 370–395. <https://doi.org/10.1177/0193841X09337228>
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Ellis, B. J., Volk, A. A., Gonzalez, J. M., & Embry, D. D. (2016). The Meaningful Roles Intervention: An Evolutionary Approach to Reducing Bullying and Increasing Prosocial Behavior. *Journal of Research on Adolescence*, 26(4), 622–637. <https://doi.org/10.1111/jora.12243>
- Evans, I. M., Harvey, S. T., Buckley, L., & Yan, E. (2009). Differentiating classroom climate concepts: Academic, management, and emotional environments. *Kotuitui*, 4(2), 131–146. <https://doi.org/10.1080/1177083X.2009.9522449>

- Gee, B., Wilson, J., Clarke, T., Farthing, S., Carroll, B., Jackson, C., King, K., Murdoch, J., Fonagy, P., & Notley, C. (2021). Review: Delivering mental health support within schools and colleges – a thematic synthesis of barriers and facilitators to implementation of indicated psychological interventions for adolescents. *Child and Adolescent Mental Health*, 26(1), 34–46. <https://doi.org/10.1111/camh.12381>
- Haataja, A., Voeten, M., Boulton, A. J., Ahtola, A., Poskiparta, E., & Salmivalli, C. (2014). The KiVa antibullying curriculum and outcome: Does fidelity matter? *Journal of School Psychology*, 52(5), 479–493. <https://doi.org/10.1016/j.jsp.2014.07.001>
- Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials—the gold standard for effectiveness research. *BJOG: An International Journal of Obstetrics & Gynaecology*, 125(13), 1–4. <https://doi.org/10.1111/1471-0528.15199>
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). *Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools?* 79(2), 181–193. <https://doi.org/10.1177/001440291307900204>
- Kaufman, T., & Huitsing, G. (2019). Implementatie van een antipestprogramma: Wat zijn de succes-ingrediënten? *Basischool Management*, 3, 10–13.
- Khalfaoui, A., García-Carrión, R., & Villardón-Gallego, L. (2021). A systematic review of the literature on aspects affecting positive classroom climate in multicultural early childhood education. *Early Childhood Education Journal*, 49(1), 71–81. <https://doi.org/10.1007/s10643-020-01054-4>
- Kirlin, M. (2005). Understanding the Relationship between Civic Skills and Civic Participation: Educating Future Public Managers. *Journal of Public Affairs Education*, 11(4), 305–314. <https://doi.org/10.1080/15236803.2005.12001404>
- Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential—but often forgotten—component of school-based interventions. *Preventing School Failure: Alternative Education for Children and Youth*, 48(3), 36–43. <https://doi.org/10.3200/psfl.48.3.36-43>
- Lenzi, M., Vieno, A., Sharkey, J., Mayworm, A., Scacchi, L., Pastore, M., & Santinello, M. (2014). How School can Teach Civic Engagement Besides Civic Education: The Role of Democratic School Climate. *American Journal of Community Psychology*, 54(3–4), 251–261. <https://doi.org/10.1007/s10464-014-9669-8>

- Lin, A. (2015). Citizenship education in American schools and its role in developing civic engagement: a review of the research. *Educational Review*, 67(1), 35–63.
<https://doi.org/10.1080/00131911.2013.813440>
- Lindenberg, S., & Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behavior. *Journal of Social Issues*, 63(1), 117–138.
- Liu, C., Cox, R. B., Washburn, I. J., Croff, J. M., & Crethar, H. C. (2017). The Effects of Requiring Parental Consent for Research on Adolescents' Risk Behaviors: A Meta-analysis. *Journal of Adolescent Health*, 61(1), 45–52.
<https://doi.org/10.1016/j.jadohealth.2017.01.015>
- Loon, A. W. G. van, Creemers, H. E., Beumer, W. Y., Okorn, A., Vogelaar, S., Saab, N., Miers, A. C., Westenberg, P. M., & Asscher, J. J. (2020). Can schools reduce adolescent psychological stress? A multilevel meta-analysis of the effectiveness of school-based intervention programs. *Journal of Youth and Adolescence*. <https://doi.org/10.1007/s10964-020-01201-5>
- Loon, A. W. G. van, Creemers, H. E., Vogelaar, S., Saab, N., Miers, A. C., Westenberg, P. M., & Asscher, J. J. (2024). Implementation and experimental evaluation of school-based intervention programs promoting adolescent mental health: Lessons learned. *School Psychology International*, 45(1), 70–84. <https://doi.org/10.1177/01430343231183999>
- Ma, X., Shen, J., & Reeves, P. (2023). Measuring integrity and fidelity of program implementation: Validating an instrument designed for school renewal. *Evaluation and Program Planning*, 100(March), 102341.
<https://doi.org/10.1016/j.evalprogplan.2023.102341>
- Manzano-Sánchez, D., & Valero-Valenzuela, A. (2019). Implementation of a model-based programme to promote personal and social responsibility and its effects on motivation, prosocial behaviours, violence and classroom climate in primary and secondary education. *International Journal of Environmental Research and Public Health*, 16(21).
<https://doi.org/10.3390/ijerph16214259>
- Mesurado, B., Guerra, P., Richaud, M. C., & Rodriguiz, L. M. (2018). Effectiveness of prosocial behavior interventions: A meta-analysis. In *Psychiatry and Neuroscience Update: From Translational Research to a Humanistic Approach: Volume III* (Vol. 3, Issue April 2020).
<https://doi.org/10.1007/978-3-319-95360-1>

- Montori, V. M., & Guyatt, G. H. (2001). Intention-to-treat principle. *Canadian Medical Association Journal*, 165(10), 1339–1341.
- Ntoumanis, N., Ng, J. Y. Y., Prestwich, A., Quested, E., Hancox, J. E., Thøgersen-Ntoumani, C., Deci, E. L., Ryan, R. M., Lonsdale, C., & Williams, G. C. (2021). A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health Psychology Review*, 15(2), 214–244. <https://doi.org/10.1080/17437199.2020.1718529>
- Pokorny, S. B., Jason, L. A., Schoeny, M. E., Townsend, S. M., & Curie, C. J. (2001). When Active Consent Procedures Replace Passive Consent. *Sage Publications*, 25(5), 567–580.
- Research randomizer. (2023). <https://www.randomizer.org/>
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68–78. <https://doi.org/10.1037/0003-066X.55.1.68>
- Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. In *Elgar Encyclopedia of Organizational Psychology*. <https://doi.org/10.4337/9781803921761.00125>
- Saarento, S., & Salmivalli, C. (2015). The role of classroom peer ecology and bystanders' responses in bullying. *Child Development Perspectives*, 9(4), 201–205. <https://doi.org/10.1111/cdep.12140>
- Sanetti, L. M. H., Cook, B. G., & Cook, L. (2021). Treatment Fidelity: What It Is and Why It Matters. *Learning Disabilities Research and Practice*, 36(1), 5–11. <https://doi.org/10.1111/ldrp.12238>
- Shin, J., & Lee, B. Y. (2021). The effects of adolescent prosocial behavior interventions: a meta-analytic review. *Asia Pacific Education Review*, 22(3), 565–577. <https://doi.org/10.1007/s12564-021-09691-z>
- Sklad, M., Diekstra, R., de Ritter, M., Ben, J., & Gravesteyn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, 49(9), 892–909. <https://doi.org/10.1002/pits>
- Tanner-Smith, E. E., Durlak, J. A., & Marx, R. A. (2018). Empirically Based Mean Effect Size Distributions for Universal Prevention Programs Targeting School-Aged Youth: A Review

- of Meta-Analyses. *Prevention Science*, 19(8), 1091–1101. <https://doi.org/10.1007/s11121-018-0942-1>
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting Positive Youth Development Through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects. *Child Development*, 88(4), 1156–1171. <https://doi.org/10.1111/cdev.12864>
- Thornberg, R. (2011). 'She's Weird!' - the social construction of bullying in school: A review of qualitative research. *Children and Society*, 25(4), 258–267. <https://doi.org/10.1111/j.1099-0860.2011.00374.x>
- Thornberg, R., Wegmann, B., Wänström, L., Bjereld, Y., & Hong, J. S. (2022). Associations between student–teacher relationship quality, class climate, and bullying roles: A Bayesian multilevel multinomial logit analysis. *Victims and Offenders*, 17(8), 1196–1223. <https://doi.org/10.1080/15564886.2022.2051107>
- Tigges, B. B. (2003). Parental consent and adolescent risk behavior research. *Journal of Nursing Scholarship*, 35(3), 283–289. <https://doi.org/10.1111/j.1547-5069.2003.00283.x>
- Villardón-Gallego, L., García-Carrión, R., Yáñez-Marquina, L., & Estévez, A. (2018). Impact of the interactive learning environments in children's prosocial behavior. *Sustainability (Switzerland)*, 10(7), 1–12. <https://doi.org/10.3390/su10072138>
- Wagner, E. F., Tubman, J. G., & Gil, A. G. (2004). Implementing school-based substance abuse interventions: Methodological dilemmas and recommended solutions. *Addiction*, 99(s2), 106–119. <https://doi.org/10.1111/j.1360-0443.2004.00858.x>
- Wang, M. Te, L. Degol, J., Amemiya, J., Parr, A., & Guo, J. (2020). Classroom climate and children's academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, 57, 100912. <https://doi.org/10.1016/j.dr.2020.100912>
- Wolfenden, L., Kypri, K., Freund, M., & Hodder, R. (2009). Obtaining active parental consent for school-based research: A guide for researchers. *Australian and New Zealand Journal of Public Health*, 33(3), 270–275. <https://doi.org/10.1111/j.1753-6405.2009.00387.x>

Mail over eerste versie

Van: "Jochum Veenstra" <jeugdinontwikkeling@outlook.com>

Aan: "Amanda van Loon" <a.w.g.vanloon@uu.nl>

Onderwerp: Redacteursbeslissing

Geachte Amanda van Loon,

Graag deel ik mee dat we uw manuscript voor het themanummer, Wat is er nodig voor het creëren van prosociale klassen? De effectiviteit van een schoolinterventie voor basisschoolkinderen, doorsturen naar peer reviewers.

