



**rijksuniversiteit
groningen**

Masterscriptie

*De betrouwbaarheid en validiteit van leerlingrapportages
over het functioneren van KiVa-leerkrachten*

Naam:	Lennard Terpstra
Studentnummer:	S4701631
E-mailadressen:	l.terpstra.4@student.rug.nl
Masterroute:	Sociologie – Criminaliteit & Veiligheid
Docentbegeleider:	Dr. G. E. (Gijs) Huitsing
Referent:	Dr. J. M. E. (Mark) Huisman
Tijdspanne scriptie:	Februari 2023 – Maart 2024

Voorwoord

In deze scriptie neem ik u graag mee door het onderzoek van mijn masterscriptie voor de opleiding Sociologie – Criminaliteit & Veiligheid. Mijn dank gaat uit naar KiVa, die mij de kans gaf om een onderzoek voor hen uit te voeren met bestaande data. Dit gaf me de mogelijkheid om meer ervaring op te doen als onderzoeker en heeft mijn expertise betreft pestgedrag en onderwijs enorm vergroot. Daarnaast gaat mijn dank uit naar Gijs Huitsing, die mij heeft begeleid en ondersteund tijdens de gehele periode. Tot slot bedank ik Mark Huisman voor zijn rol als referent en zijn ondersteuning tijdens de analytische fase.

Veel leesplezier,
Lennard Terpstra

Samenvatting

Pestgedrag vormt al decennialang een aanzienlijk maatschappelijk probleem op basisscholen, met een afname in prevalentie onder Nederlandse scholieren tussen 2000 en 2018, maar een zorgelijke stabilisatie van de cijfers sindsdien. KiVa is een effectief en theoretisch onderbouwd antipestprogramma dat dit probleem tegen probeert te gaan.

In dit onderzoek wordt de betrouwbaarheid en validiteit van leerlingrapportages over het leerkracht-functioneren onder de loep genomen en onderzocht of de beoordeling van het leerkracht-functioneren samenhangt met het welbevinden en slachtofferschap van leerlingen in de klas. De steekproef bestond uit zowel KiVa-leerkrachten als leerkrachten uit de controlegroep. Deze leerkrachten werden onderworpen aan een analyse met behulp van t-toetsen, waarbij gekeken werd naar het geslacht van de leerkracht, het geslacht van de beoordelende leerling, de toegewezen lesgroep, de aanwezigheid van meerdere leerkrachten in de groep en of de leerkracht werkzaam is op een KiVa-school. Om te onderzoeken of de beoordeling van het leerkracht-functioneren samenhangt met het welbevinden en slachtofferschap, zijn er ANOVA- en regressieanalyses uitgevoerd.

De resultaten tonen aan dat vrouwelijke leerkrachten hogere beoordelingen ontvangen dan mannen, meisjes leerkrachten hoger beoordelen dan jongens, en onderbouwleerkrachten over het algemeen beter worden beoordeeld op bekwaamheid en inspanning. Daarnaast blijkt dat KiVa-leerkrachten hogere beoordelingen ontvangen en zich beter ontwikkelen ten opzichte van de controlegroep. Tot slot blijkt dat leerkrachten positiever worden beoordeeld in klassen met een hoog welbevinden, maar er is geen directe relatie tussen het niveau van slachtofferschap en de beoordeling van het leerkracht-functioneren gevonden.

Voorzichtigheid is geboden bij het interpreteren van de aanbevelingen, gezien de bevindingen zijn verkregen uit een steekproef waarbij het verloop van leerkrachten groot was. De resultaten geven echter wel een sterke indicatie van de validiteit en betrouwbaarheid van leerlingrapportages over het functioneren van KiVa-leerkrachten. Dit onderzoek vormt een startpunt voor een diepgaander begrip van de complexiteit van leerlingrapportages over het functioneren van KiVa-leerkrachten, waarbij de resultaten als aanleiding dienen voor verdere optimalisatie en verkenning van de nauwkeurigheid van leerkrachtbeoordelingen.

Inhoudsopgave

1. Inleiding.....	6
1.1 Onderwerp en aanleiding.....	6
1.1.1 Definiëring pestgedrag	6
1.1.2 Gevolgen pestgedrag	6
1.1.3 Aanpak pestgedrag	7
1.2 Specificatie van het onderzoek	8
1.3 Wetenschappelijke relevantie	9
1.4 Sociologische relevantie	9
2. Bestaande theoretische inzichten	10
2.1 Beoordeling leerkracht-functioneren	10
2.1.1 Geslacht van de leerkracht	11
2.1.2 Geslacht van de beoordelende leerling.....	11
2.1.3 Groep van de leerkracht	12
2.1.4 Aantal leerkrachten voor de klas	13
2.1.5 KiVa-conditie van de leerkracht.....	14
2.1.6 Welbevinden van de leerling	15
2.1.7 Slachtofferschap van de leerling	16
2.2 Deze studie: conceptueel model	17
3. Methode.....	18
3.1 Data	18
3.2 Deelnemers	19
3.3 Operationalisaties	20
3.3.1 Leerkracht-functioneren	20
3.3.2 Welbevinden.....	20
3.3.3 Slachtofferschap	21
3.4 Analyse-opzet	22
3.4.1 Verschillen tussen leerkrachten.....	22
3.4.2 Ontwikkeling beoordeling leerkrachten	22
3.4.3 Samenhang klassenklimaat.....	23
4. Resultaten.....	24
4.1 Verschillen tussen leerkrachten	24
4.1.1 Wave 1.....	24
4.1.2 Wave 3.....	27

4.1.3 Wave 5.....	28
4.2 Ontwikkeling beoordeling leerkrachten	28
4.2.1 Univariate statistieken	28
4.2.2 Beoordelingen categoriseren	29
4.2.3 Stijgende en dalende beoordelingen	32
4.3 Samenhang klassenklimaat.....	32
4.3.1 Univariate & Bivariate statistieken	32
4.3.2 Welbevinden in de klas	33
4.3.3 Slachtofferschap in de klas.....	34
5. Conclusie	35
5.1 Terugkoppeling hypotheses.....	36
5.2 Discussie.....	38
5.2.1 Interpretatie resultaten.....	40
5.3 Beperkingen & Aanbevelingen	40
Literatuurlijst	44
Bijlage I. Grafische weergave variabelen leerkracht-functioneren	53
Bijlage II. Grafische weergave ontwikkeling leerkracht-functioneren	55
Bijlage III. ANOVA-analyse en regressieanalyse.....	56

1. Inleiding

1.1 Onderwerp en aanleiding

Pestgedrag is al decennialang een groot maatschappelijk probleem op basisscholen (Bradshaw et al., 2007; Craig & Pepler, 2003; Nansel et al., 2001; Swearer et al., 2010). Hoewel er gedurende de periode van 2000 tot 2018 een afname was te constateren in de prevalentie van pestgedrag onder Nederlandse scholieren, zijn de cijfers sindsdien gestabiliseerd (Stevens et al., 2018; Ministerie van Onderwijs, Cultuur en Wetenschap, 2022). In het basisonderwijs blijkt 10 tot 15 procent van de scholieren slachtoffer te zijn van een vorm van pesten; oftewel minimaal 1 op de 10 scholieren geeft aan gepest te worden (Ince, 2022). Er kunnen echter vraagtekens bij deze cijfers worden geplaatst omdat de cijfers slechts een indicatie geven van de omvang van het probleem. Pesten is namelijk een complex fenomeen wat op verschillende manieren kan worden geuit en ervaren (Bradshaw et al., 2007; Swearer et al., 2010). Zo wordt pestgedrag en agressief gedrag nog wel eens onder dezelfde noemer geschaard door scholieren, terwijl dit niet altijd het geval is. Desondanks is pesten wél een vorm van agressief gedrag (Vermande et al., 2015).

1.1.1 Definiëring pestgedrag

Om pestgedrag te onderscheiden van ander gedrag (zoals agressief gedrag), zijn er drie basiskennmerken opgesteld waaraan we pestgedrag mogen toeschrijven. Allereerst is het de intentie van de dader(s) om het slachtoffer fysiek, mentaal of emotioneel letsel toe te brengen. Ten tweede vindt het toebrengen van fysiek, mentaal of emotioneel letsel herhaaldelijk plaats over een uitgebreide tijdsspanne. Ten derde is er meestal sprake van een machtsverschil tussen de dader(s) en het slachtoffer (Salmivalli & Peets, 2009). Er zijn diverse vormen van pestgedrag mogelijk, zoals verbaal- (denk aan uitschelden, uitlachen, bedreigen etc.), fysiek- (denk aan spugen, schoppen, slaan etc.), materieel- (denk aan spullen kapot maken, afpakken, verstoppen etc.), relationeel (denk aan buitensluiten, roddelen, geruchten verspreiden etc.) en digitaal pesten (denk aan belastend materiaal verspreiden op het internet etc.) (Reijntjes et al., 2013; Salmivalli & Peets, 2009).

1.1.2 Gevolgen pestgedrag

Pestgedrag heeft een breed scala aan negatieve gevolgen voor de personen die betrokken zijn bij het pesten (Arseneault et al., 2010; Salmivalli & Peets, 2009; Ttofi et al., 2011). Zo kunnen slachtoffers van pesten te maken krijgen met psychische klachten zoals angst,

depressie en stress. Deze psychische klachten leiden in het ergste geval zelfs tot (het uitvoeren van) suïcidale gedachten (Klomek et al., 2010). Naast de slachtoffers van pesten, ondervinden ook de daders negatieve gevolgen van hun gedrag op zowel korte- als lange termijn. Op korte termijn kunnen de daders kampen met tegenvallende schoolprestaties (Ttofi et al., 2011). Tevens hebben de daders vaak moeite om empathie op te brengen voor leeftijdsgenoten waardoor ze hun negatieve gedrag nauwelijks aanpassen. Dit kan op lange termijn resulteren in een verhoogd risico op delinquent gedrag (Nansel et al., 2004; Ttofi et al., 2011). De daders dragen er echter ook aan bij dat de klasgenoten, die niks met het pesten te doen hebben, en leerkrachten last hebben van het pestgedrag. De niet-betrokken klasgenoten zijn bang om het nieuwe slachtoffer van de daders te worden waardoor ze zich onveilig voelen in de klas (Espelage & Holt, 2013). Ook de leerkrachten voelen zich onveiliger, gestrester en angstiger omdat ze last hebben van een lager welbevinden (Espelage & Holt, 2013; Mishna et al., 2008).

1.1.3 Aanpak pestgedrag

Om een effectieve preventie- en interventiestrategie te ontwikkelen waardoor bovenstaande negatieve gevolgen worden gereduceerd, is het van belang om rekening te houden met de complexiteit van pesten. Bij het ontwerpen van preventie- en interventiestrategieën tegen pesten moeten factoren zoals persoonlijke kenmerken, sociale relaties en omgevingsfactoren in acht worden genomen (Olweus, 1993; Zych et al., 2019). Een bewezen effectief en theoretisch onderbouwd antipestprogramma in Nederland is KiVa wat gekopieerd is vanuit Finland. KiVa doet er alles aan om bewustwording te creëren bij leerlingen en leerkrachten omtrent het onderwerp pesten en leren leerlingen en leerkrachten adequaat te reageren bij pestgedrag. Dit wordt gedaan middels trainingen voor leerkrachten, KiVa-lessen en een computerspel voor leerlingen en monitoring voor scholen. Tijdens deze elementen van KiVa staan twee aspecten centraal (Salmivalli et al., 2010). Eén van deze aspecten is de rolbenadering bij pesten. De rolbenadering laat zien dat pesten in de klas een groepsproces is waarin alle personen een verschillende rol aannemen; oftewel alle leerlingen, in verschillende rollen, houden het pestprobleem in stand (Veenstra & Huitsing, 2021). Het KiVa-programma richt zich daarom niet alleen op de pester en het slachtoffer maar ook op de buitenstaander (diegene die ziet dat er wordt gepest maar niets onderneemt), versterker (diegene die de pester(s) aanmoedigt), assistent (diegene die met de pester(s) meedoet) en verdediger (diegene die voor het slachtoffer opkomt) (Veenstra & Huitsing, 2021). Het tweede element is de goede sociale positie die pesters vaak innemen. Pesters verwerven door het pestgedrag een

hoge(re) status waardoor ze over een sterkere positie beschikken dan het slachtoffer (Sentse et al., 2015). KiVa probeert deze baten van de pesters te reduceren door leerlingen en leerkrachten te stimuleren om op te komen voor het slachtoffer.

Een ander bijkomend positief effect van het KiVa-programma is de beïnvloeding van de signalering en waarneming van leerkrachten (Huising et al., 2020). Door de trainingen die leerkrachten krijgen aangeboden, worden zij aangemoedigd om negatiever te denken over pesten. Indien leerkrachten over de juiste houding, vaardigheden en hulpmiddelen beschikken, gaan ze effectiever om met pestincidenten en neemt het pestgedrag af in de klas (Veenstra et al., 2014). Het leerkracht-functioneren kan worden beoordeeld aan de hand van leerlingrapportages. De leerlingrapportages bevatten vragen en stellingen over de attitude, bekwaamheid en inspanning van leerkrachten.

1.2 Specificatie van het onderzoek

In dit onderzoek wordt de validiteit en betrouwbaarheid van de leerlingrapportages over leerkrachten onder de loep genomen. Validiteit verwijst naar de mate waarin een meetinstrument daadwerkelijk meet wat het beoogt te meten (Cronbach & Meehl, 1955). Als leerlingrapportages niet valide zijn, kan het zijn dat ze niet de juiste informatie verschaffen over het leerkracht-functioneren. Dit ondermijnt de geldigheid van de leerlingrapportages. Betrouwbaarheid verwijst naar de consistentie en stabiliteit van de meetinstrumenten (Caliński et al., 1990). Als leerlingrapportages niet betrouwbaar zijn, kunnen ze leiden tot inconsistente resultaten wat de interpretatie van resultaten bemoeilijkt. De validiteit en betrouwbaarheid zijn daarnaast ook essentieel bij het beoordelen en verbeteren van het KiVa-rapport.

Daarnaast wordt onderzocht of de beoordeling van het leerkracht-functioneren samenhangt met het welbevinden en slachtofferschap van leerlingen in de klas. De beoordeling van leerlingen over het leerkracht-functioneren kunnen worden beïnvloed door het klassenklimaat. Het is aannemelijk dat positieve beoordelingen meer waarschijnlijk zijn in situaties waarin er weinig sprake is van pestgedrag en het welbevinden van leerlingen hoog is. Aan de andere kant lijken negatieve beoordelingen vaker voor te komen in situaties waarin pestgedrag frequent voorkomt en het welbevinden van leerlingen laag is. Indien vastgesteld wordt dat het slachtofferschap en het welbevinden van leerlingen in de klas daadwerkelijk van invloed zijn op de beoordeling van het leerkracht-functioneren, zou dit de interpretatie van de beoordelingen makkelijker maken omdat er een meer holistisch beeld ontstaat.

Het overkoepelende doel van dit onderzoek is inzichten krijgen over welke informatie wel of niet moeten verschijnen in het KiVa-rapport. De probleemstelling die in dit onderzoek centraal staat is: “*Wat is de betrouwbaarheid en validiteit van leerlingrapportages over het functioneren van KiVa-leerkrachten?*”.

1.3 Wetenschappelijke relevantie

Dit onderzoek is een vervolg op het artikel *The Role of Teachers in Bullying: The Relation Between Antibullying Attitudes, Efficacy, and Efforts to Reduce Bullying* van René Veenstra en collega's (2014) en geeft invulling aan twee wetenschappelijke kloven. Het artikel concludeert dat indien leerkrachten over de juiste houding, vaardigheden en hulpmiddelen beschikken, ze een primaire rol spelen bij het voorkomen en reduceren van pestgedrag (Veenstra et al., 2014).

Allereerst wordt de betrouwbaarheid en validiteit van de leerlingrapportages vastgesteld, wat in het vorige onderzoek niet gedaan is, waardoor de bevindingen geloofwaardiger en aanneembaarder zijn. Zoals hierboven staat beschreven is het vaststellen van de validiteit en betrouwbaarheid essentieel bij het beoordelen en verbeteren van het KiVa-rapport. Ten tweede worden de analyses uitgebreid door het algemene welbevinden en slachtofferschap van leerlingen mee te nemen. In tegenstelling tot het vorige artikel, dat zich uitsluitend richtte op het, door de leerling-gerapporteerde, pestgedrag (Veenstra et al., 2014). De huidige literatuur erkent dat het niveau van het algemene welbevinden een grote invloed heeft op de beoordeling van leerkrachten (Jones et al., 2013). Bovendien kan worden aangenomen dat het niveau van slachtofferschap van invloed is op de beoordeling van leerkrachten. Het is daarom van groot belang dat deze twee variabelen worden meegenomen als verklaringsmechanismen.

1.4 Sociologische relevantie

Een onderzoek kan als sociologisch relevant worden bestempeld indien het zich richt op een cohesie-, ongelijkheids- en/of rationaliseringsvraagstuk (De Graaf & Wiertz, 2019). Binnen het cohesievraagstuk wordt er onderzoek gedaan naar welke condities de sociale samenhang en integratie bevorderen of verzwakken. Binnen het ongelijkheidsvraagstuk wordt er onderzoek gedaan naar welke eigenschappen er ten grondslag liggen aan minder ongelijkheid in een samenleving. Binnen het rationaliseringsvraagstuk wordt er onderzoek gedaan naar maatschappelijke onderwerpen die de rationaliteit en efficiëntie van de moderne samenleving

beïnvloeden. Deze maatschappelijke onderwerpen worden gevormd door trends en ontwikkelingen die er in een samenleving spelen. Het onderzoek van deze masterscriptie kan worden geplaatst binnen de kaders van twee van de vraagstukken.

Allereerst past het onderzoek binnen de kaders van een cohesievraagstuk omdat pestgedrag de sociale cohesie en integratie in de klas onderdrukt. Een leerkracht die goed kan omgaan met pesten in de klas speelt een belangrijke rol in het opbouwen, onderhouden en behouden van een positief klassenklimaat. In een dergelijk klassenklimaat kunnen leerlingen worden gestimuleerd om respectvol, empathisch en tolerant met elkaar om te gaan waardoor de sociale cohesie en integratie wordt ondersteund (Bradshaw et al., 2007; Espelage & Swearer, 2011; Olweus, 2013).

Ten tweede past het onderzoek binnen de kaders van een ongelijkheidsvraagstuk omdat pesten gerelateerd kan worden aan machtsverschillen tussen leerlingen in de klas wat het welbevinden van een leerling kan doen afnemen. Machtsverschillen kunnen bijvoorbeeld ontstaan door verschillen in uiterlijk, etniciteit of populariteit. Om deze ongelijkheid in de klas te reduceren, is het van belang dat leerkrachten goed omgaan met pesten in de klas (Kärnä et al., 2011; Gini & Pozzoli, 2009).

2. Bestaande theoretische inzichten

Dit hoofdstuk biedt een theoretische verkenning voor het onderzoek wat centraal staat binnen deze scriptie. Deze verkenning maakt het mogelijk om het begrip ‘leerkracht-functioneren’ grondig te begrijpen en helpt mee bij het ontwikkelen van hypothesen die in de analyses kunnen worden getoetst. Er wordt bij het begrip ‘leerkracht-functioneren’ gekeken naar de attitude ten opzichte van pesten, bekwaamheid om pesten aan te pakken, en inspanning om dat daadwerkelijk te doen. Dit wordt onderzocht door onderscheid te maken naar het geslacht van de leerkracht, geslacht van de leerling, groep van de leerkracht, aantal leerkrachten van de klas, de KiVa-conditie van de leerkracht, het welbevinden van de leerling en het slachtofferschap van de leerling. Uiteindelijk worden de hypothesen gevisualiseerd in een conceptueel model dat wordt geanalyseerd in dit onderzoek.

2.1 Beoordeling leerkracht-functioneren

In deze studie wordt leerkracht-functioneren gemeten aan de hand van de percepties van leerlingen over de attitude tegenover pesten, de bekwaamheid om pesten te verminderen, en

de inspanning om pesten te verminderen van hun leerkracht, zoals gerapporteerd door de leerlingen zelf. De beoordelingen van leerlingen op deze drie variabelen worden verzameld via afzonderlijke vragen die gesteld worden tijdens de eerste (T1), derde (T3) en vijfde wave (T5), telkens aan het einde van het schooljaar.

2.1.1 Geslacht van de leerkracht

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot hun geslacht zijn gebaseerd op gender gerelateerde en sociaalpsychologische factoren.

In veel samenlevingen worden bepaalde gedragspatronen en -eigenschappen geassocieerd met traditionele mannelijkheid of vrouwelijkheid (Bem, 1974). Zo zijn mannen befaamd om hun kracht, dominantie en assertiviteit, terwijl vrouwen meer geacht worden geduldig, zorgzaam en empathisch te zijn (Eagly & Wood, 2012). Deze **genderstereotypen en -verwachtingen** spelen mogelijk een rol in de manier waarop leerlingen hun leerkrachten beoordelen als het gaat om het bestrijden en reduceren van pestgedrag.

Wat nauw samenhangt met de genderstereotypen en -verwachtingen zijn de **sociale normen en gedragingen** die gevormd worden door de genderrollen (Prentice & Carranza, 2002; Ridgeway, 2011). Indien de gedragingen van leerkrachten afwijken van de normen op school, kan dit leiden tot een beïnvloeding van de beoordeling over het leerkracht-functioneren. Om pestgedrag en conflicten in de klas op te lossen, vereist dat het nodige empathie, geduld en communicatie van de leerkracht wat in lijn ligt met de verwachtingen van leerlingen (Orpinas et al., 2003; Bradshaw, 2015; Merrell et al., 2008; Espelage & Swearer, 2004). Als vrouwelijke leerkrachten de vereiste eigenschappen daadwerkelijk in de praktijk brengen, wat in lijn ligt met hun genderverwachtingen, komt dit overeen met de algemeen geldende norm op school; oftewel als mannen deze eigenschappen vertonen, wijkt dit af van de traditionele mannelijkheid en mannelijke normen (Connell, 2005). De hypothese die hieruit volgt is:

H1: Vrouwelijke leerkrachten krijgen gemiddeld een hogere beoordeling op hun leerkracht-functioneren dan mannelijke leerkrachten.

2.1.2 Geslacht van de beoordelende leerling

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot het geslacht van de leerling die de beoordeling uitvoert, zijn

gebaseerd op sociale- en psychologische factoren.

De **geslacht gerelateerde percepties en verwachtingen** van leerlingen ten aanzien van hun leerkrachten, met betrekking tot geslachten, worden beïnvloedt door de significante rol van associaties (Sadker & Sadker, 1994). Bepaalde gedragspatronen en -eigenschappen worden geassocieerd met mannelijkheid of vrouwelijkheid. Door deze associaties kunnen leerlingen bevooroordeelde gedachten hebben over de verwachtingen van hun leerkrachten. Een goed voorbeeld hiervan is dat meisjes vrouwelijke leerkrachten mogelijk positiever beoordelen ten aanzien van het tegengaan van pesten omdat ze empathie hoger in het vaandel hebben staan dan jongens (Eagly & Wood, 2012).

Daarnaast spelen de **relationele vaardigheden** van zowel leerkrachten als leerlingen een essentiële functie bij het identificeren en reduceren van pestgedrag. Het reduceren van pestgedrag, wat gedaan kan worden door een interventie, vergt effectieve communicatie (Wentzel & Wigfield, 2009). Het is gemakkelijker om effectief te communiceren wanneer leerlingen beschikken over sterk ontwikkelde communicatieve vaardigheden. Eerder onderzoek wees uit dat meisjes over het algemeen een betere houding hebben in de klas dan jongens en dat hun sociale eigenschappen vaak verder gevorderd zijn (Geerdink, 2007; Driessen & Langen, 2010). Als gevolg hiervan zullen meisjes mogelijk meer geneigd zijn om de effectieve communicatie van leerkrachten te waarderen. De hypothese die hieruit volgt is:

H2: Meisjes geven leerkrachten gemiddeld een hogere beoordeling op hun leerkracht-functioneren dan jongens.

2.1.3 Groep van de leerkracht

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot de groep leerlingen waaraan ze lesgeven, zijn gebaseerd op ontwikkelings- en school gerelateerde factoren.

Allereerst kan er een verschil worden geduid in de **ontwikkelingsfase** van leerlingen die gepaard gaat met de **verwachtingen en percepties** ten aanzien van leerkrachten. Onderbouwleerlingen zitten in de vroege kindertijd waardoor ze zich op sociaal en emotioneel vlak aan het ontwikkelen zijn. Deze ontwikkeling vraagt om de nodige structuur, begeleiding en emotionele ondersteuning die de leerkracht kan bieden. Ook zijn de klassen van onderbouwleerlingen vaak kleiner waardoor leerkrachten meer tijd kunnen investeren in de interactie met leerlingen. Er is echter geen strikte regel betreft de klassengroottes; oftewel het

is óók mogelijk dat onderbouwklassen groter zijn dan bovenbouwklassen waardoor de uitspraak over meer tijd investeren in de interactie met leerlingen niet opgaat. Indien onderbouwklassen wel kleiner zijn, zorgt de versterkte interactie ervoor dat leerkrachten pestgedrag sneller signaleren en hiernaar kunnen handelen (Pellegrini & Long, 2002; Jimerson et al. 2009). Hier staat tegenover dat bovenbouwleerlingen in de late kindertijd zitten óf zich al in de adolescentie bevinden. Bovenbouwleerlingen ontwikkelen in deze fase meer onafhankelijkheid en cognitieve vermogens. Dit brengt met zich mee dat deze leerlingen zelf beter na kunnen denken en beschikken over meer zelfstandigheid. Dit kan resulteren in de vraag om minder afhankelijk te zijn van leerkrachten bij het signaleren en reduceren van pestgedrag. Tevens zijn de klassen van bovenbouwleerlingen over het algemeen groter dan in de onderbouw wat de interactie tussen de leerlingen en leerkracht moeilijker maakt (Espelage & Swearer, 2008; Bradshaw et al., 2014).

De ontwikkelingsfase correleert met de **cognitieve vermogen** van leerlingen (Piaget, 1972). Het cognitieve vermogen van onderbouwleerlingen is in tegenstelling tot bovenbouwleerlingen nog zeer beperkt waardoor ze moeite hebben om situaties te begrijpen en gedrag van anderen te beoordelen. Hierdoor hebben onderbouwleerlingen ook meer moeite om pestgedrag te begrijpen (Berk, 2015; Siegler et al., 2011). Dit zorgt er echter wel voor dat ze meer waarde hechten aan de aanpak van onderbouwleerkrachten. De hypothese die hieruit volgt is:

H3: Onderbouwleerkrachten krijgen gemiddeld een hogere beoordeling op hun leerkracht-functioneren dan bovenbouwleerkrachten.

2.1.4 Aantal leerkrachten voor de klas

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot het aantal leerkrachten die verantwoordelijk zijn voor één groep, zijn gebaseerd op sociale- en psychologische factoren.

Leerlingen verlangen **continuïteit en consistentie** tijdens het onderwijs op school (Rimm-Kaufman & Pianta, 2000). Indien leerlingen een langere periode les krijgen van één leerkracht, kan dit worden gezien als continuïteit. Tijdens deze periode is het waarschijnlijker dat de relatie tussen de leerlingen en de leerkracht stabiel wordt. Het voordeel wat dit met zich meebrengt, is het sneller signaleren en aanpakken van pestgedrag omdat de leerkracht bekend is met de sociale dynamiek onder leerlingen. Ook is het makkelijker voor deze leerkracht om de verwachtingen en regels consistent in stand te houden. Dit staat tegenover

een klas met meerdere leerkrachten (discontinuïteit) waarbij mogelijk verschillende verwachtingen en regels van verschillende leerkrachten worden gehanteerd, wat resulteert in gebrekkige consistentie of inconsistentie (Pianta & Stuhlman, 2004; Ingersoll & Strong, 2011).

Een andere factor wat continuïteit en consistentie met zich meebrengt is de **relatie en het vertrouwen** tussen leerlingen en leerkrachten (Roorda et al., 2011). De band tussen leerlingen en een leerkracht is sterker naarmate leerlingen langere tijd worden blootgesteld aan één leerkracht. De sterkere band leidt tot een groter gevoel van vertrouwen, waardoor leerlingen sneller pestgedrag bespreken en melden bij de leerkracht (Ladd & Burgess, 2001). Het is daarentegen moeilijker voor leerlingen met meerdere leerkrachten om een diepgaande band op te bouwen. Als leerlingen in een week worden onderworpen aan twee of meerdere leerkrachten, is er mogelijk sprake van een gebrek aan tijd om elkaar goed te leren kennen. Dit heeft een negatieve invloed op het vertrouwen van de leerlingen in de leerkrachten om pestgedrag te bespreken en te melden (Lynch & Cicchetti, 1997; Furrer & Skinner, 2003). De hypothese die hieruit volgt is:

H4: Leerkrachten die alleen voor de klas staan, krijgen gemiddeld een hogere beoordeling op hun leerkracht-functioneren dan leerkrachten die lesgeven aan een klas met meerdere leerkrachten.

2.1.5 KiVa-conditie van de leerkracht

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot hun betrokkenheid bij het KiVa-programma, zijn gebaseerd op school gerelateerde- en psychologische factoren.

Het **KiVa-programma** of de **KiVa-aanpak** is ontworpen om pesten op scholen te reduceren en het schoolklimaat positief te beïnvloeden en bevorderen. De KiVa-aanpak bestaat uit bewustwording, preventie en interventies om pestgedrag tegen te gaan (Veenstra & Huitsing, 2021). Leerkrachten krijgen vanuit het KiVa-programma handvaten om pestgedrag op een effectieve manier tegen te gaan.

KiVa-leerkrachten ontwikkelen een **positievere perceptie** van hun optreden ten aanzien van pestgedrag. Dit komt allereerst door de *training en ondersteuning* die KiVa-leerkrachten krijgen. Ten tweede wordt de positieve perceptie vergroot door de *feedback en evaluatie* aangezien ze hun aanpak kunnen optimaliseren. Ten slotte zorgt de *actieve*

betrokkenheid van ouders en leerlingen voor een bevordering van de positieve perceptie (Veenstra, 2015).

Naast het ontwikkelen van een positievere perceptie van hun optreden ten aanzien van pestgedrag, kunnen ze ook **daadwerkelijke effectiviteit** vergroten. Dit komt allereerst door de *effectieve interventies* die ze krijgen voorgeschoteld vanuit de KiVa-aanpak. Ten tweede wordt de daadwerkelijke effectiviteit vergroot door de *positieve klassenomgeving* waarin de leerkrachten te werk gaan. Ten slotte zorgt een *verhoogde betrokkenheid* voor een bevordering van de daadwerkelijke effectiviteit in de omgang tegen pestgedrag (Veenstra, 2015). De hypothesen die hieruit volgen zijn:

H5: Na de implementatie van het KiVa-programma krijgen KiVa-leerkrachten gemiddeld een hogere beoordeling op hun leerkracht-functioneren dan leerkrachten uit de controlegroep

2.1.6 Welbevinden van de leerling

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot het welbevinden van leerlingen, zijn gebaseerd op emotionele en psychologische factoren.