Als u vragen heeft, neem dan contact met mij op via jeugdinontwikkeling@outlook.com.

Met vriendelijke groet,

Jochum Veenstra

Mail met reviews over de eerste versie

Van: "Jochum Veenstra" <jeugdinontwikkeling@outlook.com>

Aan: "Amanda van Loon" <a.w.g.vanloon@uu.nl>

Onderwerp: Redacteursbeslissing

Geachte Amanda van Loon, Tessa Kaufman,

Uw artikel "Themanummer Wat is er nodig voor het creëren van prosociale klassen? De effectiviteit van een schoolinterventie voor basisschoolkinderen." is beoordeeld door twee leden van de redactie.

Na het lezen van de reviews en het overwege ervan stelt de themaredactie dat het onderwerp relevant is voor het themanummer en dat de steekproef en de onderzoeksopzet veelbelovend zijn. Toch twijfelden de reviewers tussen afwijzen en revisie. De voornaamste reden hiervoor is dat in de huidige versie te veel de nadruk wordt gelegd op de resultaten uit een kleine subset van de steekproef, terwijl de vooraf geplande hoofdanalyses die consistent geen significante effecten laten zien onvoldoende worden besproken. Dit wordt onvoldoende transparant gemaakt in abstract, discussie en conclusie en dit levert een te positieve interpretatie van de uitkomsten op.

Uiteindelijk is gekozen voor de mogelijkheid tot revisie, onder de voorwaarde dat de hoofdbevindingen correct en volledig worden gepresenteerd. Een herziene versie dient in alle onderdelen van het artikel de hoofdanalyses centraal te stellen (dus de nulbevindingen), en de bespreking van analyses die gedaan zijn in de subgroep dienen een meer secundair karakter te krijgen. Abstract, conclusie en aanbevelingen dienen hiermee in lijn te worden gebracht.

We willen benadrukken dat we het ook juist relevant vinden om niet-significante uitkomsten te publiceren. Zie voor andere punten de reviews hieronder.

We begrijpen dat dit om een grote revisie gaat, zou u ons daarom binnen twee weken willen mailen (op: jeugdinontwikkeling@outlook.com) of u de revisie zou willen schrijven? Ik vraag dit in verband met de planning van ons themanummer en strakke deadlines waar we aan gebonden zijn.

Mocht u de grote revisie willen schrijven dan is de deadline voor de volgende versie 16 september.

Voeg bij u nieuwe versie zowel een antwoordbrief als een versie van het manuscript mét track and trace (of arceringen).

Wilt u een nieuwe versie indienen dan kan dat via jeugdinontwikkeling.nl

Met vriendelijke groeten,
Jochum Veenstra (redactie-secretaris)

Reviewer 1:

Aanbeveling: Weiger inzending

Titel

De titel is aantrekkelijk, kort en bondig. Het onderzoek gaat meer in op effectiviteit dan op "wat is er nodig", maar ik zie daar niet direct een probleem.

Abstract

onvoldoende

Abstract - toelichting

Het abstract is goed gestructureerd, maar is geen goede weergave van de resultaten en zelfs misleidend doordat er een veel te rooskleurig beeld geschetst wordt van de resultaten. Volgens het abstract zouden er tenminste 7 significante uitkomsten zijn. In werkelijkheid was er in de steekproef in zijn geheel geen enkel significant resultaat. En op het einde van de interventie in de implementatiegetrouwe subgroep maar 1 significant effect en 7 niet-significante effecten wanneer we naar de primaire variabelen kijken en is dit effect ook nog eens zeer klein (.10). Daarbij ging het ook nog eens om een voor deze studie ontworpen uitkomstmaat die niet voor deze populatie gevalideerd is. Ook opvallend: volgens de tabellen zijn er ook negatieve resultaten (negatieve beta coëfficiënten, maar dit wordt nergens besproken) . Niet significante resultaten dienen evenveel aandacht te krijgen als sign resultaten.

Met andere woorden: er is maar zeer geringe evidentie voor effectiviteit.

Het abstract dient te vermelden hoe groot de steekproef was waarin sensitiviteitsanalyses met implementatiegetrouwheid zijn gedaan.

Inleiding

meerdere wijzigingen zijn nodig

Inleiding - toelichting

- De relevante literatuur wordt besproken en de argumentatie wordt goed opgebouwd.
- Het theoretische kader wordt zeer beknopt besproken en niet gekoppeld aan de inhoud van de interventie. Ook in de methode wordt niet echt duidelijk wat de interventie inhoudt. Terwijl de lezer wel nieuwsgierig is. Er mag meer uitgelegd worden over Meaningful Roles.
- in relatie tot de vorige opmerking: Er wordt inhoudelijk geen vergelijking gemaakt met bestaande interventieprogramma's (in Nederland), waardoor het niet duidelijk is wat de uniciteit is van dit programma tov bestaande theory-based/informed of evidence-based/informed interventies (bv. het NJI heeft 3 universele interventies voor klasklimaat, 7 universele interventies voor sociale vaardigheden op school en 7 voor schoolklimaat en 1 universele, schoolse interventie voor burgerschap opgenomen in de database).
- De verantwoording van de afhankelijke variabelen is summier. De indeling in primaire en secundaire variabelen moet verantwoord worden in de inleiding
- Uitleg bij figuur 1 ontbreekt in de tekst. Elk concept in de figuur moet duidelijk zijn voor de lezer. Ik miste vooral informatie over rollen, vergaderingen, complimenten. Wat wordt hiermee bedoeld? Waarom is dit belangrijk? Wat is de relatie met de zelfdeterminatietheorie?
- de aandacht voor implementatiegetrouwheid is een sterkte.

Methode

meerdere wijzigingen zijn nodig

Methode - toelichting

- het betreft een sterk design (RCT) met herhaalde metingen, een grote steekproef, een redelijke responsgraad. De onderzoekers doen ook sensitiviteits analyses en analyseren drop-out en de randomisatie met de beschikbare gegevens. Dit alles maakt een sterke indruk.
- Een aantal zaken ontbreken:
- inhoud en duur van de trainingssessies voor leerkrachten worden niet besproken.
 - inhoud en format van de interventie zelf: wat moeten leerkrachten implementeren in hun klassen en hoe?
 - ik vond de tabel helpend, maar er ontbreekt nog informatie zoals de antwoordschaal (essentieel om de statistiek te kunnen begrijpen). Kan dit nog opgenomen worden in de tabel? Ook een korte definitie is nodig (of een voorbeeld item). Meetmoment en Type kunnen weggelaten worden, zodat er ruimte komt in de tabel. Een andere manier om ruimte te maken in de tabel is om de verschillende benamingen van de variabelen in dezelfde kolom te plaatsen, waarbij de uitkomstcategorieën (bv prosociaal gedrag) vet gedrukt worden en de namen niet-vet (bijvoorbeeld). De meetmomenten kunnen direct achter de variabelenaam geplaatst worden, zodat er geen aparte kolom nodig is. Als er aparte tabellen gemaakt worden voor werkzame elementen en moderatoren, dan is het ook niet nodig om een kolom met type te voorzien.
- appendix A: beschrijving instrumenten. Hier dienen nog de coderingen van de antwoorden toegevoegd te worden, inclusief de coderingen van de demografische variabelen (immers

essentieel om de statistiek te kunnen interpreteren).

-de beschrijving van de steekproef dient vermeld te worden in de methode en niet in de resultaat. Ook schoolkenmerken van de deelnemende scholen moeten hier besproken worden.

ANALYSES

-Is er een multi-level model geanalyseerd met conditie op cluster (school) level?