Het welbevinden van leerlingen is een belangrijk aspect van hun schoolervaring en heeft een directe invloed op verschillende aspecten van hun sociaal-emotionele- en academische ontwikkeling (Bradshaw et al., 2012). Positief welbevinden kan worden geassocieerd met hogere motivatie, betere leerprestaties en een gezonde sociaal-emotionele ontwikkeling (Suldo & Shaffer, 2008; Wang & Eccles, 2011).

Een hoog niveau van leerkracht-functioneren draagt bij aan een positief welbevinden van leerlingen, wat resulteert in een positief klassenklimaat waarin leerlingen eerder geneigd zijn de ondersteuning en begeleiding van leerkrachten te waarderen. Anderzijds beoordelen leerlingen met een laag welbevinden mogelijk het leerkracht-functioneren kritischer. Dit fenomeen kan worden toegeschreven aan het feit dat leerlingen met een laag welbevinden **vatbaarder** zijn voor **negatieve ervaringen** of **interacties** met leerkrachten, waardoor ze meer geneigd zijn om aandacht te schenken aan tekortkomingen van deze leerkrachten (Holfve-Sabel, 2014). Deze negatieve aandacht van leerlingen met een laag welbevinden resulteert in minder positieve beoordelingen.

Welbevinden oefent tevens invloed uit op de vorming van een emotionele band tussen leerlingen en leerkrachten (Holfve-Sabel, 2014). Indien leerkrachten een bijdrage leveren aan het welbevinden van hun leerlingen, kan er een **sterke vertrouwensband** ontstaan. Een sterke

vertrouwensband draagt bij aan goede communicatie en samenwerking; leerlingen zullen sneller om hulp vragen en vragen stellen wanneer ze zich op hun gemak voelen bij hun leerkrachten. De hypothese die hieruit volgt is:

H6: Leerkrachten ontvangen in klassen met een hoog welbevinden een hogere beoordeling op hun leerkracht-functioneren in vergelijking met klassen met een laag welbevinden.

2.1.7 Slachtofferschap van de leerling

De bestaande theoretische inzichten met betrekking tot de beoordeling van het functioneren van leerkrachten in relatie tot het slachtofferschap van leerlingen, zijn gebaseerd op emotionele en psychologische factoren.

Slachtofferschap van leerlingen verwijst naar het ervaren van pesten, dit kan door de leerling fysiek, mentaal of emotioneel letsel toe te brengen (Salmivalli & Peets, 2009). Het letsel kan leiden tot ernstige emotionele en psychologische gevolgen voor de getroffen leerlingen, waaronder angst, depressie en stress (Arseneault et al., 2010; Salmivalli & Peets, 2009; Ttofi et al., 2011). Daarnaast kan het slachtofferschap ook invloed hebben op hoe leerlingen hun leerkrachten beoordelen.

Leerlingen die slachtoffer zijn van pesten, en dus ook een **lager welbevinden** hebben, beoordelen mogelijk het leerkracht-functioneren kritischer. Zoals eerder is beschreven, zijn leerlingen met een lager welbevinden vatbaarder voor negatieve ervaringen of interacties met leerkrachten, waardoor ze meer geneigd zijn om aandacht te schenken aan tekortkomingen van deze leerkrachten (Holfve-Sabel, 2014).

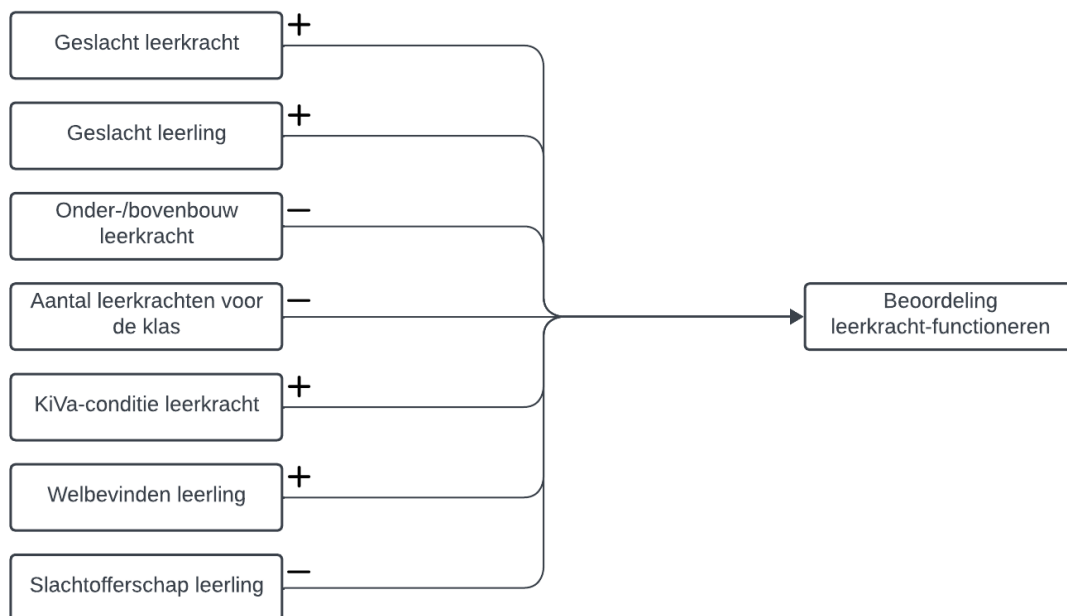
Daarnaast heeft een onderzoek van Cook en collega's (2010) aangetoond dat slachtofferschap van pesten correleert met **verminderde sociale en emotionele competenties** bij leerlingen. Leerlingen die gepest worden, kunnen moeite hebben met het opbouwen van relaties met hun leerkrachten. Dit resulteert in een verminderde bereidheid om begeleiding en ondersteuning vanuit de leerkracht te accepteren. Deze verminderde bereidheid heeft negatieve gevolgen voor de beoordeling van het leerkracht-functioneren.

Bovendien bestaat er een negatief verband tussen slachtofferschap en het zelfbeeld van leerlingen (Hawker & Boulton, 2000). Het **zelfbeeld** van gepeste leerlingen is **lager** waardoor ze terughoudender zijn in het vormen van emotionele banden met anderen, waaronder leerkrachten. Dit komt doordat ze twijfelen aan hun eigenwaarde en bekwaamheid. Ook dit draagt bij aan een lagere beoordeling van het leerkracht-functioneren. De hypothese die hieruit volgt is:

H7: Leerkrachten ontvangen in klassen waarin meer wordt gepest een lagere beoordeling op hun leerkracht-functioneren in vergelijking met klassen waarin minder wordt gepest.

2.2 Deze studie: conceptueel model

Figuur 1 illustreert het conceptuele model dat wordt gebruikt ter visuele representatie van de hypothesen die in dit onderzoek worden onderzocht. Een '+' geeft aan dat er een positief verband wordt verwacht; een '-' geeft aan dat er een negatief verband wordt verwacht. Ter verduidelijking staat de '+' bij geslacht voor het vrouwelijke geslacht en de '-' bij onder-/bovenbouwleerkracht voor bovenbouwleerkrachten.



Figuur 1 Conceptueel model

De volgende deelvragen zullen worden beantwoord door middel van bovenstaande hypothesen om een diepgaander inzicht te krijgen: 1. *'Zijn er duidelijke verschillen tussen leerkrachten: zijn er hoog- en laagpresteerders?'*, 2. *'Hoe is de ontwikkeling van de beoordeling van leerkrachten over de tijd, en is deze stabiel?'* en 3. *'Wat is dan vervolgens de samenhang met het klassenklimaat?'*

Deelvraag 1 onderzoekt of er duidelijke verschillen zijn tussen leerkrachten, met een focus op het identificeren van hoog- en laag presterende leerkrachten. Om de hoog- en laag presterende leerkrachten te identificeren, wordt de gemiddelde beoordeling op attitude, bekwaamheid en inspanning als maatstaf gebruikt. Hierbij wordt ook gekeken naar verschillen op basis van geslacht van de leerkracht, geslacht van de beoordelende leerling, de

groep waaraan de leerkracht lesgeeft, of de groep meerdere leerkrachten heeft en of de leerkracht op een KiVa-school werkt. De hypothesen 1 t/m 5 worden getoetst met behulp van t-toetsen, specifiek tijdens de eerste wave (T1). De t-toetsen richten zich op het beoordelen van mogelijke verschillen tussen de groepen op basis van de genoemde kenmerken en variabelen, met bijzondere aandacht voor de impact van het KiVa-programma op hypothese 5, die ook tijdens de derde (T3) en vijfde wave (T5) wordt getoetst om de vergelijking met de controlegroep te maken.

Deelvraag 2 onderzoekt hoe de beoordelingen van leerkrachten zich over de tijd ontwikkelen en of deze ontwikkeling stabiel is. De gemiddelde scores van attitude, bekwaamheid en inspanning worden samengevoegd tot één gemiddelde score voor het leerkracht-functioneren. Deze scores worden vervolgens onderverdeeld in vijf categorieën op basis van hun plaats in het eerste, tweede t/m vierde of vijfde kwintiel. Dit categoriseren van beoordelingen in kwintielen maakt het mogelijk om leerkrachten in verschillende groepen in te delen, zoals laag stabiel, dalers, middenmoters, verbeteraars en hoog stabiel. Door de kwintielen van elke meetronde aan elkaar te koppelen, kan worden onderzocht hoeveel leerkrachten van lage naar hoge beoordelingen gaan en vice versa. Hierdoor wordt ook hypothese 5 opnieuw getoetst.

Deelvraag 3 onderzoekt in hoeverre er samenhang is tussen het klassenklimaat, gemeten aan de hand van welbevinden en slachtofferschap, en de beoordeling van leerkrachten op het leerkracht-functioneren. Hierbij worden de eerder gecreëerde groepen uit de vorige deelvraag (deelvraag 2) meegenomen en wordt middels een ANOVA-analyse onderzocht of er significante verschillen zijn in groepsgemiddelden van de variabelen welbevinden en slachtofferschap. Daarnaast maakt een regressieanalyse het mogelijk de mate te beoordelen waarin deze variabelen de variabiliteit in leerkracht-functioneren verklaren en om hun impact op het leerkracht-functioneren vast te stellen. De bevindingen dragen bij aan het toetsen van hypothese 6 en 7.

3. Methode

3.1 Data

Data die in dit onderzoek gebruikt worden, zijn voortgekomen uit de evaluaties van het KiVa-programma in Nederland tussen mei 2012 en mei 2014. Het KiVa-antipestprogramma is een evidence-based interventie en streeft ernaar het pestgedrag te reduceren door de groepsdynamiek te optimaliseren, empathie en sociale vaardigheden te bevorderen en het

pestgedrag aan te pakken op individueel-, klassen- en schoolniveau. De interventie is oorspronkelijk afkomstig uit Finland waar het zijn effectiviteit reeds heeft aangetoond. Het KiVa-programma is in navolging van het succes in Finland ook in Nederland geïmplementeerd en geëvalueerd.

De evaluaties van het KiVa-programma zorgden ervoor dat tijdens vijf meetperiodes (waves) veranderingen op het gebied van pestgedrag, slachtofferschap, attitudes en welzijn op school geëvalueerd kon worden. Tevens kon aan de hand van deze evaluaties de effectiviteit van het programma worden vastgesteld en verbeterd. De evaluaties bestaan zowel uit leerling- als leerkrachtrapportages. Voor dit onderzoek is echter alleen gebruik gemaakt van leerlingrapportages over de leerkrachten.

3.2 Deelnemers

Dit onderzoek maakt gebruik van gegevens die verzameld zijn in drie opeenvolgende meetmomenten van gegevensverzameling in het kader van de evaluatie van het KiVa-antipestprogramma. De gegevens zijn verzameld in mei 2012 (*wave 1*: groepen 4-7, leeftijden 7-11), mei 2013 (*wave 3*: groepen 5-8, leeftijden 8-12) en mei 2014 (*wave 5*: groepen 5-8, leeftijd 8-12). Scholen, die kunnen worden beschouwd als deelnemers in het onderzoek, werden geworven middels een brief waarin het KiVa-project werd beschreven. Deze brief is in de herfst van 2011 verstuurd naar alle 6.966 basisscholen in Nederland. Van alle scholen gaven 132 scholen aan vrijwillig te willen deelnemen. Er voltooiden echter 100 scholen de eerste meetperiode (*wave 1*). Deze scholen werden daarna willekeurig toegewezen aan een interventie- of controlegroep. Tijdens deze toewijzing vielen er uiteindelijk nog twee scholen af. De interventiegroepen kregen voorafgaand van het nieuwe schooljaar in augustus 2012 een tweedaagse KiVa-training terwijl de controlegroepen juist gevraagd werden om hun standaard aanpak tegen pesten voort te zetten tot juni 2014. De leerlingen werden geworven middels een toestemmingsformulier die vanuit de scholen naar alle ouders en voogden van de leerlingen werden gestuurd. Voor alle waves lag de passieve toestemming om deel te nemen boven de 95%, waarbij passieve toestemming verwijst naar het verkrijgen van goedkeuring van de ouders of verzorgers door het ontbreken van bezwaar tegen de deelname.

De uiteindelijke steekproef omvatte tijdens alle drie de waves 98 scholen. Het aantal klassen en leerlingen tijdens de waves variëren doordat er na elke wave leerlingen om verscheidene redenen zowel de school/klas verlaten als -betreden. Tijdens *wave 1* (T1) zitten er 457 klassen en 2404 leerlingen (waarvan 50,5% jongens en 49,5% meisjes) in de

steekproef, tijdens *wave 3* (T3) zijn dit 394 klassen en 2554 leerlingen (waarvan 51,3% jongens en 48,7% meisjes) en tijdens *wave 5* (T5) zijn dit 391 klassen en 3482 leerlingen (waarvan 50,2% jongens en 49,8% meisjes). Het aantal leerkrachten voor de klas varieert van één leerkracht tot vijf leerkrachten. Er zijn slechts weinig klassen met drie of meer leerkrachten (W1 = 15, W3 = 18 en W5 = 20), wat ervoor heeft gezorgd dat het onderzoek zich beperkt tot leerkracht 1 en leerkracht 2.

3.3 Operationalisaties

3.3.1 Leerkracht-functioneren

Leerkracht-functioneren kan worden gedefinieerd aan de hand van de percepties van leerlingen over de attitude tegenover pesten, bekwaamheid om pesten te verminderen en inspanning om pesten te verminderen van hun leerkracht (Veenstra et al., 2014). De variabele leerkracht-functioneren werd gemeten door alle leerlingen een afzonderlijke vraag te stellen over zowel de attitude, de bekwaamheid als de inspanning. Allereerst werd de vraag gesteld: ‘Hoe denkt meester/juf {naam} over pesten?’. Leerlingen konden de vraag beantwoorden met: 1 = het is iets goeds, 5 = het is heel erg slecht. Deze antwoorden gaven de percepties van leerlingen weer over de attitude van leerkrachten tegenover pesten. Ten tweede werd de vraag gesteld: ‘Wat kan meester/juf {naam} doen om het pesten te stoppen?’. Leerlingen konden de vraag beantwoorden met: 1 = niets, 5 = heel veel. Deze antwoorden gaven de percepties van leerlingen weer over de bekwaamheid van leerkrachten om pesten te verminderen. Ten slotte werd de vraag gesteld: ‘Hoeveel heeft meester/juf {naam} sinds de kerstvakantie gedaan om pesten te stoppen?’. Leerlingen hadden dezelfde antwoordmogelijkheden als bij vraag twee. Deze antwoorden gaven de percepties van leerlingen weer over de inspanning van leerkrachten om pesten te verminderen. Om de antwoorden op de drie vragen gemakkelijker te kunnen duiden is het 5-punts rapportcijfer omgezet naar een 10-punts rapportcijfer waardoor bijvoorbeeld antwoordmogelijkheden 1 = niets, 5 = heel veel veranderen in 2 = niets, 10 = heel veel. Vervolgens wordt het oordeel van alle leerlingen over een leerkracht gemiddeld genomen voor elke klas op alle drie de meetmomenten afzonderlijk. Deze aanpak resulteert in een gemiddelde score per klas voor elk meetmoment en wordt gebruikt in de analyses op leerkrachtniveau. Gedurende de drie waves varieert de Cronbach’s alpha voor de gecombineerde gemiddelde score van de drie items tussen 0,67 en 0,73.

3.3.2 Welbevinden

Welbevinden kan worden gedefinieerd aan de hand van zowel de emotionele kant als de sociale kant. De emotionele kant van welbevinden is gericht op het leuk vinden en genieten van school; de sociale kant van welbevinden is gericht op gevoelens van veiligheid en acceptatie op school (Adi et al., 2007). De variabele welbevinden werd gemeten door de leerlingen stellingen te laten beantwoorden over de emotionele- en sociale kant van welbevinden. Om de emotionele kant van welbevinden te achterhalen werden 3 items gebruikt. Voorbeelden van de items zijn: “Op school vind ik het leuk” en “Ik ben blij als ik op school ben”. Leerlingen konden de vraag beantwoorden met: 1 = nooit, 4 = altijd. Om de sociale kant van welbevinden te achterhalen werden 4 items gebruikt. Voorbeelden van de items zijn: “Ik voel me veilig op school” en “Op school hoor ik erbij zoals ik ben”. Leerlingen hadden dezelfde antwoordmogelijkheden als bij de emotionele kant. Om de variabele welbevinden bruikbaar te maken, wordt de variabele samengesteld door de som te nemen van alle acht items, waarvan het gemiddelde wordt berekend. Deze samengestelde welbevindenscore wordt toegepast op leerkrachtniveau, waarbij het gemiddelde welbevinden voor elke leerkracht wordt berekend op basis van de individuele welbevindenscores van de leerlingen. Gedurende de drie waves varieert de Cronbach’s alpha voor de acht items tussen 0,77 en 0,87.

3.3.3 Slachtofferschap

Slachtofferschap kan worden gedefinieerd als de ervaring van een leerling van herhaaldelijk(e) geweld, agressie, intimidatie, uitsluiting of andere negatieve vormen van sociale interacties door één of meerdere leerlingen (Olweus, 1996). De variabele slachtofferschap werd in dit onderzoek gemeten door leerlingen een globale slachtoffersvraag te stellen. De vraag luidde: ‘Nu je weet wat pesten is, kun je dan ook zeggen hoe vaak je X (X staat voor het meetmoment: X1 = sinds de kerstvakantie 2011, X2 = sinds de zomervakantie 2012, X3 = sinds de kerstvakantie 2012 etc.) bent gepest?’. Leerlingen konden de vraag beantwoorden met: 1 = ik ben niet gepest 2 = ik ben één of twee keer gepest, 3 = ik ben twee of drie keer per maand gepest, 4 = ik ben ongeveer één keer per week gepest en 5 = ik ben meerdere keren per week gepest. Deze antwoorden gaven de frequentie weer waarin leerlingen gepest zijn. Om de variabele slachtofferschap bruikbaar te maken, zijn de antwoordmogelijkheden opnieuw gecodeerd en omgezet naar een dummy-variabele. Antwoordmogelijkheden 1 en 2 zijn gecodeerd als 0, wat aangeeft dat het kind niet wordt gepest, en antwoordmogelijkheden 3, 4 en 5 zijn gecodeerd als 1, wat betekent dat het kind wel wordt gepest. Het aantal gepeste kinderen per klas wordt berekend door de som van

de gepeste leerlingen af te zetten tegen alle leerlingen in de klas. De totale som van gepeste leerlingen in de klas wordt vervolgens omgezet naar een percentage van het totaal aantal leerlingen in die klas. Dit percentage fungeert als de variabele voor slachtofferschap op klasniveau (leerkrachtniveau).

3.4 Analyse-opzet

De opzet van de analyses wordt opgesplitst in drie losstaande benaderingen, aangezien ze zich toespitsen op het beantwoorden van de drie specifieke deelvragen. Hieronder worden de drie deelvragen met de bijbehorende analyse-opzet besproken.

3.4.1 Verschillen tussen leerkrachten

De eerste deelvraag luidt: *‘Zijn er duidelijke verschillen tussen leerkrachten: zijn er hoog- en laagpresteerders?’*. Om tot een antwoord te komen op deze deelvraag, wordt er aanvankelijk gekeken naar de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning om inzicht te verschaffen in de verdeling van deze variabelen. Beoordeling worden niet alleen gebaseerd op het gemiddelde, maar er wordt ook onderscheid gemaakt op basis van verschillende factoren. Dit omvat het geslacht van de leerkracht, waarbij afzonderlijke gemiddelde beoordelingen worden berekend voor mannelijke en vrouwelijke leerkrachten (1). Daarnaast wordt er onderscheid gemaakt op basis van het geslacht van de beoordelende leerling, waarbij aparte gemiddelde beoordelingen worden berekend voor de leerkracht beoordeeld door jongens versus meisjes (2). Verder wordt er gekeken naar de groep waaraan de leerkracht lesgeeft (3), of de groep meerdere leerkrachten heeft (4) en of de leerkracht werkzaam is op een KiVa-school (5). Vervolgens worden hypothesen 1 t/m 5 getoetst door middel van t-toetsen op het eerste meetmoment (T1), aangezien op dat moment het KiVa-programma nog niet is geïmplementeerd. Daarna wordt de relatie tussen het gemiddelde en de spreiding van deze variabelen onderzocht. Dit onderzoek verschaft inzicht in de mate van overeenstemming tussen leerlingen binnen een klas, waarbij de focus ligt op de vraag of leerlingen het eens zijn over het functioneren van hun leerkracht. Tot slot wordt *hypothese 5* opnieuw getoetst tijdens de derde (T3) en vijfde wave (T5) om de impact van het KiVa-programma ten opzichte van de controlegroep te beoordelen.

3.4.2 Ontwikkeling beoordeling leerkrachten

De tweede deelvraag luidt: ‘*Hoe is de ontwikkeling van de beoordeling van leerkrachten over de tijd, en is deze stabiel?*’. Om tot een antwoord te komen op deze deelvraag, wordt er allereerst gekeken naar de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning tijdens de meetmomenten van wave 1 (T1), wave 3 (T3) en wave 5 (T5). Deze scores worden visueel weergegeven in lijngrafieken voor zowel de KiVa-leerkrachten als de leerkracht uit de controlegroep. Vervolgens wordt de gecombineerde gemiddelde beoordeling van attitude, bekwaamheid en inspanning onderverdeeld in vijf categorieën op basis van hun plaats in het eerste, tweede t/m vierde of vijfde kwintiel. Beoordelingen die behoren tot het eerste kwintiel worden geclassificeerd als de laagste beoordelingen en ontvangen een score van 1, die in het tweede t/m het vierde kwintiel als middelste beoordelingen en ontvangen een score van 2 en die in het vijfde kwintiel als hoogste beoordelingen en krijgen een score van 3. Door de scores van de kwintielen (1,2 en 3) van de drie waves aan elkaar te koppelen, ontstaan verschillende groepen leerkrachten: laag stabiel, dalers, middenmoters, verbeteraars en hoog stabiel. Hierbij wordt een honderdtafschaal gebruikt om de scores van de kwintielen te combineren en verschillende groepen van leerkrachten te identificeren op basis van hun prestaties over de drie meetmomenten. Daarnaast kan er aan de hand van de kwintielen worden onderzocht hoeveel KiVa-leerkrachten en leerkrachten uit de controlegroep van een lage(re) naar een hoge(re) beoordeling gaan over tijd en vice versa, wat de mogelijkheid geeft om de impact van het KiVa-programma ten opzichte van de controlegroep (*hypothese 5*) te toetsen.

3.4.3 Samenhang klassenklimaat

De derde deelvraag luidt: ‘*Wat is dan vervolgens de samenhang met het klassenklimaat?*’. Om tot een antwoord te komen op deze deelvraag, wordt er eerst gekeken naar de gemiddelde scores van de leerkrachten uit de groepen 1 t/m 5 tijdens wave 5 (T5), gevormd tijdens de uitwerking van deelvraag 2. Deze leerkrachten hebben beoordelingen ontvangen op de variabelen attitude, bekwaamheid, inspanning, welbevinden en slachtofferschap. Het betreft hier leerling gerelateerde variabelen, wat impliceert dat de scores gebaseerd zijn op de gemiddelde beoordelingen van de leerlingen die de betreffende leerkrachten hebben beoordeeld; oftewel de leerkrachtcores weerspiegelen het gemiddelde welbevinden en slachtofferschap in de klas van de leerkracht. Na deze analyse wordt de relatie tussen deze variabelen bekeken om inzicht te krijgen in hun onderlinge verbanden. Hierna wordt onderzocht of er significante verschillen zijn in de gemiddelde scores van welbevinden en slachtofferschap tussen de groepen 1 t/m 5. Indien er significante verschillen worden

geïdentificeerd, wordt er vervolgens onderzoek gedaan naar welke specifieke groepen significant van elkaar verschillen (ANOVA-analyse). Tot slot wordt een regressieanalyse uitgevoerd. Hierbij worden de variabelen welbevinden, slachtofferschap en leerkracht-
conditie betrokken. Het doel van deze analyse is om de mate te beoordelen waarin deze variabelen bijdragen aan de variabiliteit in het functioneren van leerkrachten en om de invloed ervan op het leerkracht-functioneren vast te stellen. De bevindingen van de analyses geven de mogelijkheid om te toetsen of leerkrachten in klassen met een hoog welbevinden (*hypothese 6*) en waarin minder wordt gepest (*hypothese 7*) een hogere beoordeling op hun leerkracht-functioneren ontvangen.

4. Resultaten

4.1 Verschillen tussen leerkrachten

4.1.1 Wave 1

Tabel 1 geeft een overzicht van de gemiddelde scores op variabelen attitude, bekwaamheid en inspanning van leerkracht 1 (N=453, waarvan 19,2% mannen en 80,8% vrouwen) en leerkracht 2 (N=221, waarvan 14,7% mannen en 85,3% vrouwen), die door 2404 leerlingen (waarvan 50,5% jongens en 49,5% meisjes) tijdens wave 1 (T1) zijn gerapporteerd.

De gemiddelde scores zijn vergeleken door middel van t-toetsen. De t-toetsen zijn uitgevoerd op basis van verschillende factoren, waaronder het geslacht van de leerkracht, het geslacht van de beoordelende leerling, de groep waaraan de leerkracht lesgeeft, of de groep meerdere leerkrachten heeft, en of de leerkracht al dan niet op een KiVa-school werkt.

Tabel 1 Vergelijking van gemiddelde scores op leerkracht-functioneren op basis van verschillende factoren (T1)

	Leerkracht 1 (N)			Leerkracht 2 (N)		
Geslacht leerkracht:	Man (86)	Vrouw (358)	t-test	Man (32)	Vrouw (182)	t-test
<i>Attitude (2-10)</i>	8,52 (0,75)	8,62 (0,77)	$t(442) = -1.01; p = 0.32$	8,32 (0,66)	8,48 (0,68)	$t(212) = -1.25; p = 0.21$
<i>Bekwaamheid (2-10)</i>	7,20 (0,92)	7,41 (0,84)	$t(442) = -2.09; p = 0.04$	6,76 (1,02)	7,05 (0,92)	$t(212) = -1.63; p = 0.11$
<i>Inspanning (2-10)</i>	7,44 (0,69)	7,64 (0,66)	$t(442) = -2.43; p = 0.02$	7,13 (0,92)	7,48 (0,80)	$t(212) = -2.23; p = 0.03$
Geslacht leerling:	Jongen (450)	Meisje (450)	t-test	Jongen (219)	Meisje (219)	t-test
<i>Attitude (2-10)</i>	8,58 (0,92)	8,67 (0,87)	$t(449) = -1.90; p = 0.06$	8,46 (0,88)	8,49 (0,82)	$t(218) = -0.57; p = 0.57$
<i>Bekwaamheid (2-10)</i>	7,28 (1,02)	7,45 (1,04)	$t(449) = -3.16; p < 0.01$	6,93 (1,12)	7,11 (1,10)	$t(218) = -2.20; p = 0.03$
<i>Inspanning (2-10)</i>	7,53 (0,85)	7,64 (0,87)	$t(449) = -2.08; p = 0.04$	7,34 (1,01)	7,52 (0,98)	$t(218) = -2.52; p = 0.01$
Bouw:	Onder (103)	Boven (350)	t-test	Onder (43)	Boven (178)	t-test
<i>Attitude (2-10)</i>	8,21 (0,89)	8,73 (0,69)	$t(451) = -6.24; p < 0.01$	8,18 (0,65)	8,53 (0,70)	$t(219) = -3.00; p < 0.01$
<i>Bekwaamheid (2-10)</i>	7,53 (0,86)	7,33 (0,86)	$t(451) = 2.03; p = 0.04$	7,38 (0,77)	6,91 (0,95)	$t(219) = 3.00; p < 0.01$
<i>Inspanning (2-10)</i>	7,93 (0,61)	7,51 (0,66)	$t(451) = 5.74; p < 0.01$	7,83 (0,66)	7,33 (0,87)	$t(219) = 3.49; p < 0.01$

Aantal:	Enige leerkracht (232)			Meerdere leerkrachten (221)		t-test
<i>Attitude (2-10)</i>	8,65 (0,77)			8,56 (0,77)		$t(451) = 1.23; p = 0.22$
<i>Bekwaamheid (2-10)</i>	7,39 (0,88)			7,37 (0,84)		$t(451) = 0.15; p = 0.88$
<i>Inspanning (2-10)</i>	7,57 (0,62)			7,64 (0,72)		$t(451) = -1.08; p = 0.28$
Conditie:	Controle (150)	KiVa (303)	t-test	Controle (64)	KiVa (157)	t-test
<i>Attitude (2-10)</i>	8,57 (0,90)	8,63 (0,70)	$t(451) = -0.73; p = 0.47$	8,44 (0,72)	8,47 (0,70)	$t(219) = -0.33; p = 0.74$
<i>Bekwaamheid (2-10)</i>	7,33 (0,87)	7,41 (0,86)	$t(451) = -0.89; p = 0.38$	6,91 (0,91)	7,04 (0,94)	$t(219) = -0.99; p = 0.32$
<i>Inspanning (2-10)</i>	7,59 (0,63)	7,61 (0,69)	$t(451) = -0.30; p = 0.76$	7,41 (0,87)	7,43 (0,85)	$t(219) = -0.21; p = 0.84$

Onafhankelijke t-toetsen werden uitgevoerd om de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning van leerkrachten te vergelijken op basis van het geslacht van de leerkracht voor zowel leerkracht 1 als leerkracht 2. Op deze manier kon worden getoetst of vrouwelijke leerkrachten een hogere beoordeling krijgen op hun leerkracht-functioneren dan mannelijke leerkrachten (H1). Tabel 1 toont aan dat vrouwelijke leerkrachten bij leerkracht 1 gemiddeld hogere scores behalen op de variabelen bekwaamheid (7.20 voor mannen vs. 7.41 voor vrouwen, $t(442) = -2.09, p=.04$) en inspanning (7.44 voor mannen vs. 7.64 voor vrouwen, $t(442) = -2.43, p=.02$) in vergelijking met mannelijke leerkracht. Voor attitudes was er ook een verschil (8.52 voor mannen vs. 8.62 voor vrouwen), maar dit was niet statistisch significant, $t(442) = -1.01, p=.32$. Deze bevindingen zijn vergelijkbaar voor leerkracht 2. De resultaten liggen deels in lijn met de verwachting van *hypothese 1*.