Leerlingenvragenlijsten: is er gecontroleerd voor clustervariantie op klasniveau en schoolniveau?

Leerkrachtvragenlijsten: is er gecontroleerd voor clustervariantie op schoolniveau?

-er wordt in Tabel 1 gesproken over moderatoren, maar er zijn geen hypothesen over moderatie effecten geformuleerd in de inleiding. Demografische gegevens zouden opgenomen moeten worden als covariaten in de regressie analyse.

- Een alpha van .35 -.42 is bijzonder laag. Die uitkomstmaat zou eigenlijk niet meegenomen mogen worden in het onderzoek. Omdat het een secundaire en geen primaire uitkomst betreft, zou ik deze weglaten.

-tabellen met details over de steekproef per conditie zouden naar de appendix kunnen (als de gelijkenissen en verschillen tussen condities dan wel kort in de tekst besproken worden). Dat maakt ook ruimte voor de beschrijving van de (codering van) de afhankelijke variabelen , zoals eerder genoemd (essentieel om de statistiek te kunnen beschrijven .

-de sensitiviteits- en drop-out analyses en de aandacht voor implementatiegetrouwheid zijn sterktes in dit onderzoek.

Resultaten

onvoldoende

Resultaten - toelichting

-de tabellen vond ik zeer lastig te lezen en de presentatie van de resultaten is niet compleet en niet volgens APA richtlijnen. Er ontbreekt essentiële informatie (vooral in de regressietabellen), terwijl andere informatie overbodig is. Daardoor is het niet mogelijk om de resultaten van de interventie-effecten goed te reviewen.

Ik licht hieronder verder toe:

-analyse van mogelijke outliers ontbreekt. Ook wordt er geen informatie gegeven over de scheefheid en gepiektheid van de verdelingen van de afhankelijke variabelen.

-tabel 2: het is niet gebruikelijk om percentages te geven per antwoordcategorie. Voor interval/ratio variabelen moet gerapporteerd worden: M, SD, scheefheid, gepiektheid (voor de hele steekproef en voor subgroepen).

-het zou de lezer helpen wanneer eerst alle controlechecks en beschrijvende statistiek gerapporteerd wordt, en de toetsende statistiek nadien apart gerapporteerd worden.

-het zou de lezer helpen wanneer de resultaten geordend worden volgens primaire en secundaire uitkomstvariabelen

-in de tabellen staan veel negatieve effecten (negatieve beta coëfficiënten), maar toch worden deze geïnterpreteerd als positieve effecten. Dat is zeer verwarrend. Kunnen de onderzoekers dat uitleggen?

- de opmaak van de tabellen is niet volgens APA (bijv de wijze waarop significantie gerapporteerd wordt).
- tabellen met regressie analyses zijn incompleet en moeten het volgende bevatten: alle predictoren inclusief conditie, covariaten (demografische variabelen) en de T1 meting in de eerste kolom. Daarnaast moeten volgens APA minimaal de ongestandaardiseerde beta en de standaardfout (SE) vermeld worden . Verder dienen tabellen met regressieanalyses de afhankelijke variabele en de steekproefgrootte (N of n) te vermelden in de titel vd tabel. Zonder deze informatie is het niet mogelijk voor een reviewer om de interventie-effecten te reviewen.
- het is vreemd dat dezelfde informatie wordt gegeven in meerdere tabellen en dat de cijfers afwijken terwijl ze exact hetzelfde zouden moeten zijn (als ik het allemaal goed begrijp). Bijvoorbeeld: tabellen 11 en 12 vermelden beide de M en SD van de afhankelijke variabelen in de controle en experimentele condities en toch zijn de cijfers niet exact hetzelfde. Nog verontrustender: beide tabellen vermelden ook dezelfde toetsende statistiek voor verschillen op T1 en rapporteren desondanks zelfs verschillende toetswaarden. Hetzelfde geldt voor tabellen 13 en 14.
- hoe werd de effectgrootte berekend? Zijn alle cijfers dubbel gecheckt? Ik had de indruk dat de effectgrootte soms niet klopte als ik naar het gemiddelde verschil en de SD keek. Maar dat kan aan een andere berekening te wijten zijn.
- de tabellen rapporteren apart over pesten en slachtofferschap, maar ik kon deze variabele niet terugvinden in Tabel 1 in de methode. Daar wordt gesproken van 1 variabele voor pesten en slachtofferschap

Discussie

onvoldoende

Discussie - toelichting

- de discussie doet geen recht aan de resultaten en brengt een te rooskleurig beeld van de resultaten (zie feedback op het abstract). Niet significante resultaten moeten minimaal evenveel aandacht krijgen als significante resultaten en direct bij aanvang van de discussie vermeld worden (en in abstract en conclusie). Ook negatieve effecten moeten besproken worden. Ook de uitdovende effecten tussen T2 en T3 verdienen meer aandacht. Als er amper meer positieve effecten zijn op T3 (slechts 1 van de 8 primaire variabelen waarbij het ook nog eens een schaal betreft die zelf ontworpen is en niet gevalideerd is voor de populatie), kan men niet anders dan concluderen dat de interventie niet effectief is (in deze vorm). Men kan pas spreken van bewezen effectiviteit wanneer er ook effecten gevonden worden (op T3) op gevalideerde vragenlijsten (wat hier niet het geval was). Uiteraard zal JiO een artikel niet afwijzen omwille van nulresultaten wanneer de methodologische kwaliteit van het onderzoek voldoende sterk is.
- het zou helpen wanneer er in de discussie onderscheid werd gemaakt tussen primaire en secundaire uitkomsten
- de zwakten in de methode worden onvoldoende besproken, te weten:
- leerkrachtenuitkomsten: leerkrachten waren niet blind voor feit dat ze deelnamen aan hun

interventie. Waarschijnlijk is hier sprake van bias in leerkrachtrapportages.

-veel vragenlijsten waren niet gevalideerd voor deze populatie

-in relatie tot vorige opmerking: veel vragenlijsten waren zelf ontworpen specifiek voor deze studie. Het is bekend dat er grotere effecten of eerder significante resultaten gevonden worden met zelfgeconstrueerde, interventiespecifieke vragenlijsten. Men kan pas spreken van bewezen effectiviteit wanneer er ook effecten gevonden worden (op T3) op gevalideerde vragenlijsten (wat hier niet het geval was).

-niet alle vragenlijsten/variabelen konden betrouwbaar gemeten worden. Een alpha van .35 -.42 is bijzonder laag (zie eerdere suggestie om deze weg te laten, omdat het een secundaire variabele betreft).

-het is opvallend dat deze interventie geïmplementeerd werd in KIVA scholen. Is hier sprake van selectie- of steekproefbias? Hoe kan dit de resultaten beïnvloeden hebben?

Conclusies

onvoldoende

Conclusies - toelichting

De conclusies lijken misleidend (zie commentaar bij abstract en discussie).

Tabellen en figuren

De tabellen zijn lastig te begrijpen en de informatie is incompleet (zie feedback op de resultaten).
De figuren zijn mooi.

Referenties

oke

Ruimte voor kleine opmerkingen

Algemeen oordeel

weigeren

Reviewer 2:

Aanbeveling: Revisies benodigd

Titel

De titel suggereert dat de besproken schoolinterventie effectief is, maar dat lijkt een wat te rooskleurige presentatie. Verzoek aan de auteurs om een titel te kiezen die beter past bij de uitkomsten van deze studie.

Abstract

onvoldoende

Abstract - toelichting

Het abstract is qua structuur en taal overwegend goed geschreven, maar een belangrijke tekortkoming is dat alleen positieve uitkomsten worden gepresenteerd zonder de uitkomsten van de hoofdanalyses te noemen die allemaal niet statistisch significant waren. De analyses die gaan over de hele steekproef van 2237 leerlingen worden helemaal niet gerapporteerd in het abstract en deze resultaten zouden juist als hoofdresultaten gepresenteerd moeten worden in plaats van de resultaten van de sensitiviteitsanalyses binnen een kleine subgroep. Zie ook de opmerkingen bij de resultatensectie en discussie over dit punt. Verzoek aan de auteurs is om in de abstract een meer evenwichtig beeld van de resultaten te geven, waarin positieve resultaten maar ook welke resultaten allemaal niet werden gevonden worden genoemd en ook de (kleine!) effectgroottes aan bod komen.

Kleinere punten in het abstract:

De eerste zin leest niet zo makkelijk. Misschien op te lossen door er 'evenals voor het optimaliseren van een effectieve leeromgeving' van te maken.

op-evidentie-gebaseerde -> In het Nederlands wordt dit los geschreven: op evidentie gebaseerde

Derde zin: 'de intrinsieke motivatie van leerlingen te verhogen' -> intrinsieke motivatie voor wat?

Graag noemen hoe groot de subgroep met ideale programmatrouw was.