Gepaarde t-toetsen werden uitgevoerd om de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning van zowel leerkracht 1 als leerkracht 2 te vergelijken op basis van het geslacht van de leerling. Op deze manier kon worden getoetst of meisjes een hogere beoordeling geven dan jongens (H2). Tabel 1 toont aan dat meisjes leerkracht 1 gemiddeld hogere scores geven op bekwaamheid (7.28 voor jongens vs. 7.45 voor meisjes, $t(449) = -3.16, p<.01$) en inspanning (7.53 voor jongens vs. 7.64 voor meisjes, $t(449) = -2.08, p=.04$) in vergelijking met jongens. Voor attitudes was er ook een verschil (8.58 voor jongens vs. 8.67 voor meisjes), maar dit was niet statistisch significant, $t(449) = -1.90, p=.06$. Deze resultaten zijn vergelijkbaar voor leerkracht 2. De bevindingen zijn overeenstemmend met *hypothese 2*. Desalniettemin is er geen bewijs gevonden dat er verschillen zijn in de beoordelingen van attitude tussen meisjes en jongens.

Onafhankelijke t-toetsen werden uitgevoerd om de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning van leerkrachten te vergelijken op basis van de groep waaraan de leerkracht lesgeeft (onderbouw of bovenbouw) voor zowel leerkracht 1 als leerkracht 2. Op deze manier kon worden getoetst of onderbouwleerkrachten gemiddeld een

hogere beoordeling krijgen op hun leerkracht-functioneren dan bovenbouwleerkrachten (H3). Tabel 1 toont aan dat zowel bij leerkracht 1 (8.21 voor onderbouw vs. 8.73 voor bovenbouw, $t(451) = -6.24, p < .01$) als bij leerkracht 2 (8.18 voor onderbouw vs. 8.53 voor bovenbouw, $t(219) = -3.00, p < .01$) bovenbouwleerkrachten hoger scoren op attitude in vergelijking met onderbouwleerkrachten. Daarentegen scoren onderbouwleerkrachten bij leerkracht 1 juist hoger op bekwaamheid (7.53 voor onderbouw vs. 7.33 voor bovenbouw, $t(451) = 2.03, p = .04$) en inspanning (7.93 voor onderbouw vs. 7.51 voor bovenbouw, $t(451) = 5.74, p < .01$) in vergelijking met bovenbouwleerkrachten. Deze resultaten zijn vergelijkbaar voor leerkracht 2 en komen overeen met *hypothese 3*; de bevindingen van de variabele attitude geven echter een tegengesteld verband waarbij bovenbouwleerkrachten beter worden beoordeeld dan onderbouwleerkrachten.

Onafhankelijke t-toetsen werden uitgevoerd om de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning van leerkrachten te vergelijken, gebaseerd op het aantal leerkrachten dat voor de klas staat. Op deze manier kon worden getoetst of leerkrachten die alleen voor de klas staan gemiddeld een hogere beoordeling krijgen op hun leerkracht-functioneren dan leerkrachten die lesgeven aan een klas met meerdere leerkrachten (H4). Tabel 1 toont aan dat er geen significant verschil is in de gemiddelde scores op de variabelen attitude (8.65 voor enige leerkracht vs. 8.56 voor meerdere leerkrachten, $t(451) = 1.23, p = .22$), bekwaamheid (7.39 voor enige leerkracht vs. 7.37 voor meerdere leerkrachten, $t(451) = 0.15, p = .88$) en inspanning (7.57 voor enige leerkracht vs. 7.64 voor meerdere leerkrachten, $t(451) = -1.08, p = .28$) tussen leerkrachten die alleen voor de klas staan en leerkrachten die lesgeven aan een klas met meerdere leerkrachten. Dit strookt niet met de verwachting van *hypothese 4*.

Onafhankelijke t-toetsen werden uitgevoerd om de gemiddelde scores op de variabelen attitude, bekwaamheid en inspanning van leerkrachten te vergelijken op basis van de conditie van de school (KiVa-conditie of niet) voor zowel leerkracht 1 als leerkracht 2. Op deze manier kon worden getoetst of KiVa-leerkrachten bij de voormeting gemiddeld een hogere beoordeling krijgen op hun leerkracht-functioneren dan leerkrachten uit de controlegroep (H5). Er wordt echter verwacht dat er tijdens wave 1, voor de interventie plaatsvindt, geen verschillen zijn tussen de twee groepen. Tabel 1 toont aan dat *Hypothese 5* niet opgaat tijdens wave 1 (T1). De testen zijn allemaal niet significant, wat betekent dat gemiddelde scores van de KiVa-leerkrachten en leerkrachten uit de controlegroep niet van elkaar verschillen. Dit komt tevens wel overeen met de verwachting bij wave 1.

Tabel 1 Bivariate statistieken variabelen attitude, bekwaamheid en inspanning van en tussen leerkracht 1 (N = 453) en 2 (N = 220) (T1)

Variabelen	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12
1. M attitude lk1	-	-,76**	,32**	-,26**	,23**	-,27**	,62**	-,58**	-,00	-,21**	-,02	-,25**
2. SD attitude lk1		-	-,09	,29**	-,05	-,35**	-,51**	,78**	,06	,28**	,05	,29**
3. M bekwaamheid lk1			-	-,43**	,50**	-,11*	,18**	-,05	,38**	-,21**	,18**	-,19**
4. SD bekwaamheid lk1				-	,98	,40**	-,17**	,35**	-,08	,61**	,09	,36**
5. M inspanning lk1					-	-,31**	,18**	-,07	,29**	-,09	,53**	-,37**
6. SD inspanning lk1						-	-,17**	,33**	-,06	,38**	-,17*	,64**
7. M attitude lk2							-	-,76**	,38**	-,25**	,42**	-,31**
8. SD attitude lk2								-	-,11	,26**	-,18**	,33*
9. M bekwaamheid lk2									-	-,38**	,71**	-,27**
10. SD bekwaamheid lk2										-	-,15*	,49**
11. M inspanning lk2											-	-,43**
12. SD inspanning lk2												-

* = $p < 0,05$, ** = $p < 0,01$

Tabel 2 geeft een overzicht van de correlaties tussen het gemiddelde en de bijbehorende standaarddeviatie van de variabelen attitude, bekwaamheid en inspanning van leerkrachten. De relaties tussen de gemiddelden en de standaarddeviaties van de variabelen zijn allemaal negatief, dit betekent dat een hoog gemiddelde samengaat met een lagere standaarddeviatie. De gemiddelde score van attitude bij leerkracht 1 heeft een significant sterke correlatie met de standaarddeviatie; oftewel er is meer overeenstemming tussen leerlingen bij goede beoordelingen op attitude van leerkracht 1 en minder overeenstemming bij slechte beoordelingen. Dit geldt tevens ook voor de beoordeling van leerkracht 2. Ook laat tabel 2 de relatie zien tussen de gemiddelden van leerkracht 1 en leerkracht 2. De gemiddelde attitudescores en inspanningsscores bij leerkracht 1 hebben een positief en significante, maar middelmatige correlatie met de scores van dezelfde variabelen bij leerkracht 2. Wat betekent dat hoge scores van attitude en inspanning bij leerkracht 1 over het algemeen samengaan met hoge scores van dezelfde variabelen bij leerkracht 2, hoewel deze correlatie niet erg sterk is.

4.1.3 Wave 3

Tijdens wave 3 (T3) zijn gegevens verzameld van 2554 leerlingen, waarvan 51,3% jongens en 48,7% meisjes. In tabel 3 zijn de resultaten van de onafhankelijke t-toetsen opgenomen, waarbij de variabelen attitude, bekwaamheid en inspanning worden vergeleken op basis van de conditie van de school. De gepresenteerde resultaten betreffen leerkracht 1 (N=391,

waarvan 20,1% mannen en 79,9% vrouwen) en leerkracht 2 (N=222, waarvan 19,5% mannen en 80,5% vrouwen).

Tabel 3 Onafhankelijke t-toets om variabelen te vergelijken op basis van de conditie van de school (T3)

Conditie:	Leerkracht 1 (N)			Leerkracht 2 (N)		
	Controle (121)	KiVa (270)	t-test	Controle (62)	KiVa (160)	t-test
Attitude (2-10)	8,78 (0,66)	9,04 (0,59)	t (389) = -4.01; p < 0.01	8,67 (0,56)	8,97 (0,58)	t (220) = -3.51; p < 0.01
Bekwaamheid (2-10)	6,60 (1,09)	7,38 (1,03)	t (389) = -6.86; p < 0.01	6,32 (1,00)	7,22 (1,03)	t (220) = -5.87; p < 0.01
Inspanning (2-10)	7,31 (0,78)	7,77 (0,75)	t (389) = -5.59; p < 0.01	7,17 (0,75)	7,75 (0,85)	t (220) = -4.69; p < 0.01

Tabel 3 toont aan dat de bevindingen overeenkomen met *Hypothese 5*. De testen zijn zowel bij leerkracht 1 als leerkracht 2 significant; oftewel KiVa-leerkrachten scoren gemiddeld hoger op alle variabelen dan leerkrachten uit de controlegroep.

4.1.5 Wave 5

Tijdens wave 5 (T5) zijn gegevens verzameld van 3482 leerlingen, waarvan 50,1% jongens en 49,8% meisjes. In tabel 4 zijn de resultaten van de onafhankelijke t-toetsen opgenomen, waarbij de variabelen attitude, bekwaamheid en inspanning worden vergeleken op basis van de conditie van de school. De gepresenteerde resultaten betreffen leerkracht 1 (N=381, waarvan 22,7% mannen en 77,3% vrouwen) en leerkracht 2 (N=231, waarvan 18,1% mannen en 81,9% vrouwen).

Tabel 4 Onafhankelijke t-toets om variabelen te vergelijken op basis van de conditie van de school (T5)

Conditie:	Leerkracht 1 (N)			Leerkracht 2 (N)		
	Controle (119)	KiVa (262)	t-test	Controle (68)	KiVa (163)	t-test
Attitude (2-10)	8,84 (0,69)	9,05 (0,62)	t (379) = -2.95; p < 0.01	8,70 (0,80)	8,96 (0,60)	t (229) = -2.77; p < 0.01
Bekwaamheid (2-10)	6,45 (1,00)	7,05 (1,08)	t (379) = -5.13; p < 0.01	6,19 (1,21)	6,73 (1,13)	t (229) = -3.29; p < 0.01
Inspanning (2-10)	7,17 (0,76)	7,59 (0,81)	t (379) = -4.81; p < 0.01	7,10 (1,13)	7,57 (0,88)	t (229) = -3.40; p < 0.01

Tabel 4 toont aan dat de bevindingen overeenkomen met *Hypothese 5*. De testen vertonen significantie op een vergelijkbare wijze als tijdens de derde wave, zowel bij leerkracht 1 als leerkracht 2, wat betekent dat KiVa-leerkrachten gemiddeld hoger scoren op alle variabelen dan leerkrachten uit de controlegroep.

4.2 Ontwikkeling beoordeling leerkrachten

4.2.1 Univariante statistieken

Voor het onderzoek naar de ontwikkeling van de beoordeling van leerkrachten zijn de leerkrachten geselecteerd die gedurende alle waves beoordelingen hebben gekregen. Hierdoor is het oorspronkelijke aantal leerkrachten van 976 gereduceerd tot 259 leerkrachten. Tabel 5 geeft een overzicht van de univariate statistieken van de variabelen attitude, bekwaamheid en inspanning van KiVa-leerkrachten (N=178) en leerkrachten uit de controlegroep (N=81) die door leerlingen zijn gerapporteerd.

De KiVa-leerkrachten scoren, los van de variabele attitude in wave 1, op alles hoger dan leerkrachten uit de controlegroep. Wat daarnaast opvalt is dat leerkrachten gemiddeld een hoge beoordeling halen op de variabele attitude, deze varieert van 8,70 tot 9,10. Dit wijst uit dat leerlingen aangeven dat leerkrachten pesten ‘iets slechts’ tot ‘iets heel slechts’ vinden. Tot slot ligt de standaarddeviatie bij de variabele bekwaamheid in wave 3 en 5 rond 1, wat betekent dat de gegeven beoordelingen flink uiteenlopen.

Tabel 5 Univariate statistieken variabelen attitude, bekwaamheid en inspanning KiVa- (N=178) en controlegroep leerkrachten (N=81)

Conditie:	Wave: 1		3		5	
	Controle	KiVa	Controle	KiVa	Controle	KiVa
<i>Attitude (2-10)</i>	8,76 (0,67)	8,70 (0,69)	8,90 (0,56)	9,10 (0,53)	8,90 (0,65)	9,07 (0,61)
<i>Bekwaamheid (2-10)</i>	7,17 (0,86)	7,24 (0,81)	6,79 (0,96)	7,51 (0,97)	6,64 (1,04)	7,05 (1,08)
<i>Inspanning (2-10)</i>	7,52 (0,68)	7,53 (0,67)	7,48 (0,72)	7,84 (0,67)	7,31 (0,81)	7,66 (0,83)

4.2.2 Beoordelingen categoriseren

De gecombineerde gemiddelde beoordeling van leerkrachten op attitude, bekwaamheid en inspanning, oftewel de gemiddelde score op leerkracht-functioneren, zijn onderverdeeld in vijf categorieën afhankelijk van hun plaats in het eerste, tweede tot vierde of vijfde kwintiel per wave. Beoordelingen die behoren tot het eerste kwintiel worden geclassificeerd als de laagste beoordelingen en ontvangen een score van 1, die in het tweede t/m het vierde kwintiel als middelste beoordelingen en ontvangen een score van 2 en die in het vijfde kwintiel als hoogste beoordelingen en krijgen een score van 3. De leerkrachten in de groep ‘stabiel laag’ (N=64) behalen scores die zich meermaals bevinden in het eerste kwintiel, wat duidt op constante lage beoordelingen. De leerkrachten in de groep ‘dalers’ (N=30) behalen in de eerste wave hogere scores ten opzichte van de andere waves, wat wijst op een negatieve trend. De leerkrachten in de groep ‘middenmoters’ (N=53) behalen scores die balanceren in én rondom het tweede kwintiel tot vierde kwintiel, wat laat zien dat de gemiddelde scores gehandhaafd worden. De leerkrachten in de groep ‘verbeteraars’ (N=42) behalen in de laatste waves hogere scores in vergelijking met de voorgaande waves, wat een positieve trend

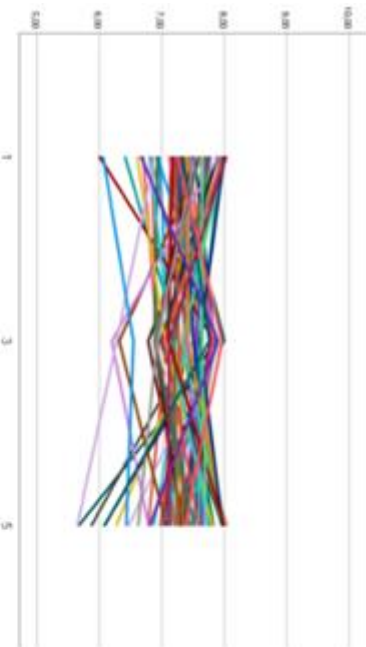
aangeeft. De leerkrachten in de groep ‘stabiel hoog’ (N=53) behalen scores die zich meermaals in het vijfde kwintiel, wat duidt op constante hoge beoordelingen. Tabel 6 presenteert een overzicht van de groepsverdeling, waarbij de corresponderende kwintielscores (1,2 en 3) per meetmoment aan elkaar zijn gekoppeld en uitgedrukt wordt in een honderdtal. Daarnaast worden het aantal leerkrachten in zowel de KiVa-groep als de controlegroep weergegeven. Ter verduidelijking van de honderdtallen impliceert bijvoorbeeld de score 111 dat de beoordeling van de leerkracht tijdens meetmoment 1, 3 en 5 in het eerste kwintiel valt. Dit honderdtal wordt toegewezen aan de groep 'stabiel laag' omdat de scores herhaaldelijk in het eerste kwintiel vallen, wat wijst op consistente lage beoordelingen. Een ander voorbeeld is 312, waarbij de beoordeling van de leerkracht tijdens meetmoment 1 in het vijfde kwintiel ligt, tijdens meetmoment 3 in het eerste kwintiel valt en tijdens meetmoment 5 tussen het tweede en vierde kwintiel in zit. Dit honderdtal wordt toegewezen aan de groep ‘dalers’ vanwege de hogere score in de eerste wave hoger in vergelijking met de waves die daarop volgen. De leerkrachten met kwintielscores 131, 231 en 313 (N=17) zijn uitgesloten van de steekproef, aangezien deze scores niet adequaat konden worden toegewezen tot één van de groepen. Als gevolg hiervan is de steekproef gereduceerd tot 242 leerkrachten.

Tabel 6 Overzicht van de groepsindeling (N=242)

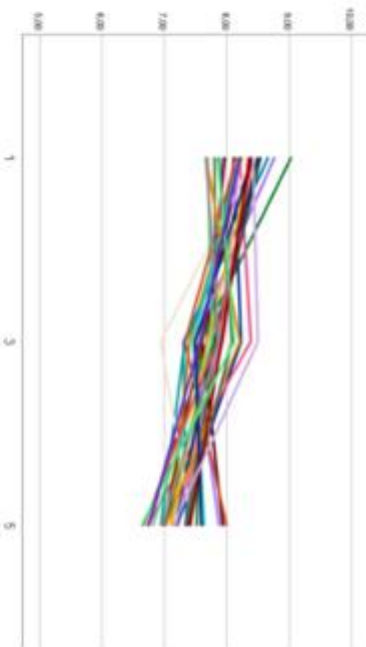
Groepen en aantal leerkrachten (Controle/KiVa)										
Groepen:	Stabiel laag		Dalers		Middenmoters		Verbeteraars		Stabiel hoog	
	111	16 (8/8)	221	8 (3/5)	122	11 (2/9)	113	2 (0/2)	323	6 (3/3)
	112	11 (5/6)	311	6 (4/2)	132	9 (2/7)	123	9 (2/7)	332	5 (1/4)
	121	14 (7/7)	312	6 (4/2)	222	14 (3/11)	133	11 (0/11)	333	42 (5/38)
	211	9 (5/4)	321	7 (3/4)	232	7 (5/2)	213	1 (0/1)		
	212	14 (10/4)	331	3 (0/3)	322	12 (4/8)	223	5 (2/3)		
							233	14 (3/11)		
Totaal:	5	64 (35/29)	5	30 (13/17)	5	53 (16/37)	6	42 (7/35)	3	53 (9/44)

Cutoff-points van de beoordelingen: (T1) 1 t/m een 7,58, 2 t/m een 8,05 en 3 vanaf een 8,06 – (T3) 1 t/m een 7,64, 2 t/m een 8,22 en 3 vanaf een 8,23 – (T5) 1 t/m een 7,49, 2 t/m een 8,05 en 3 vanaf een 8,06

Om de groeps categorieën te verduidelijken, zijn visuele representaties gemaakt op basis van de gemiddelde beoordelingen om het verloop van de scores op het leerkracht-functioneren te illustreren. Allereerst is er een algehele visualisatie gemaakt voor zowel KiVa-leerkrachten als de leerkrachten uit de controlegroep (zie bijlage II). Daarnaast zijn hieronder de vijf groepen visueel weergegeven, waarbij het gemiddelde leerkracht-functioneren op de Y-as is uitgezet en de waves op de X-as staan (zie figuur 2 t/m 6).



Figuur 2 Visualisatie groep 'Stabiel laag'



Figuur 3 Visualisatie groep 'Dalers'



Figuur 4 Visualisatie groep 'Middenmoters'



Figuur 5 Visualisatie groep 'Stabiel hoog'



Figuur 6 Visualisatie groep 'Stabiel laag'

4.2.3 Stijgende en dalende beoordelingen

Tabel 7 Overzicht aantal stijgende en dalende beoordelingen van leerkrachten uit controlegroep (N=78) en KiVa-leerkrachten (N=164)

Score	Aantal (% van N)	
	Controle	KiVa
Stabiël laag	35 (44%)	29 (18%)
Dalers	13 (16%)	17 (11%)
Middenmoters	16 (20%)	37 (23%)
Verbeteraars	7 (9%)	35 (21%)
Stabiël hoog	9 (11%)	44 (27%)
N	78	164

Tabel 7 en een chi-kwadraat toets, $X^2(4, N = 242) = 29.38, p < 0.01$, tonen aan dat de bevindingen overeenkomen met *Hypothese 5*. Het aantal KiVa-leerkrachten is percentueel sterker vertegenwoordigd in de groep verbeteraars (21%) dan leerkrachten uit de controlegroep (9%), wat betekent dat KiVa-leerkrachten over de tijd gemiddeld een hogere beoordeling op hun leerkracht-functioneren krijgen dan leerkrachten uit de controlegroep. De chi-kwadraat toets toont tevens aan dat er een significant verschil is in de score tussen de controle- en KiVa-leerkrachten.

4.3 Het klassenklimaat

4.3.1 Univariante & Bivariate statistieken

Voor het onderzoek naar het klassenklimaat zijn dezelfde leerkrachten geselecteerd die bij alle metingen betrokken waren, waardoor de steekproef 242 leerkrachten bevat. Tabel 8 geeft een overzicht van de univariate statistieken van de variabelen attitude, bekwaamheid, inspanning, welbevinden en slachtofferschap van KiVa-leerkrachten (N=162) en leerkrachten uit de controlegroep (N=80) die door leerlingen zijn gerapporteerd.

Tabel 8 Univariante statistieken variabelen attitude, bekwaamheid, inspanning, welbevinden en slachtofferschap KiVa- (N=162) en controlegroep leerkrachten (N=80) (T5)

Variabelen (schaal)	Gemiddelde (standaarddeviaties)	
	Controle	KiVa
Attitude (2-10)	8,87 (0,65)	9,11 (0,58)
Bekwaamheid (2-10)	6,62 (1,05)	7,12 (1,05)
Inspanning (2-10)	7,28 (0,80)	7,71 (0,84)
Welbevinden (0-4)	3,14 (0,18)	3,15 (0,21)
Slachtofferschap (0-100)	17,22%	13,05%
Slachtofferschap max.	53,8%	71,40%

KiVa-leerkrachten scoren hoger op de variabelen die het leerkracht-functioneren representeren in vergelijking met leerkrachten uit de controlegroep. Verder valt het op dat het verschil in de scores op welbevinden tussen de twee groepen nihil is; de gemiddelde score geeft aan dat leerlingen overwegend ‘meestal’ (antwoordmogelijkheid 3) antwoordden op vragen met betrekking tot welbevinden in de klas van de leerkracht. Gebruikmakend van een 4-puntsschaal impliceert dit gemiddelde dat het emotionele- en sociale welbevinden redelijk hoog ligt. Ten slotte is het gemiddelde percentage slachtofferschap iets lager in de klassen van KiVa-leerkrachten, met uitzondering van een klas met het hoogste percentage.

Tabel 9 geeft een overzicht van de correlaties tussen het gemiddelde van de variabelen attitude, bekwaamheid, inspanning, welbevinden en slachtofferschap van leerkrachten tijdens wave 5. De gemiddelde score van attitude heeft een significant positieve en matige correlatie met de gemiddelde score van welbevinden, wat betekent dat een hogere attitudescore samengaat met een verhoogde welbevindenscore. Daarnaast heeft de gemiddelde score van attitude een significant negatieve en matige correlatie met het percentage slachtofferschap. Met andere woorden, een hogere attitudescore van de leerkracht gaat mogelijk gepaard met een lager percentage slachtofferschap. Tot slot geldt deze correlatie ook voor de gemiddelde score van welbevinden en het percentage slachtofferschap. Kortom, hogere welbevindenscores lijkt samen te gaan met een lager percentage slachtofferschap.

Tabel 9 Bivariate statistieken variabelen attitude, bekwaamheid, inspanning, welbevinden en slachtofferschap (N=242) (T5)

Variabelen	1.	2.	3.	4.	5.
1. M attitude	-	,36**	,34**	,43**	-,40**
2. M bekwaamheid		-	,71**	,18**	,08
3. M inspanning			-	,31**	,06
4. M welbevinden				-	-,41**
5. Pct. slachtofferschap					-

* = $p < 0,05$, ** = $p < 0,01$

4.3.2 Welbevinden in de klas

Een ANOVA-analyse (zie bijlage III) vertoont een significant verschil in het welbevinden in de klas is tussen de beoordeelde leerkrachten tijdens wave 5, $F(4,237) = 4.33$; $p = .002$. Post-hoc analyses laten zien dat de klassen van de groep verbeteraars een significant hogere welbevindenscores behalen dan de groep stabiel laag ($p = .002$) en dalers ($p = .019$). Tussen de rest van de groepen zijn geen significante verschillen in het welbevinden gevonden. Deze resultaten ondersteunen *hypothese 6*, waarbij wordt gesuggereerd dat er een samenhang is

tussen de beoordelingen van leerkrachten op hun functioneren en het welbevinden in klassen; de bevindingen duiden op een relatie tussen deze factoren, waarbij gunstigere welbevindenscores samenhangen met hogere welbevindenscores.

Tabel 10 Overzicht groepsgemiddelden welbevinden (T5)

Groepen	Gemiddelde (1-4)
1. Stabiël laag	3,09
2. Dalers	3,10
3. Middenmoters	3,15
4. Verbeteraars	3,24
5. Stabiël hoog	3,17

Tabel 11 Modelschatting leerkracht-functioneren – hypothese 6 (T5)

	b(SE)	p
(Constant)	3,99(0,64)	<0,01
Welbevinden (1-4)	1,41(0,20)	<0,01
Conditie (0-1) ^a	1,03(0,09)	<0,01
R ²	0,18	
Adjusted R ²	0,17	
F-change	26,05	<0,01

^a = variabele is een dummy

Een regressie-analyse (zie bijlage III en tabel 11) met het gemiddelde leerkracht-functioneren als afhankelijke variabele en welbevinden en leerkracht-conditie als verklarende variabele is significant, $F(2.239) = 26.05$, $p < .001$. Ongeveer 17,9% van de variantie in het leerkracht-functioneren kan worden verklaard door zowel het welbevinden als de leerkracht-conditie. Het welbevinden vertoont een positieve samenhang met het leerkracht-functioneren ($b = 1,15$; $p = <0,01$), wat ook geldt voor de leerkracht-conditie ($b = 0,37$; $p = <0,01$). Dit suggereert dat naarmate het welbevinden in een klas toeneemt, de waarschijnlijkheid van een hogere beoordeling op het leerkracht-functioneren eveneens stijgt. Bovendien wordt deze kans verder vergroot wanneer de leerkracht tot de KiVa-conditie behoort, wat de consistentie ondersteunt met *hypothese 6*.

4.3.3 Slachtofferschap in de klas

Een ANOVA-analyse (zie bijlage III) vertoont geen significant verschil in het percentage slachtofferschap in de klas tussen de beoordeelde leerkrachten, $F(4.237) = 0.40$; $p = .811$. Deze bevindingen suggereren dat de beoordelingen van het leerkracht-functioneren niet

significant samengaan met het niveau van slachtofferschap. Het ontbreken van een significant verschil duidt erop dat de variaties in de beoordelingen van leerkracht-functioneren niet direct gerelateerd is aan de mate van slachtofferschap, wat afwijkt van *hypothese 7*.

Tabel 12 Overzicht groepsgemiddelden welbevinden (T5)

Groepen	Gemiddelde (0-100)
1. Stabiel laag	15,08%
2. Dalers	15,59%
3. Middenmoters	12,73%
4. Verbeteraars	14,87%
5. Stabiel hoog	14,17%

Tabel 13 Modelschatting leerkracht-functioneren – hypothese 7 (T5)

	b(SE)	p
(Constant)	7,61(0,10)	<0,01
Slachtofferschap (0-100)	-,001(,004)	0,83
Conditie (0-1) ^a	0,38(0,09)	<0,01
R ²	0,07	
Adjusted R ²	0,06	
F-change	8,69	<0,01

^a = variabele is een dummy

Een regressie-analyse (zie bijlage III en tabel 13) met het gemiddelde leerkracht-functioneren als afhankelijke variabele en slachtofferschap en leerkracht-conditie als verklarende variabele is significant, $F(2,239) = 8,69$, $p < .001$. Ongeveer 6,8% van de variantie in het leerkracht-functioneren kan worden verklaard door zowel het slachtofferschap als de leerkracht-conditie. Het slachtofferschap vertoont een minimale negatieve samenhang met het leerkracht-functioneren ($b = -,001$; $p = 0,83$), wat echter niet-significant is. Deze bevindingen suggereren dat, hoewel er een minimale daling is in de waarschijnlijkheid van een hogere beoordeling op het leerkracht-functioneren bij toenemend slachtofferschap in een de klas, deze stijging niet significant is. Dit wijst uit dat er onvoldoende statistisch bewijs is voor *hypothese 7* om te concluderen dat slachtofferschap in de klas significant van invloed is op de beoordelingen van het leerkracht-functioneren.

5. Conclusie

De conclusie brengt samen wat de bevindingen zijn over de betrouwbaarheid en validiteit van leerlingrapportages over het functioneren van KiVa-leerkrachten. Het onderzoek heeft zich

gericht op de effectiviteit en bruikbaarheid van deze rapportages, met specifieke aandacht voor de invloed van het klassenklimaat op de beoordelingen van leerlingen over het leerkracht-functioneren. Door deze bevindingen hoop ik een bijdrage te leveren aan de optimalisatie van het KiVa-rapport. De resultaten van dit onderzoek geven antwoord op de centrale probleemstelling: "*Wat is de betrouwbaarheid en validiteit van leerlingrapportages over het functioneren van KiVa-leerkrachten?*".

5.1 Terugkoppeling hypothesen

In dit onderzoek zijn verschillende hypothesen geformuleerd om een dieper begrip te verschaffen in de beoordelingen van leerkrachten. Hypothese 1 t/m 5 zijn onderzocht door middel van t-toetsen, waarbij verschillende kenmerken werden onderscheiden. *Hypothese 1* en *hypothese 2* richtten zich op geslachtsverschillen bij leerkrachten en leerlingen. Uit de resultaten blijkt dat vrouwelijke leerkrachten gemiddeld hogere beoordelingen krijgen dan mannelijke leerkrachten (H1), en dat meisjes leerkrachten gemiddeld hogere scores geven dan jongens (H2). Deze verschillen zijn echter niet significant voor de variabele attitude. *Hypothese 3* suggereerde dat onderbouwleerkrachten gemiddeld hogere beoordelingen ontvangen dan bovenbouwleerkrachten. Deze hypothese wordt ondersteund door de resultaten, behalve voor de attitudevariabele, waar het effect juist tegenovergesteld bleek. *Hypothese 4* stelde dat leerkrachten die alleen voor de klas staan, gemiddeld hogere beoordelingen krijgen ten opzichte van leerkrachten die lesgeven aan een klas met meerdere leerkrachten. De resultaten stroken niet met de hypothese, aangezien de leerkrachten die alleen voor de klas staan lagere beoordelingen ontvingen betreft de inspanning die ze leveren. *Hypothese 5* opperde dat KiVa-leerkrachten gemiddeld hogere beoordelingen ontvangen ten opzichte van leerkrachten uit de controlegroep. Hoewel er in de eerste wave (T1) geen significant verschil werd gevonden, wat overeenkwam met de verwachting, werd er in wave 3 (T3) en 5 (T5) wél een significant verschil gevonden waaruit bleek dat KiVa-leerkrachten hogere beoordelingen ontvangen van leerlingen.