Laatste zin: wat wordt hier precies mee bedoeld? Na lezing van het hele manuscript ga ik ervan uit dat bedoeld wordt dat het belangrijk is beter te begrijpen hoe optimale implementatie van dit basisschoolprogramma kan worden bevorderd omdat de effectiviteit daar deels van af hangt. Dat is wat concreter dan 'inzicht in optimale implementatie is essentieel'.

Inleiding

meerdere wijzigingen zijn nodig

Inleiding - toelichting

De eerste twee alinea's zijn duidelijk en goed geschreven. De derde alinea zou baat hebben bij een duidelijker structuur. Nu de bespreking van eerdere studies van specifiek naar algemeen in plaats van andersom en algemenere effecten zoals welzijn en specifiekere effecten zoals prosociale vaardigheden en effecten voor volwassenen en kinderen staan door elkaar beschreven.

Verzoek is om een duidelijker structuur aan te brengen in deze alinea. Het wordt soms ook niet goed duidelijk of een besproken resultaat over de 'Meaningful roles' interventie gaat of meer in het algemeen over 'dit soort programma's'. Als eerst andere soortgelijke programma's zouden worden besproken en daarna pas specifieke resultaten van 'Meaningful roles' dan wordt het wat makkelijker de verschillende resultaten uit elkaar te houden.

De inhoud van de interventie die getoetst wordt in de huidige studie en hoe deze interventie verschilt van andere interventies die ook vallen onder 'dit soort programma's' blijft ook wat onduidelijk. Een mogelijke oplossing zou kunnen zijn om naast een korte aanvullende beschrijving in het manuscript zelf een bijlage over de specifiekere inhoud van de interventie toe te voegen.

Ook wordt beschreven dat dit een antipestprogramma voor middelbare scholieren is maar het is in deze studie gebruikt op basisscholen. Dit vereist ook nog wel wat extra toelichting.

In alinea 4 zou het goed zijn de onderscheid te maken tussen primaire en secundaire uitkomstmaten (zoals in figuur 1 wel duidelijk wordt gedaan). Ook worden niet alle hypothesen onderbouwd.

Methode

meerdere wijzigingen zijn nodig

Methode - toelichting

De methode is over het algemeen duidelijk en goed beschreven.

Belangrijk punt: Waar in de methode duidelijk wordt beschreven dat de intention-to-treat analyses (dus de analyses op de gehele steekproef) de hoofdanalyses zijn en de analyses op de subgroep die de interventie optimaal heeft doorgevoerd de sensitiviteitsanalyses, matcht dit niet met hoe in abstract en discussie met de resultaten wordt omgegaan.

Wat niet helemaal duidelijk was:

1) Waarom is het feit dat dezelfde organisatie die KiVa heeft ontwikkeld ook de in de huidige studie geteste interventie heeft ontwikkeld een argument om alleen KiVa scholen te werven? Dit werd mij niet duidelijk. En wat zijn de andere implicaties van KiVa scholen werven? Betekent het dat deze scholen ook meedoen aan KiVa? Wordt hier dan eigenlijk niet het effect van 2 ipv 1 interventie getoetst? Of juist het effect van de huidige interventie bovenop het effect van KiVa? Graag een toelichting hierop toevoegen in de methode sectie.

In 2.1 graag verwijzen naar figuur 2.

Tabel 1: Moderators -> moderatoren

2.5 Statistische analyse: Graag de tweede zin van alinea 2 wat duidelijker formuleren. met baseline (T1) worden neem ik aan de baseline van de uitkomstmaten bedoelt? Graag iets

duidelijker omschrijven. Ook graag uitleggen met welk doel de covariantie van conditie en de specifieke variabele(n) is opgenomen in de modellen en hoe.

Resultaten

meerdere wijzigingen zijn nodig

Resultaten - toelichting

In alinea 2 van paragraaf 3.1 staat N=2337, terwijl in het abstract 2237 staat.

Tabel 4: Goed om ook bij deze tabel aan te geven dat vetgedrukte statistieken statistisch significante resultaten weergeven zodat men in 1 oogopslag kan zien dat er geen significante bevindingen in de tabel staan.

3.2 Randomisatie en dropouts

De verwijzing naar tabel 7 en 8 staat nu aan het eind van een zin waarvan het eerste deel over tabel 7 gaat en het tweede deel over tabel 8. Graag de tabel direct achter het relevante zinsdeel noemen.

Tabel 11: Verschillen op T1 -> Verschillen op T1

Tabel 11: Voor 3 van de 7 significante resultaten lijkt er eigenlijk geen verbetering te hebben plaatsgevonden voor de interventiegroep (en soms zelfs een verslechtering) maar een nog ergere verslechtering voor de controlegroep. Dit wordt niet besproken in de resultatensectie of in de discussie.

Ook scoorde de experimentele conditie als ik het goed zie op alle uitkomstmaten op T1 (baseline) al beter dan de controlegroep. De implicaties hiervan worden niet of in ieder geval niet voldoende besproken. In de discussie wordt wel dit genoemd "Hoewel we in de analyses rekening hielden met de verschillen op baseline tussen de experimentele- en controlegroep, suggereren deze verschillen dat de randomisatie niet geheel succesvol was", maar ik mis wat de implicaties hiervan zijn volgens de auteurs.

Mijn opmerkingen over tabel 11 gelden gedeeltelijk ook voor tabel 12.

Laatste zin -> en kwamen vaker uit de experimentele conditie

Discussie

onvoldoende

Discussie - toelichting

Net als in het abstract dienen ook in de discussie de uitkomsten van de hoofdanalyses (intention-to-treat analyses) te worden besproken. Die analyses horen centraal te staan en natuurlijk kunnen daarna de uitkomsten van de sensitiviteitsanalyses ook worden genoemd, maar een

hoofdboodschap zou toch zeker moeten zijn dat er geen aanwijzingen zijn dat de interventie werkt als naar de intention-to-treat resultaten wordt gekeken. En ook moeten de sensitiviteitsanalyses voorzichtiger worden geïnterpreteerd. De groepen waren op baseline niet vergelijkbaar, voor een aantal van de statistisch significante effecten was er in ongeveer de helft van de effecten helemaal geen verbetering of zelfs een verslechtering in de interventie conditie (alleen een nog sterkere verslechtering in de controlegroep). Is het dan wel gerechtvaardigd om te claimen dat de interventie heeft gewerkt? Misschien als er redenen zijn om aan te nemen dat leerlingen altijd verslechteren in de periode tussen T1 en T2, maar hier zeggen de auteurs niks over. Bovendien waren de gevonden effecten ook nog eens allemaal erg klein.

De discussie bevat delen die mijns inziens echt overinterpretatie zijn en niet te rechtvaardigen op basis van de uitkomsten van de analyses:

1) heel alinea 2: verreweg de meeste resultaten ondersteunen de claim niet dat "het vervullen van basispsychologische behoeften ter bevordering van prosociale klassen ook effectief kan zijn bij jongere leeftijdsgroepen en binnen een breder scala aan interventies". Er wordt ook gerefereerd aan de grote steekproef maar die grote steekproef is helemaal niet gebruikt voor de sensitiviteitsanalyses waar het steeds over gaat in de discussie.

2) "Onze bevindingen hebben belangrijke implicaties voor leerkrachten en andere belanghebbenden. Ten eerste lieten onze resultaten zien dat strategieën om autonomie, competentie, en verbondenheid te stimuleren effectief kunnen zijn bij het creëren van prosociale klassen. Daarom wordt het aanbevolen dat leerkrachten op een basisbehoeften-stimulerende manier werken die de basisbehoeften van leerlingen versterkt, bijvoorbeeld door leerlingen meer autonomie en verantwoordelijkheid te geven bij taken of opdrachten (i.e., betekenisvolle rollen)." -> de 0-bevindingen van alle intention-to-treat analyses die eigenlijk de hoofdanalyses waren, gecombineerd met de hele kleine effecten bij de subgroep die de interventie helemaal zoals bedoeld heeft doorgevoerd, en daarbij nog het feit dat voor een aantal van die kleine effecten de interventiegroep helemaal niet verbeterde maar alleen minder verslechterde dan de controlegroep, maken dat deze claims niet gerechtvaardigd zijn op basis van de gepresenteerde resultaten.

3) "Onze resultaten toonden aan dat het programma alleen effectief was in een subgroep van leerlingen en leerkrachten die een ideale programmatrouw rapporteren." -> Op basis van de punten die ik hierboven ook noem zou ik zeggen dat er eigenlijk nauwelijks aanwijzingen zijn gevonden dat het programma effectief was.

Ik mis in de discussie helemaal wat de KiVa component betekent voor de interpretatie van de bevindingen van het huidige onderzoek. Moeten we het zien als resultaten van de huidige interventie bovenop de resultaten van KiVa?

4.1 -> Ook hier weer word $N > 2000$ aangehaald om voldoende power te beargumenteren maar de subgroepresultaten zijn gebaseerd op een veel lagere steekproef.