De bevindingen uit dit onderzoek bieden een antwoord op de deelvraag: "*Zijn er duidelijke verschillen tussen leerkrachten: zijn er hoog- en laagpresteerders?*". Uit de resultaten blijkt dat er variaties zijn in de beoordelingen van de leerkrachten op basis van verschillende kenmerken. Gender-gerelateerde kenmerken tonen aan dat vrouwelijke leerkrachten gemiddeld hogere beoordelingen ontvangen dan mannelijke leerkrachten, en dat meisjes aan leerkrachten hogere beoordelingen toekennen dan jongens. Daarnaast blijkt dat er

differentiatie is tussen onderbouw- en bovenbouwleerkrachten, waarbij onderbouwleerkrachten een hogere beoordeling ontvangen op bekwaamheid en inspanning dan bovenbouwleerkrachten. Dit geldt tevens niet voor de variabele attitude. Tot slot laten de resultaten zien dat de deelname aan het KiVa-programma verbeterde beoordelingen oplevert voor KiVa-leerkrachten vergeleken met leerkrachten uit de controlegroep.

Hypothese 5 is niet alleen beoordeeld door middel van een t-toets, maar ook onderzocht door de KiVa-leerkrachten en leerkrachten uit de controlegroep in te delen in groepen op basis van hun gecombineerde gemiddelde score op leerkracht-functioneren. De resultaten uit dit onderzoek liggen in lijn met de hypothese (H5).

De bevindingen uit dit onderzoek bieden een antwoord op de deelvraag: “*Hoe is de ontwikkeling van de beoordeling van leerkracht over de tijd, en is deze stabiel?*”. De resultaten suggereren dat KiVa-leerkrachten sterker vertegenwoordigd zijn in de groep ‘stijgers’ dan leerkrachten uit de controlegroep, terwijl het omgekeerde geldt voor de groep ‘dalers’. Dit wijst op een positievere ontwikkeling in de beoordeling voor KiVa-leerkrachten ten opzichte van leerkrachten uit de controlegroep.

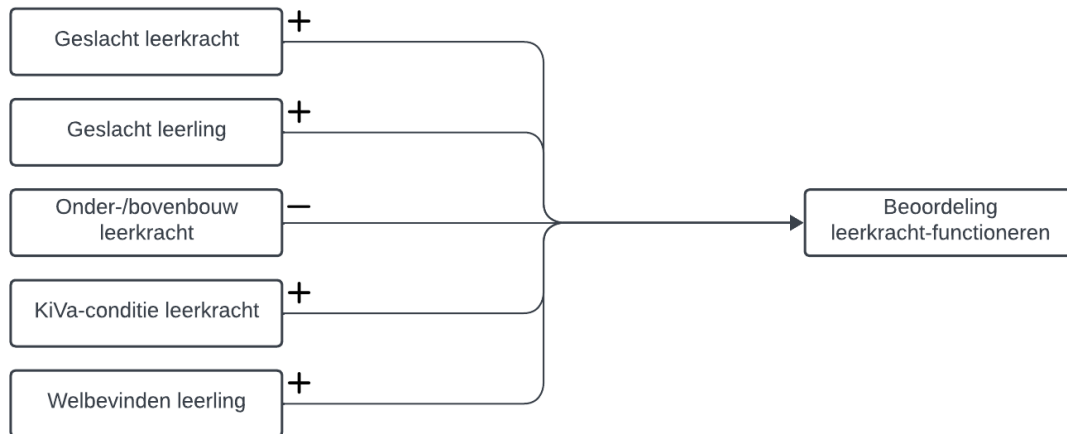
Hypotheses 6 en 7 zijn onderzocht door middel van een ANOVA- en regressieanalyse.

Hypothese 6 wordt ondersteund doordat leerkrachten in klassen met een hoog welbevinden gemiddeld een hogere beoordeling ontvangen op hun leerkracht-functioneren ten opzichte van leerkrachten in een klassen met een laag welbevinden. Deze relatie wordt tevens versterkt wanneer het een KiVa-leerkracht betreft (H6). Er is echter voor *hypothese 7* geen significant bewijs gevonden, wat betekent dat het niveau van slachtofferschap in de klas niet direct correleert met beoordelingen van het leerkracht-functioneren (H7).

De bevindingen uit dit onderzoek bieden een antwoord op de deelvraag: “*Wat is dan vervolgens de samenhang met het klassenklimaat?*”. Uit de resultaten blijkt dat het welbevinden in de klas een positieve samenhang heeft op de beoordelingen van het leerkracht-functioneren. Er is echter geen relatie gevonden tussen de mate van slachtofferschap in de klas en de beoordelingen van het leerkracht-functioneren.

Figuur 7 illustreert een vernieuwd conceptueel model, geïntegreerd met de resultaten van dit onderzoek. Vanwege het ontbreken van significante resultaten zijn de effecten van het aantal leerkrachten en de mate van slachtofferschap niet opgenomen. Een ‘+’ geeft aan dat er een

positief verband wordt verwacht; een ‘-’ geeft aan dat er een negatief verband wordt verwacht. Het ‘+’ symbool bij geslacht vertegenwoordigt vrouwelijke geslacht, en het ‘-’ symbool bij onder-/bovenbouwleerkracht staat voor bovenbouwleerkrachten. Bovendien zijn er geen significante resultaten gevonden voor de attitudebeoordeling bij deze twee kenmerken, zoals eerder vermeld.



Figuur 7 Vernieuwd conceptueel model

5.2 Discussie

Attitude is een veelzijdig en complex concept met een subjectieve aard. Deze kenmerken van attitude kunnen mogelijk verklaren waarom er geen verschillen zijn gevonden in de beoordelingen ervan met betrekking tot het geslacht van zowel de leerlingen als de leerkrachten. Iemands attitude wordt beïnvloed door een grote verscheidenheid aan individuele factoren. Voorbeelden van deze factoren zijn sociale invloeden, culturele achtergrond, individuele percepties en persoonlijke ervaringen (Ajzen & Fishbein, 1977; Eagly & Chaiken, 1993). De grote subjectiviteit van individuele factoren, die sterk afhankelijk zijn van persoonlijke ervaringen en interpretaties, draagt mogelijk bij aan de variabiliteit in beoordelingen van de attitude. Deze variabiliteit maakt het lastig om, binnen de context van geslachtsverschillen, duidelijke en significante patronen te ontdekken. Het is aannemelijk dat deze variabiliteit een negatieve impact heeft op de waarneming van verschillen tussen mannen en vrouwen, wat mogelijk de niet significante resultaten verklaart.

Het leerjaar en de verwachtingen en beoordelingscriteria van leerlingen zijn nauw met elkaar verbonden. Dit kan mogelijk verklaren waarom bovenbouwleerkrachten een hogere attitudebeoordeling ontvangen dan onderbouwleerkrachten. De verschillen van onderwijsmethoden en benaderingen in de onder- en bovenbouw zijn groot en kunnen impact hebben op de vorming en beoordeling van attitudes (Hattie, 2008). Beoordelingen worden

gevormd door relationele vaardigheidsaspecten zoals betrokkenheid, empathie, beschikbaarheid maar ook door gebruikte lesmaterialen. Indien leerlingen in verschillende leerjaren hun beoordeling baseren op andere aspecten, kan er een afwijking ontstaan in de attitudebeoordelingen en draagt dit bij aan de waargenomen verschillen tussen onder- en bovenbouwleerlingen.

Er zijn geen verschillen gevonden in de beoordelingen van het leerkracht-functioneren tussen leerkrachten die alleen voor de klas staan en leerkrachten in klassen met meerdere klassen. Deze bevinding kan worden mogelijk worden toegeschreven aan de complexiteit van teamteaching. Onderzoek van Ermeling (2010) wijst uit dat teamteaching, waarbij meerdere leerkrachten gezamenlijk verantwoordelijk zijn voor een klas, kan leiden tot diverse onderwijsstijlen en dynamieken in de klas. Hoewel de interactie tussen leerkrachten kan de leeromgeving voor kinderen kan verrijken, kan het ook leiden tot variabiliteit in de manier waarop leerlingen de prestaties van individuele leerkrachten beoordelen. Deze variabiliteit kan bijdragen aan het ontbreken van significante resultaten.

De positieve beoordelingen van KiVa-leerkrachten ten opzichte van leerkrachten uit de controlegroep komen overeen met de bevindingen uit het onderzoek van Huising & Veenstra (2011). In hun studie observeerden zij eveneens dat leerkrachten op KiVa-scholen een negatievere houding ten opzichte van pesten hadden (attitude), effectiever optraden tegen pestgedrag (bekwaamheid) en daadwerkelijk veel inspanningen leverden om pesten te verminderen (inspanning).

Het algemene welbevinden van leerlingen in de klas beïnvloedt de beoordelingen die leerlingen geven over het leerkracht-functioneren op positieve wijze. Deze conclusie wordt ondersteund door eerdere onderzoeken. Jones et al. (2013) bevestigen de impact van het algemene welbevinden op de beoordeling van leerkrachten. Daarnaast lijken deze resultaten parallel te lopen met de bevindingen uit de studies van Levine & Lezotte (1990) en Sammons (1997), die de invloed van welbevinden op onderwijsprestaties van leerkrachten onderzochten. Hoewel de focus van deze onderzoeken niet specifiek gericht waren op de beoordeling van leerkrachten, is het redelijk om aan te nemen dat een positief welbevinden van leerlingen kan bijdragen aan een positievere perceptie van de onderwijsomgeving, wat op zijn beurt invloed kan hebben op de beoordeling van leerkrachten.

Er ontbreekt een relatie tussen de mate van slachtofferschap in de klas en beoordelingen van leerlingen over het leerkracht-functioneren. Een mogelijke verklaring hiervoor zou kunnen zijn dat deze twee variabelen verschillende constructen meten die geen

onderlinge verbinding tonen. De literatuur wijst uit dat verschillende factoren de waarneming van leerkracht-functioneren beïnvloeden. Voorbeelden van deze factoren zijn een pedagogische aanpak, het klasmanagement en communicatievaardigheden (Brophy, 1983; Stronge, 2002). Slachtofferschap wordt daarentegen meestal beïnvloed door interpersoonlijke relaties die leerlingen met elkaar hebben en kunnen daarnaast voortkomen uit sociale dynamieken binnen de klas (Salmivalli, 2010). Het ontbreken van een significante correlatie tussen deze variabelen kan worden verklaard door de complexiteit van de onderliggende factoren die zowel de perceptie van het leerkracht-functioneren als de ervaring van slachtofferschap beïnvloeden.

5.2.1 Interpretatie resultaten

De resultaten kunnen worden geïnterpreteerd als een sterke indicatie van de validiteit en betrouwbaarheid van de leerlingrapportages over het functioneren van KiVa-leerkrachten. De bevindingen die in lijn zijn met de hypothesen, waaronder de verwachting rondom hogere beoordelingen voor KiVa-leerkrachten, geslachtsverschillen bij leerkrachten en leerlingen, de invloed van het KiVa-programma op beoordelingen, en de relatie tussen leerkracht-functioneren en klassenklimaat, versterken het vertrouwen in de nauwkeurigheid en consistentie van de door leerlingen verstrekte beoordelingen.

De consistentie tussen verwachtingen en gerapporteerde ervaringen van leerlingen onderstrepen dat leerlingrapportages een waardevol instrument is voor het beoordelen van het functioneren van KiVa-leerkrachten. Tegelijkertijd valt het op dat de verkregen beoordelingen soms weinig van elkaar verschillen, wat enige voorzichtigheid vereist bij de interpretatie van de resultaten. Ondanks deze kleine variabiliteit tonen de positieve effecten van het KiVa-programma op de percepties van leerlingen over het functioneren van hun leerkracht aan dat deze rapportages kunnen dienen als een betrouwbaar en valide instrumenten om de effectiviteit van het KiVa-programma te meten.

Deze bevindingen dragen bij aan het onderzoek van Veenstra et al. (2014), aangezien het de uitspraak ondersteunt dat wanneer leerkrachten over de juiste houding, vaardigheden en hulpmiddelen beschikken, ze een primaire rol spelen bij het voorkomen en reduceren van pestgedrag. De resultaten van dit onderzoek maken deze bewering geloofwaardiger en aannemelijker in de context van het KiVa-programma.

5.3 Beperkingen & Aanbevelingen

Bij het implementeren van de aanbevelingen, is het belangrijk om rekening te houden met drie kanttekeningen die van toepassing kunnen zijn op het onderzoek. Deze kanttekeningen beïnvloeden de interpretatie, representativiteit, validiteit en generaliseerbaarheid van de bevindingen. Op basis van deze worden tevens aanbevelingen gedaan voor toekomstig onderzoek.

Een **eerste kanttekening** die in dit onderzoek geplaatst kan worden, heeft betrekking op de beperkte beschikbaarheid van gegevens voor de steekproef. Deze beperking wordt voornamelijk veroorzaakt doordat veel leerkrachten niet op alle drie de meetmomenten hebben deelgenomen aan de monitor, mogelijk door het gevolg van het verloop op scholen. Hierdoor konden slechts een beperkt aantal leerkrachten worden opgenomen in de steekproef. Voor toekomstig onderzoek is het aan te bevelen om vooraf een strategie vast te stellen om dit probleem te voorkomen, met als doel een meer consistentere en representatievere steekproef te verkrijgen.

De complexiteit van de dataset, voortkomend uit het feit dat er rekening gehouden moet worden met verschillende niveaus, vormde een uitdaging in dit onderzoek en levert meteen een **tweede kanttekening** op. De focus van dit onderzoek richt zich op het leerkracht-niveau, terwijl andere relevante niveaus zoals het school-, klassen- en individueel niveau buiten beschouwing zijn gelaten. Deze keuze kan resulteren in onvolledige resultaten, gezien het feit dat het KiVa-programma ontworpen is om pestgedrag aan te pakken op alle niveaus. De complexiteit werd verder vergroot door de statistische afhankelijkheden binnen de data, wat de interpretatie bemoeilijkt. Een illustratief voorbeeld hiervan is de roulatie van leerkrachten gedurende het onderzoek, waarbij sommige leerkrachten lesgeven aan verschillende klassen en dus door verschillende leerlingen werden beoordeeld. Aan de andere kant werden sommige leerkrachten, bijvoorbeeld in gecombineerde klassen zoals groep 5/6, tweemaal beoordeeld door dezelfde leerlingen. Deze situatie deed zich voor tijdens de eerste wave (T1) door leerlingen van groep 5 en opnieuw tijdens de derde wave (T3) door dezelfde leerlingen die inmiddels in groep 6 zaten. Het is belangrijk om op te merken dat indien deze klassen als uitdagend werden ervaren, dit mogelijk een ongunstige invloed had op de beoordeling van de betreffende leerkrachten en tevens de generalisatie van de bevindingen beperken. Een gedifferentieerde analyse, waarin alle verschillende niveaus worden meegenomen, zou een dieper inzicht verschaffen kunnen bieden in de dynamiek en effecten van de KiVa-interventie. Daarnaast beperkt het analyseren van alleen het leerkracht-niveau de mogelijkheid om de interacties tussen de verschillende niveaus in kaart te brengen. Een mogelijk voorbeeld van een dergelijke interactie is dat, ondanks dat er positieve

veranderingen zijn in de houding van leerlingen, er geen zichtbare veranderingen zijn binnen de klassendynamiek. Het is daarom aan te bevelen om in toekomstig onderzoek een analyse uit te voeren op verschillende niveaus van het KiVa-programma om een optimaal beeld te krijgen van de invloeden van de interventie. Om de statistische afhankelijkheden te minimaliseren, is het aan te bevelen bij het ontwerp van het onderzoek strategieën te implementeren die de invloed van bijvoorbeeld roulerende leerkrachten op de data beperken. Dit kan gedaan worden door methoden toe te passen die rekening houden met deze specifieke dynamieken, aangezien het niet haalbaar lijkt om leerkrachten gedurende het gehele onderzoek aan dezelfde klassen toe te wijzen. Deze aanpak zou de mogelijkheid bieden om toekomstige onderzoeken te versterken, resulterend in robuustere bevindingen en een verhoogde algemene geldigheid van de resultaten.