"De lage power geeft aan dat slechts een klein deel van de leerlingen [...]" -> Dit is een wat vreemde formulering. Ik zou het stukje over dat de lage power dit aangeeft weglaten en gewoon beschrijven dat slechts een klein deel van de leerkrachten en leerlingen ervaren dat het

programma was doorgevoerd zoals bedoeld. Een belangrijke vraag is: waarom is deze interventie zo vaak niet doorgevoerd zoals bedoeld.

"Om de implementatie van schoolprogramma's te verbeteren, adviseerde eerder onderzoek om positieve relaties tussen externen en schoolpersoneel te creëren en te onderhouden [...]" -> Onduidelijk wat voor externen hier bedoeld worden.

Conclusies

onvoldoende

Conclusies - toelichting

De conclusie is veel te sterk geformuleerd en bevat op deze manier geen correcte en evenwichtige weergave van de bevindingen.

Tabellen en figuren

De tabellen en figuren zijn opzich duidelijk. Ik heb mijn feedback op de tabellen al bij de resultaten sectie gezet. De figuren zijn duidelijk en informatief.

Referenties

Appendix A: Ik mis referenties van Appendix A. Graag een referentielijst toevoegen.

Ruimte voor kleine opmerkingen

Beschikbaarheid van data: Begrijp ik het goed dat de data echt openbaar beschikbaar worden gemaakt in DANS? Dus iedereen kan de data vrij gebruiken? Dus niet dat alleen de meta-data beschikbaar worden gemaakt en dat als onderzoekers de data willen gebruiken dat ze dan contact op moeten nemen met de auteur? Misschien staat het er gewoon goed nu hoor, maar ik begon te twijfelen doordat er aan de ene kant staat dat de data zelf openbaar beschikbaar worden gemaakt en aan de andere kant dat contact met de auteur kan worden opgenomen.

Zal de syntax van de analyses ook publiek beschikbaar worden gemaakt (bijvoorbeeld op het Open science framework)?

Algemeen oordeel

revisie

Waarom is dit onderzoek wel of niet relevant voor de JiO-lezer?

Deze studie is in potentie erg interessant vanwege de grote steekproef en het gerandomiseerde cluster ontwerp. De interventie beoogde door versterking van de drie basispsychologische behoeften autonomie, competentie, en verbondenheid de intrinsieke motivatie van leerlingen en daarmee hun prosocialiteit in de klas te bevorderen. De interventie past goed binnen het themanummer. In de huidige vorm is dit manuscript echter niet geschikt voor publicatie, omdat de resultaten veel te positief worden geïnterpreteerd en weergegeven.

Mail met reviews over de tweede versie

Van: "Jochum Veenstra" <jeugdinontwikkeling@outlook.com>

Aan: "Amanda van Loon" <a.w.g.vanloon@uu.nl>

Onderwerp: Redacteursbeslissing

Geachte Amanda van Loon, Tessa Kaufman,

Uw artikel "Themanummer Wat is er nodig voor het creëren van prosociale klassen? De effectiviteit van een schoolinterventie voor basisschoolkinderen." is beoordeeld door twee leden van de redactie.

Beiden reviewers erkennen de relevantie van de studie. De themaredactie schrijft: 'Deze studie is erg interessant vanwege de grote steekproef en het gerandomiseerde cluster ontwerp. De interventie beoogde door versterking van de drie basispsychologische behoeften autonomie, competentie, en verbondenheid de intrinsieke motivatie van leerlingen en daarmee hun prosocialiteit in de klas te bevorderen en past erg goed binnen het huidige themanummer. Er is nauwelijks bewijs gevonden dat de interventie werkt zoals beoogd, slechts voor een kleine groep leerlingen met ideale programmatrouw zijn kleine positieve effecten gevonden die veelal kortdurend waren. In de eerste versie waren de resultaten wat te rooskleurig gepresenteerd, maar dat is in de huidige versie sterk verbeterd.

In de meer specifieke opmerkingen van de reviewers staat nog wel een aantal suggesties om dit nog iets verder door te voeren door bijvoorbeeld in de samenvatting niet enkel aan te geven dat een kleine groep leerlingen beter af was, maar dit te nuanceren door aan te geven dat ook daar de effecten veelal klein waren en kortdurend. Er is verder vrij veel tekst toegevoegd en hoewel dit over het geheel genomen het manuscript sterk heeft verbeterd riep het soms ook weer nieuwe vragen op. '

Zou u naar het nuanceren en de opmerkingen in de reviews willen kijken?

Zou u over een twee weken een nieuwe versie willen uploaden. De deadline voor de nieuwe versie is: 21 november. Mocht 21 november niet haalbaar zijn zou u me dan willen mailen op: jeugdinontwikkeling@outlook.com

Voeg bij u nieuwe versie zowel een antwoordbrief als een versie van het manuscript mét track and trace (of arceringen).

Wilt u een nieuwe versie indienen dan kan dat via jeugdinontwikkeling.nl

Met vriendelijke groeten,
Jochum Veenstra (redactie-secretaris)

Reviewer 1:

Aanbeveling: Revisies benodigd

Titel

Prima

Abstract

een kleine wijziging is nodig

Abstract - toelichting

Ten opzichte van de vorige versie is het abstract goed verbeterd en is de weergave van de resultaten waarheidsgetrouwer. Er is nog een onduidelijkheid voor mij: resultaten in de gehele steekproef laten geen significant resultaat zien, maar vervolgens wordt gesteld dat leerkrachten een meer positieve houding rapporteren. Dit lijkt mij tegenstrijdig. Dezelfde tegenstrijdigheid komt ook terug in de bijsluiters en in de discussie.

Ik zou in het abstract zeker nog vermelden dat het om KiVA scholen gaat (ook in de bijsluiters). Daarnaast vroeg ik mij of de auteurs duiden op beschikbaarheid van internationale of Nederlandse schoolprogramma's als ze zeggen dat er weinig opties zijn (dit komt ook terug op andere plaatsen in de tekst)

Bij bullet 5 van de bijsluiters zou ik nog toevoegen "maar niet op het einde van het schooljaar" voor de volledigheid

Inleiding

goed/voldoende

Inleiding - toelichting

De inleiding is veel uitgebreider en informatiever dan in de vorige versie. Het programma is nu inhoudelijk veel duidelijker. Nog een vraagje: Welke aanpassingen zijn precies gedaan om het programma geschikt te maken voor basisschoolleerlingen?

Methode

meerdere wijzigingen zijn nodig

Methode - toelichting

Een aantal zaken verdienen nog wat nadere toelichting of opheldering:

- Het verschil tussen sekse jongen, sekse meisje en gender is mij niet duidelijk.
- Het is mij niet helemaal duidelijk wat bedoeld wordt in tabel 3 met bespreken van pesten (gaat dit over het klassikaal bespreken van pesten?), houding tov pesten (bv of de leerkracht pesten afkeurt?), effectiviteit om pesten te verminderen (gaat het hier om gevoelens van zelfeffectiviteit?), houding over stimulatie prosociaal gedrag (gaat dit over of leerkrachten dit als hun taak zien misschien?)?

Analyses:

- Het is mij niet duidelijk wat er bedoeld wordt met aparte modellen waarbij de onderlinge samenhang van de effecten werd meegenomen? Gaat hier over een onderlinge samenhang tussen bepaalde variabelen? Zo ja, welke? En: hoe kun je onderlinge samenhang meenemen als er aparte modellen gedraaid werden?

- Het klinkt mij wat vreemd in de oren dat de regressie analyses geclusterd werden op schoolniveau. Bedoelen jullie misschien dat er in de analyses gecorrigeerd werd voor clustervariantie in de afhankelijke variabelen op schoolniveau? Ik maak hieruit op dat er geen multilevel analyses gedaan zijn met conditie als level-2 predictor, maar dat er enkel gecorrigeerd is voor clustervariantie?

- omdat er nu een onderzoeksvraag is die betrekking heeft op effectiviteit onder conditie van implementatietrouw (doel 2), zou ik niet langer spreken van sensitiviteitsanalyses en dit overal weghalen (ook in de resultaten en tabellen).

-Implementatietrouw

Het kan helpen om in de methode een aparte paragraaf en titel te gebruiken voor "implementatietrouw". Nu moet de lezer zoeken naar de uitleg van de metingen van implementatietrouw.

Eerder was het me niet opgevallen dat implementatietrouw 2x gemeten is, namelijk op T2 en T3. Zou het niet interessant zijn om apart te kijken naar klassen waar zowel op T2 als T3 implementatietrouw gerapporteerd wordt en waarvan men dus kan aannemen dat het programma goed geïmplementeerd is gedurende de gehele interventieperiode? Indien dat niet wordt gedaan, moet in de discussie vermeld worden dat het niet vastgesteld kon worden of implementatietrouw voor de hele interventieperiode gold. (zie ook mijn feedback op de discussie over de beperkingen in de metingen van implementatietrouw).

Resultaten

meerdere wijzigingen zijn nodig

Resultaten - toelichting

-Ondanks dat MLR een robuuste methode is, is het wel regel dat de verdeling van continue variabelen (scheefheid en gepiektheid) opgenomen wordt in tabellen met beschrijvende statistiek (eventueel in bijlage).

-ook is het belangrijk dat de lezer inzage krijgt in de volledige modellen met alle predictoren, dat is inclusief conditie, covariaten (demografische variabelen) en de T1 (baseline) meting .

Aangezien er al veel tabellen zijn, is het te overwegen om dit in bijlage te rapporteren.

-Tabel 4: De kolom met de gestandaardiseerde beta bij T2: zou dit niet p-waarde moeten zijn (zoals bij T3)?

-Tabel 5 presenteert.... voor leerkrachten. Graag dit veranderen in: voor leerkrachtgerapporteerde uitkomsten.

-Graag steeds bij Tabel 6 aangeven of men verwijst naar A of B. Idem voor Tabel 7.

-bij Tabel 7 en 8 gaat het over analyses waarbij de leerkracht aangeeft dat er wel/niet sprake is van implementatietrouw. Waarom werden deze analyses niet voor alle afhankelijke variabelen gedaan? Dat is nodig voor een volledig beeld van de resultaten.

Zijn de resultaten voldoende statisch onderbouwd?

..

Discussie

meerdere wijzigingen zijn nodig

Discussie - toelichting

-In de inleiding wordt de nadruk gelegd op het belang van verandering in sociale norm om een blijvend effect op klimaat te verkrijgen. Aangezien er geen effect van het programma lijkt te zijn op de norm, vraag ik mij af of dat een verklaring kan zijn voor de beperkte effecten in het algemeen en voor de uitdovende effecten bij implementatietrouw (dat is: wel effecten op T2 maar minder op T3) en of dit besproken kan worden in de discussie.

-implementatietrouw is een variabele op klasniveau (een leerkracht implementeert het programma goed of niet in een klas), maar werd gemeten en geanalyseerd op leerlingniveau. Dus wanneer 1 leerling zegt dat de interventie goed geïmplementeerd werd dan verwachten we al een effect voor die leerling. Er had ook gekozen kunnen worden voor een meer objectieve maat waarbij bijvoorbeeld ten minste 50% van de leerlingen in een klas de implementatietrouw als goed beoordelen. Dat had een meer robuuste maat opgeleverd.

-daarnaast valt op dat 1 van de criteria was dat een leerling minimaal 1 compliment moet hebben gekregen. Impliceert dat niet dat er indirect wordt geselecteerd op de uitkomstmaten (prosociaal

gedrag)?

Daarnaast werd de perceptie van leerlingen van implementatietrouw los gezien van de perceptie van leerkrachten van implementatietrouw. En ook werd dit voor meetmomenten afzonderlijk bekeken en niet voor het hele jaar.

Kortom, we weten dus niet of er meer robuuste effecten waren in klassen waar het programma trouw geïmplementeerd werd gedurende het hele jaar en volgens een meerderheid van de deelnemers (of volgens zowel leerkrachten en leerlingen). Deze vragen rond implementatietrouw en welke conclusies we hier nu wel of niet uit kunnen trekken, verdienen meer aandacht in de discussie. Dit maakt ook dat de er geen conclusies getrokken kunnen worden over de effectiviteit van het programma bij implementatietrouw. Ook kan denk ik met dit in gedachten niet gesproken worden van “optimale” implementatietrouw.

-een aantal zinnen begreep ik niet direct: 1) kunnen jullie toelichten wat jullie bedoelen met lage variabiliteit en hoe dit kan leiden tot nulresultaten? Was er misschien sprake van een plafond effect? 2) het is mij niet duidelijk wat jullie precies “hebben gezien in onze resultaten” (p.31). Dit vraagt nog wat toelichting zodat de lezer jullie goed kan volgen.

-Het leest wat vreemd dat de effecten die gevonden werden bij programmatrouw veroorzaakt worden door motivatieproblemen, aandachtsproblemen, Het zou logischer zijn wanneer het uitdoven van deze effecten door deze zaken veroorzaakt wordt. Misschien is dat ook wat jullie bedoelen?

-Kunnen jullie nog kort toelichten waarom er op T3 minder power was (p.32)

Conclusies

goed/voldoende

Conclusies - toelichting

In zijn algemeenheid zou ik vragen goed op te letten dat er niet gesuggereerd wordt dat het programma effectief zou zijn wanneer implementatietrouw hoog genoeg is. Die conclusie komt nog te vroeg (zie ook eerdere opmerkingen bij de meting van implementatietrouw)

Tabellen en figuren

Volgens APA staan de afhankelijke variabelen boven de kolommen. In deze tabellen is dat niet het geval. Dat kan verwarrend zijn voor de lezer. Kan dit ergens toegelicht worden?

Tabel 6b: bij de laatste kolom is het informatiever om ja/nee te vervangen door de p-waarde van de toets

Tabel 6B, voetnoot b: hoe is dit getoetst?

Referenties

de referentielijst ziet er goed uit, maar niet alle referenties zijn volgens APA (sommige titels begint elke woord met een hoofdletter)

Ruimte voor kleine opmerkingen

Ik denk dat het nodig is om de tekst nog eens goed na te lezen. In sommige gevallen zijn de zinnen lang of is er sprake van spreektaal (bv. “beter af zijn”).

In zijn algemeenheid heb ik de indruk dat de auteurs het artikel heel grondig herwerkt hebben. Er is veel informatie en uitleg toegevoegd en de resultaten worden met de nodige zorgvuldigheid en nuance geïnterpreteerd. Ook in de discussie zijn een aantal belangrijke reflecties toegevoegd. Ik zou dan ook graag een positief advies geven in het vertrouwen dat ook de huidige feedback goed verwerkt zal worden.

Algemeen oordeel

revisie

Reviewer 2:

Aanbeveling: Revisies benodigd

Titel

Ja

Abstract

meerdere wijzigingen zijn nodig

Abstract - toelichting

Er staat nu dat voor een kleine subgroep die ideale programmatrouw rapporteerde gold dat de leerlingen uit de interventieconditie beter af waren dan de leerlingen uit de controleconditie. Twee opmerkingen hierover: (1) Dit is een groepsgemiddelde dus graag aangeven dat leerlingen gemiddeld genomen beter af waren; (2) Ik mis dat ook voor deze groep de gevonden effecten veelal slechts klein waren en kortdurend (wel na 6 maanden maar niet meer na 12 maanden). Dit graag toevoegen en hierop ook de 1-na-laatste concluderende zin aanpassen.

De zin dat leerkrachten een toegenomen positieve houding tov stimulering van prosociaal gedrag en autonomie lieten zien is wat onduidelijk. Er staat niet duidelijk genoeg dat het gaat om hun eigen positieve houding (en dus niet die van de leerlingen) en als ik het goed begrijp uit de rest van het manuscript was die toename niet te zien in de groep met ideale programmatrouw maar dat wordt hier nergens genoemd. Dit laatste graag toevoegen of de hele zin weglaten uit de samenvatting.

Deze opmerkingen ook graag toepassen in de Engelstalige samenvatting, en ook in de bijsluiter en de conclusie.

Inleiding

goed/voldoende

Inleiding - toelichting

De inleiding kan qua structuur nog verbeterd worden. Het begint nu met beschrijving van meaningful roles en de interventie die gebruikt is in de huidige studie en pas daarna wordt ingegaan op vergelijkbare Nederlandse interventies en pas daarna op internationale nog bredere interventies. Qua opbouw zou het logischer zijn van breed naar smal te werken en de eigen specifieke interventie pas aan het einde van de inleiding te beschrijven.

Specifiekere punten:

- alinea 3 is in de huidige vorm echt te lang, dus graag opsplitsen
- 'een schoolbreed programma met rollen' (p. 4) - rollen kunnen hier beter weg want wordt in de volgende zin uitgelegd. Is nu dubbelop en pas te begrijpen na lezing van de volgende zin.
- veiligheidsingenieurs -> veiligheidsingenieur
- in publiek -> te Engels, beter om voor iedereen zichtbaar oid te gebruiken
- 'focust zich niet specifiek op pesters, maar op de gehele groep' -> wordt hier de hele klas bedoeld?
- En wat houdt de aanpassing van het programma voor basisschoolleerlingen in? En wie heeft deze aanpassingen gemaakt? Het zou goed zijn in het supplemental material een tabel of beschrijving toe te voegen van de aanpassingen die gemaakt zijn t.o.v. meaningful roles.
- Deze zin loopt niet: "Eerder onderzoek laat zien dat deze (losse) componenten positieve effecten op verantwoordelijkheid, competentie, verbondenheid, interpersoonlijke vaardigheden, erkenning, prosociaal gedrag, en probleemgedrag (Battistich, 2005; Edwards & Mullis, 2003; Ellis et al., 2016; Embry, 2004; Embry & Biglan, 2008; Gfroerer et al., 2024)." Ook zijn 3 refs wel voldoende.
- Het is een aanpak [...] wat uitgevoerd wordt -> die uitgevoerd wordt
- Kunnen ook wetenschappelijke referenties worden toegevoegd in de alinea over de werkzaamheid van Nederlandse interventies?
- Het valt op dat er zowel voor de Nederlandse als voor de internationale interventies alleen studies worden aangehaald die aanwijzingen hebben gevonden dat de interventies werkten. Er zijn dus helemaal geen studies die hebben gevonden dat een interventie niet of minder goed werkte?
- (3) hoger prosociaal gedrag -> toename van prosociaal gedrag
- (6) een toegenomen positieve houding van leerkrachten ten opzichte van pesten -> graag herformuleren want ik neem aan dat leerkrachten niet ineens pesten als iets positiefs zijn gaan zien?
- Als tweede doel hebben we ook de effecten van programmatrouw bepaald -> Verderop worden dit sensitiviteitsanalyses genoemd. Graag in de inleiding ook deze term gebruiken.

Methode

een kleine wijziging is nodig

Methode - toelichting

De methode is over het algemeen duidelijk en goed beschreven.

Specifieke punten:

- (net zoals bij anderen die effectieve anti-pestprogramma's implementeren). -> net zoals bij andere scholen
- In par 2.2 wordt Indonesië onder Westerse landen geschaard. Dit is wat verwarrend. Graag aanpassen of uitleg toevoegen.
- MLR corrigeert/is robuust tegen afwijkingen van niet-normaliteit en niet-onafhankelijkheid -> graag 'corrigeert' weglaten en wat bedoel je in deze context met niet-onafhankelijkheid en waarom verwacht je dit binnen jullie studie?
- Effectgroottes werden berekend door middel van een online programma -> dit is een beetje vaag.
- Voor programmatrouw is ook meegenomen of studenten tenminste een compliment per week ontvingen. Waarom is gekozen voor ontvingen ipv voor gegeven? En zou dit niet de resultaten iets te positief kunnen beïnvloeden omdat dit toch eigenlijk al een teken kan zijn voor prosociaal gedrag? Was het niet beter/zuiverder geweest programmatrouw op het niveau van de klas te bepalen? Dit punt graag sowieso bespreken in de discussie.

Resultaten

een kleine wijziging is nodig

Resultaten - toelichting

3.2: graag wat duidelijker opschrijven wat hiermee bedoeld wordt: "Voor leerkrachthouding over de stimulatie van basisbehoeften op T3 werden significante effecten van het programma gevonden voor het stimuleren van prosociaal gedrag en autonomie." Wat betekent dit nu precies? Dat leerkrachten vergeleken met baseline op T3 aangaven prosociaal gedrag en autonomie van hun leerlingen meer te stimuleren? En graag ook expliciet noemen dat er op T2 nog geen verbetering zichtbaar was.

3.3: Ik mis dat er eigenlijk alleen kleine effecten zijn gevonden en ook vind ik de bespreking van T2 en T3 resultaten in de huidige vorm niet zo inzichtelijk. Mijn suggestie zou zijn om dit meer systematisch te bespreken, bijv.

- 1) de primaire en secundaire uitkomsten die helemaal geen effect lieten zien
- 2) de primaire en secundaire uitkomsten die een effect op T2 lieten zien en dan ook aangeven welk van deze uitkomsten nog steeds een sig effect op T3 liet zien en welke uitkomsten niet + hoe groot al deze effecten waren
- 3) Welke primaire en secundaire uitkomsten pas op T3 een effect lieten zien + hoe groot al deze effecten waren.

Het zou goed zijn deze structuur ook in de discussie te laten terugkomen.

Verzoek aan de auteurs is om te zorgen dat naar alle tabellen verwezen wordt in de tekst (voor Tabel 9a zie ik bijvoorbeeld dat dit niet het geval is)

Zijn de resultaten voldoende statisch onderbouwd?

Ja

Discussie

meerdere wijzigingen zijn nodig

Discussie - toelichting

De discussie sluit veel beter aan bij de resultaten dan in de eerste versie. Wel valt op dat niet genoemd wordt dat ook de effecten die gevonden zijn voor de groep met ideale implementatie veelal slechts klein waren en ook kortdurend. Dit laatste wordt onderaan p. 31 wel genoemd maar het zou beter zijn dit ook al te noemen in de eerste alinea.

Andere punten:

- "Leerkrachten die [gemaskeerd] hadden geïmplementeerd rapporteerden een positievere houding ten opzichte van het stimuleren van prosociaal gedrag en autonomie." -> graag duidelijker opschrijven. Nu staat er bijv. niet expliciet dat het om hun eigen positieve houding gaat en niet om die van de leerlingen.
- "Wanneer leerkrachten aangaven dat ze de interventie hadden uitgevoerd zoals bedoeld, werden er geen significante effecten gevonden." -> graag een reflectie van de auteurs hierop. Hoe zou dit kunnen komen? Waren al al op baseline verschillen tussen leerkrachten met lage en hoge programmatrouw op het gebied van hun houding tov het stimuleren van prosociaal gedrag en autonomie? Kan het zijn dat dit er nu juist voor gezorgd heeft dat deze leerkrachten de interventie goed doorvoerden?
- In alinea 2 zou het feit dat het KIVA scholen zijn misschien ook kunnen meespelen
- "Ten slotte is het ook mogelijk dat er minder effecten op T3 werden gevonden vanwege een lagere power." -> graag even toelichten waarom er een lagere power was op T3 en hoe groot het verschil dan was tussen de steekproefgroottes van T2 en T3. Ook gaat dit argument alleen op als er niet significante resultaten waren maar de beta coëfficiënten van de T3 analyses nog steeds wel aanleiding geven om te denken dat er een effect is, bijvoorbeeld als ze vergelijkbaar zijn met de beta coëfficiënten op T2.

Limitaties:

- "Voor de analyses met ideale programmatrouw rapporteerden leerlingen met hoge programmatrouw over het algemeen hogere en positievere scores dan leerlingen met lage programmatrouw of leerlingen in de controlegroep." -> graag nog iets duidelijker toevoegen dat dit verschil er ook al op baseline was want ik moest dit stukje een paar keer lezen voordat het duidelijk was.
- "Echter waren de verschillen op baseline klein, en is het in lijn met eerder onderzoek dat laat zien dat hogere levels op baseline geassocieerd zijn met sterkere effecten (e.g., Sterneklar et al.,

2019)." -> wordt hier sterke vertekening van effecten bedoeld ipv sterkere effecten
- Was het niet beter/zuiverder geweest programmatrouw op het niveau van de klas te bepalen?
Nu zou een kritiekpunt kunnen zijn dat effecten van deze subgroep mogelijk vertekend zijn doordat leerlingen zijn geselecteerd die zelf complimenten ontvingen.
- Beperkingen van de zelfgeconstrueerde vragenlijst: het zou toch ook kunnen dat dit heeft geresulteerd in kleinere effecten? Bijv als de interne consistentie van de items laag was.

Praktische implicaties:

- Ook hier moet worden genuanceerd dat de positieve effecten maar klein en kortdurend waren

Niet lopende zin: "Ondanks deze mogelijke biases zijn de deelnemende interventie- en controlescholen waren verspreid over heel Nederland"

Conclusies

een kleine wijziging is nodig

Conclusies - toelichting

Graag duidelijk aangeven dat ook de ideale programmatrouw leerlingen maar iets beter af waren en dat dit maar kortdurend was.

Tabellen en figuren

De tabellen en figuren zijn duidelijk en informatief. Het zijn er wel wat veel geworden, dus de auteurs zouden kunnen overwegen een aantal tabellen naar het supplemental material te verplaatsen.

Referenties

Ruimte voor kleine opmerkingen

Algemeen oordeel

revisie

Mail met reviews over de derde versie

Van: "Jochum Veenstra" <jeugdinatorontwikkeling@outlook.com>

Aan: "Amanda van Loon" <a.w.g.vanloon@uu.nl>

Onderwerp: Redacteursbeslissing

Geachte Amanda van Loon, Tessa Kaufman,

Uw artikel "Themanummer Wat is er nodig voor het creëren van prosociale klassen? De effectiviteit van een schoolinterventie voor basisschoolkinderen." is beoordeeld door twee leden van de redactie.

Een van de twee reviewers heeft nog een paar kleine verbeterpunten die vooral over formuleringen gaan. Zou u de opmerkingen van deze reviewer willen verwerken en mij een versie willen sturen waarin u de maskeringen opheft? Mochten eerdere opmerkingen tegenstrijdig zijn met de opmerkingen die in de review van reviewer 1 staan dan wil ik u vragen om de opmerkingen van reviewer 1 als leidend op te vatten.

De reviewer uit de themareadactie kijkt nog een laatste keer mee naar dit niet-geanonimiseerde review en vervolgens zal ik het zo snel mogelijk uitzetten bij de corrector.

Zou u over twee weken de nieuwe versie willen uploaden? De deadline voor de nieuwe versie is: 9 dec

Voeg bij u nieuwe versie zowel een antwoordbrief als een versie van het manuscript mét track and trace (of arceringen). NB: ik zou u willen vragen om de maskeringen alvast op te heffen.

Onder deze mail kan u de reviews lezen. Een nieuwe versie kan u indienen via jeugdinatorontwikkeling.nl

Zou u de ontvangst van deze mail willen bevestigen door te mailen naar: jeugdinatorontwikkeling@outlook.com

Mocht u de revisie niet binnen twee weken kunnen schrijven zou u dan contact met ons willen opnemen via: jeugdinontwikkeling@outlook.com

Met vriendelijke groeten,
Jochum Veenstra (redactie-secretaris)

Reviewer 1:

Aanbeveling: Revisies benodigd

Titel

ja

Abstract

een kleine wijziging is nodig

Abstract - toelichting

In de laatste zin van de samenvatting, de bijsluiters en de conclusie wordt aangegeven dat het belangrijk is om meer inzicht te krijgen over programma-implementatie omdat programmatrouw belangrijk lijkt te zijn voor effectiviteit. Het is natuurlijk prima om dit te noemen maar het klinkt nu wel alsof de sleutel tot succes van dit programma programmatrouw is. Om niet voorbij te gaan aan het punt dat er zelfs bij ideale programma trouw slechts kleine en kortdurende effecten zijn gevonden zou ik op alle drie de plekken ook toevoegen dat het daarnaast ook belangrijk is om meer inzicht te krijgen in waarom de effecten zelfs bij ideale programmatrouw klein en kortdurend waren.

Inleiding

een kleine wijziging is nodig

Inleiding - toelichting

“Als tweede doel hebben we ook de effecten van programmatrouw bepaald – wanneer een programma wordt geïmplementeerd zoals bedoeld – waardoor onderzoekers accurate conclusies

kunnen trekken over de effectiviteit van een programma (Ma et al., 2023; Sanetti et al., 2021).” (p7-8) -> Dit impliceert dat het accurater is om te kijken naar ideale programmatrouw dan naar intention-to-treat. Dit is wat verwarrend. Kan dit weg? ‘waardoor onderzoekers accurate conclusies kunnen trekken over de effectiviteit van een programma (Ma et al., 2023; Sanetti et al., 2021).’ Het stukje is ook niet echt nodig want de zin erna is wel duidelijk en to-the-point en geeft al heel goed het belang van deze analyses aan.

Methode

meerdere wijzigingen zijn nodig

Methode - toelichting

Op p. 13 wordt verwezen naar appendix A. Graag iets uitgebreider in 1 zin aangeven wat er in appendix A te vinden is.

Voor tabellen in de appendix graag een letter aan de tabelnummering toevoegen, voor tabel 1 uit appendix B graag Tabel B1 ipv Tabel 1. Graag overal doorvoeren.

Er is op p. 18 informatie over scheefheid en gepiektheid toegevoegd. Dit is nuttige informatie maar past niet goed binnen de paragraaf Randomisatie en drop-outs. Graag ergens anders noemen. En ook graag expliciet aangeven of dit voor pesten en slachtofferschap problemen geeft voor de analyses (ik neem aan van niet vanwege MLR).

Resultaten

meerdere wijzigingen zijn nodig

Resultaten - toelichting

Ik vind de bespreking van de uitkomsten van ideale programmatrouw in 3.3 nog steeds wat warrig. Het zou erg helpen om het iets verhalender op te schrijven. Bijv. Bij ideale programmatrouwe werden de volgende effecten van primaire uitkomstmaten gevonden De meeste van deze effecten waren maar kortdurend en waren op T3 weer verdwenen, met uitzondering van effect x dat op zowel T2 als T3 werd gevonden. Er zijn ook een aantal effecten gevonden die wel op T3 maar nog niet op T2 aanwezig waren, namelijk.... Ik vind dat de lezer nu nog iets te hard moet werken in deze paragraaf. Hetzelfde geldt voor de beschrijving van de uitkomsten voor de secundaire maten.

Ik zou alle informatie over de covariaten voor de leesbaarheid verplaatsen naar de bijlage en daarnaar verwijzen in de hoofdtekst.

Zijn de resultaten voldoende statisch onderbouwd?

ja

Discussie

een kleine wijziging is nodig

Discussie - toelichting

“Alle effecten waren klein en voornamelijk halverwege het schooljaar (T2).” -> graag hier al expliciet aangeven dat alle effecten klein en voornamelijk kortdurend waren

Ik vind het in de discussie juist wat verwarring oproepen dat wordt gesproken over het 'ervaren' van optimale programmatrouw. Dit is namelijk niet wat letterlijk is uitgevraagd. Ik zou dit in de discussie oplossen door op p. 31 even 1 keer wat meer informatie te geven en daarna gewoon te spreken over optimale programmatrouw. Kan bijv bij deze zin: “Als laatste is een mogelijke reden dat de interventie niet geïmplementeerd is zoals bedoeld, wat een belangrijke voorwaarde kan zijn voor de effectiviteit van een interventie (e.g., Durlak & DuPre, 2008), wat we ook hebben gezien in onze resultaten (i.e., alleen effecten in de subgroep van leerlingen die ervaarden dat het programma optimaal was geïmplementeerd).” -> alleen effecten bij leerlingen met ideale programmatrouw, dwz leerlingen die zelf aangaven dat x, y, en z.

“ondanks een positieve trend voor de meeste uitkomsten bij de groep leerkrachten met hoge programmatrouw” -> ik vind het niet zo'n goed idee om over positieve trend te spreken bij een studie waarin al zo ontzettend veel analyses zijn gedaan en effecten zo klein zijn. Algemene opmerking mbt alle suggesties dat power een rol kan spelen: Is power hier echt van belang zodat nog kleinere effecten kunnen worden gevonden? En als verklaring voor verschillen tussen T2 en T3 zou het nog kunnen, maar dan moeten de beta's wel echt vrijwel identiek zijn en niet toch wel wat lager zoals in veel gevallen zo is. Een verschil tussen .08 en .14 (zoals genoemd in de response op reviewer 2 punt 11D) is daarvoor echt te groot.

Conclusies

een kleine wijziging is nodig

Conclusies - toelichting

“Leerkrachten rapporteerden een positievere (eigen) houding ten opzichte van het stimuleren van prosociaal gedrag en autonomie.” -> suggestie voor leesbaarheid: Leerkrachten rapporteerden wel een positievere eigen houding ten opzichte van het stimuleren van prosociaal gedrag en autonomie

Laatste zin conclusie: “Het is belangrijk om meer inzicht te krijgen over programma-implementatie, aangezien de mate van programmatrouw belangrijk lijkt te zijn voor programma-effectiviteit.”

Inzicht over -> inzicht in

Lijkt -> lijkt te

Om niet voorbij te gaan aan het punt dat er zelfs bij ideale programma trouw slechts kleine en

kortdurende effecten zijn gevonden zou ik ook toevoegen dat het daarnaast ook belangrijk is om meer inzicht te krijgen in waarom de effecten zelfs bij ideale programmatrouw klein en kortdurend waren.

Tabellen en figuren

Goed om nog even alle tabeltitels bij langs te gaan. Ik vind ze wel duidelijker geworden maar zou voor tabellen die alleen descriptieve statistieken en verschillen op baseline weergeven liever niet regressieanalyse (zoals bijv in tabel 9A) of Analyse (overal) gebruiken. Iets als Beschrijvende statistieken en verschillen op T1....etc is wat mij betreft duidelijk.

Referenties

Ruimte voor kleine opmerkingen

Aanpak wat -> aanpak die (komt aantal keer voor in het manuscript)

De figuren missen bij de laatste versie, reminder om niet te vergeten die te uploaden

Een of meerdere -> leestekens missen

Algemeen oordeel

revisie

Reviewer 2:

Aanbeveling: Accepteer inzending

Titel

Abstract

goed/voldoende

Abstract - toelichting

Inleiding

goed/voldoende

Inleiding - toelichting

Methode

goed/voldoende

Methode - toelichting

Resultaten

goed/voldoende

Resultaten - toelichting

Zijn de resultaten voldoende statisch onderbouwd?

Discussie

goed/voldoende

Discussie - toelichting

Conclusies

goed/voldoende

Conclusies - toelichting

Tabellen en figuren

Referenties

Ruimte voor kleine opmerkingen

Algemeen oordeel

accepteren