Een **derde**, en tevens laatste, **kanttekening** die in dit onderzoek geplaatst kan worden, betreft het gebruik van verouderde data, waarbij de gegevens dateren van vóór 2014. Het onderwijs kenmerkt zich als een dynamische omgeving waarin voortdurend veranderingen plaatsvinden. In 2015 trad de wet sociale veiligheid in werking, wat resulteerde in wijzigingen in de context waarin het KiVa-programma werd geïmplementeerd. Deze wet schrijft voor dat scholen een beleid voor sociale veiligheid moeten uitvoeren, een specifiek aanspreekpunt voor pesten moeten en iemand moeten aanstellen die het anti-pestbeleid op school coördineert. Bovendien dienen scholen de ervaringen van veiligheid en het welzijn van hun leerlingen actief te monitoren. Deze focus op de sociale veiligheid kan resulteren in een positieve invloed op de beoordeling van het leerkracht-functioneren door leerlingen omdat het leerkrachten stimuleert om proactief bij te dragen aan een veilige en ondersteunde leeromgeving. Leerlingen zouden mogelijk de betrokkenheid, effectievere communicatie en actievere deelname van leerkrachten aan het anti-pestbeleid positiever kunnen beoordelen, waardoor het belangrijk is om in acht te nemen dat het onderzoek een momentopname was die mogelijk niet gelijk is aan de huidige onderwijssituatie. Om de daadwerkelijke impact van de wetsinvoering op de beoordeling van het leerkracht-functioneren te begrijpen, wordt aanbevolen om toekomstig onderzoek uit te voeren met gegevens die beschikbaar zijn vanaf 2015 en later.

Hoewel de resultaten van dit onderzoek een sterke indicatie geven van de validiteit en betrouwbaarheid van leerlingrapportages over het functioneren van KiVa-leerkrachten, vormt het ook een basis voor verdere verkenning van de nauwkeurigheid van deze rapportages. De bevindingen wijzen uit dat de beoordelingen soms weinig van elkaar verschillen, wat als

vertretpunt kan dienen voor toekomstig onderzoek om diepgaander in te gaan op de complexiteit van leerlingrapportages en leerkrachtbeoordelingen. Het doel van dit onderzoek is bij te dragen aan het begrip en verdere optimalisatie van het KiVa-rapport.

Literatuurlijst

- Adi, Y., Killoran, A., Janmohamed, K., & Stewart-Brown, S. (2007). Systematic review of the effectiveness of interventions to promote mental wellbeing in children in primary education: Report 1: Universal Approaches Non-violence related outcomes. *The University of Warwick*. <https://www.ncbi.nlm.nih.gov/books/NBK73674/>
- Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918.
<https://doi.org/10.1037/0033-2909.84.5.888>
- Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems: ‘Much ado about nothing’? *Psychological Medicine*, *40*(5), 717–729.
<https://doi.org/10.1017/s0033291709991383>
- Bem, L. S. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. <https://doi.org/10.1037/h0036215>
- Berk, L. (2015). *Child development*. Pearson Higher Education AU.
- Bradshaw, C. P. (2015). Translating research to practice in bullying prevention. *American Psychologist*, *70*(4), 322–332. <https://doi.org/10.1037/a0039114>
- Bradshaw, C. P., Sawyer, A., & O’Brennan, L. M. (2007). Bullying and Peer Victimization at School: Perceptual Differences Between Students and School Staff. *School Psychology Review*, *36*(3), 361–382.
<https://doi.org/10.1080/02796015.2007.12087929>
- Bradshaw, C. P., Waasdorp, T. E., & Johnson, S. L. (2014). Overlapping verbal, relational, physical, and electronic forms of bullying in adolescence: Influence of school context. *Journal of Clinical Child and Adolescent Psychology*, *44*(3), 494–508.
<https://doi.org/10.1080/15374416.2014.893516>

- Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of School-Wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*(5), e1136–e1145. <https://doi.org/10.1542/peds.2012-0243>
- Brophy, J. (1983). Research on the self-fulfilling prophecy and teacher expectations. *Journal of Educational Psychology*, *75*(5), 631–661. <https://doi.org/10.1037/0022-0663.75.5.631>
- Caliński, T., Moore, D. S., & McCabe, G. P. (1990). Introduction to the practice of statistics. *Biometrics*, *46*(3), 884. <https://doi.org/10.2307/2532120>
- Connell, R. W. (2005). *Masculinities: Second Edition*. Univ of California Press.
- Cook, C. R., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: A meta-analytic investigation. *School Psychology Quarterly*, *25*(2), 65–83. <https://doi.org/10.1037/a0020149>
- Craig, W. M., & Pepler, D. (2003). Identifying and Targeting Risk for Involvement in Bullying and Victimization. *The Canadian Journal of Psychiatry*, *48*(9), 577–582. <https://doi.org/10.1177/070674370304800903>
- Cronbach, L. J., & Meehl, P. E. (1955). *Construct validity in psychological tests*. (Vol. 52). American Psychological Association. <https://doi.org/10.1037/h0040957>
- De Graaf, N. D., & Wiertz, D. (2019). *Societal Problems as Public Bads*. Routledge.
- Driessen, G., & Van Langen, A. (2010). De onderwijssachterstand van jongens. Omvang, oorzaken en oplossingen. *Radboud Universiteit Nijmegen*. <https://doi.org/10.13140/RG.2.1.2402.9848>
- Eagly, A. H., & Chaiken, S. (1997). The psychology of attitudes. *Journal of Marketing Research*, *34*(2), 298. <https://doi.org/10.2307/3151869>

- Ermeling, B. A. (2010). Tracing the effects of teacher inquiry on classroom practice. *Teaching and Teacher Education*, 26(3), 377–388. <https://doi.org/10.1016/j.tate.2009.02.019>
- Espelage, D. L., & Holt, M. K. (2013). Suicidal Ideation and School Bullying Experiences After Controlling for Depression and Delinquency. *Journal of Adolescent Health*, 53(1), S27–S31. <https://doi.org/10.1016/j.jadohealth.2012.09.017>
- Espelage, D. L., & Swearer, S. M. (2004). Bullying in American schools : a social-ecological perspective on prevention and intervention. In *HAL (Le Centre pour la Communication Scientifique Directe)*. Lawrence Erlbaum Associates Publishers.
- Espelage, D. L., & Swearer, S. M. (2008). Addressing research gaps in the intersection between homophobia and bullying. *School Psychology Review*, 37(2), 155–159.
- Espelage, D. L., & Swearer, S. M. (2011). Bullying in North American Schools. In *Routledge eBooks*. Informa. <https://doi.org/10.4324/9780203842898>
- Furrer, C., & Skinner, E. A. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148–162. <https://doi.org/10.1037/0022-0663.95.1.148>
- Geerdink, G. (2007). *Diversiteit op de pabo. Sekseverschillen in motivatie, curriculumperceptie en studieresultaten*. <http://repository.ubn.ru.nl/handle/2066/54565>
- Gini, G., & Pozzoli, T. (2009). Association Between Bullying and Psychosomatic Problems: A Meta-analysis. *Pediatrics*, 123(3), 1059–1065. <https://doi.org/10.1542/peds.2008-1215>
- Hattie, J. (2008). *Visible Learning: a synthesis of over 800 Meta-Analyses relating to Achievement*. https://minerva-access.unimelb.edu.au/bitstream/11343/31622/1/281182_161493.pdf

- Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years' Research on Peer Victimization and Psychosocial Maladjustment: a meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry*, *41*(4), 441–455.
<https://doi.org/10.1111/1469-7610.00629>
- Holfve-Sabel, M. (2014). Learning, Interaction and Relationships as Components of Student Well-being: Differences Between Classes from Student and Teacher Perspective. *Social Indicators Research*, *119*(3), 1535–1555. <https://doi.org/10.1007/s11205-013-0557-7>
- Huitsing, G., Lodder, G. M. A., Browne, W. J., Oldenburg, B., Van Der Ploeg, R., & Veenstra, R. (2020). A Large-Scale Replication of the Effectiveness of the KiVa Antibullying Program: a Randomized Controlled Trial in the Netherlands. *Prevention Science*, *21*(5), 627–638. <https://doi.org/10.1007/s11121-020-01116-4>
- Huitsing, G., & Veenstra, R. (2012). Het KiVa-antipestprogramma: de cruciale rol van de groep. In: Goossens, F., Vermande, M., & Van der Meulen, M. (Eds.), *Pesten op school* (pp. 152-159). Amsterdam: Boom.
- Ince, D. (2022, September 22). *Cijfers over pesten: slachtoffers* | Nederlands Jeugdinstituut. NJI. <https://www.nji.nl/cijfers/pesten-slachtoffers>
- Ingersoll, R., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers. *Review of Educational Research*, *81*(2), 201–233.
<https://doi.org/10.3102/0034654311403323>
- Jimerson, S. R., Swearer, S. M., & Espelage, D. L. (2009). *International Handbook of School Bullying: An International Perspective*. Routledge.
- Jones, S. M., Bouffard, S. M., & Weissbourd, R. (2013). Educators' Social and Emotional Skills Vital to Learning. *Phi Delta Kappan*, *94*(8), 62–65.
<https://doi.org/10.1177/003172171309400815>

- Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A Large-Scale Evaluation of the KiVa Antibullying Program: Grades 4-6. *Child Development, 82*(1), 311–330. <https://doi.org/10.1111/j.1467-8624.2010.01557.x>
- Klomek, A. B., Sourander, A., & Gould, M. S. (2010). The association of suicide and bullying in childhood to young adulthood: a review of cross-sectional and longitudinal research findings. *The Canadian Journal of Psychiatry, 55*(5), 282–288. <https://doi.org/10.1177/070674371005500503>
- Ladd, G. W., & Burgess, K. B. (2001). Do Relational Risks and Protective Factors Moderate the Linkages between Childhood Aggression and Early Psychological and School Adjustment? *Child Development, 72*(5), 1579–1601. <https://doi.org/10.1111/1467-8624.00366>
- Levine, D. U., & Lezotte, L. W. (1990). *Unusually effective schools: A review and analysis of research and practice*. Madison, WI: The National Center for Effective Schools Research and Development.
- Lynch, M., & Cicchetti, D. (1997). Children's relationships with adults and peers: An examination of elementary and junior high school students. *Journal of School Psychology, 35*(1), 81–99. [https://doi.org/10.1016/s0022-4405\(96\)00031-3](https://doi.org/10.1016/s0022-4405(96)00031-3)
- Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23*(1), 26–42. <https://doi.org/10.1037/1045-3830.23.1.26>
- Ministerie van Onderwijs, Cultuur en Wetenschap. (2022, March 10). *Evaluatie Wet veiligheid op school Eindrapportage*. Kamerstuk | Rijksoverheid.nl. <https://www.rijksoverheid.nl/documenten/kamerstukken/2021/11/11/evaluatie-wet-veiligheid-op-school-eindrapport>

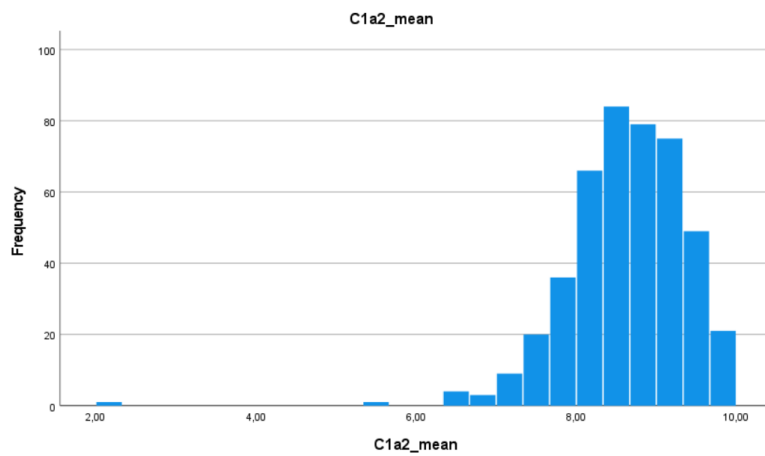
- Mishna, F., Wiener, J., & Pepler, D. (2008). Some of My Best Friends: Experiences of Bullying Within Friendships. *School Psychology International*, 29(5), 549–573.
<https://doi.org/10.1177/0143034308099201>
- Nansel, T. R., Craig, W. M., Overpeck, M. D., Saluja, G., & Ruan, W. J. (2004). Cross-national Consistency in the Relationship Between Bullying Behaviors and Psychosocial Adjustment. *Archives of Pediatrics & Adolescent Medicine*, 158(8), 730.
<https://doi.org/10.1001/archpedi.158.8.730>
- Nansel, T. R., Overpeck, M. D., Pilla, R. S., Ruan, W. J., Simons-Morton, B. G., & Scheidt, P. C. (2001). Bullying Behaviors Among US Youth. *JAMA*, 285(16), 2094.
<https://doi.org/10.1001/jama.285.16.2094>
- Olweus, D. (1994). Bullying at school: what we know and what we can do. *Choice Reviews Online*, 32(02), 32–1080. <https://doi.org/10.5860/choice.32-1080>
- Olweus, D. (1996). Bully/victim problems in school. *Prospects*, 26(2), 331–359.
<https://doi.org/10.1007/bf02195509>
- Olweus, D. (2013). School Bullying: Development and Some Important Challenges. *Annual Review of Clinical Psychology*, 9(1), 751–780. <https://doi.org/10.1146/annurev-clinpsy-050212-185516>
- Orpinas, P., Horne, A. M., & Staniszewski, D. (2003). School bullying: Changing the problem by changing the school. *School Psychology Review*, 32(3), 431–444.
<https://doi.org/10.1080/02796015.2003.12086210>
- Pellegrini, A. D., & Long, J. D. (2002). A longitudinal study of bullying, dominance, and victimization during the transition from primary school through secondary school. *British Journal of Development Psychology*, 20(2), 259–280.
<https://doi.org/10.1348/026151002166442>

- Piaget, J. (1972). Intellectual Evolution from Adolescence to Adulthood. *Human Development, 15*(1), 1–12. <https://doi.org/10.1159/000271225>
- Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-Child relationships and children's success in the first years of school. *School Psychology Review, 33*(3), 444–458. <https://doi.org/10.1080/02796015.2004.12086261>
- Prentice, D. A., & Carranza, E. (2002). What Women and Men Should Be, Shouldn't be, are Allowed to be, and don't Have to Be: The Contents of Prescriptive Gender Stereotypes. *Psychology of Women Quarterly, 26*(4), 269–281. <https://doi.org/10.1111/1471-6402.t01-1-00066>
- Reijntjes, A., Vermande, M., Goossens, F. A., Olthof, T., Van De Schoot, R., Aleva, L., & Van Der Meulen, M. (2013). Developmental trajectories of bullying and social dominance in youth. *Child Abuse & Neglect, 37*(4), 224–234. <https://doi.org/10.1016/j.chiabu.2012.12.004>
- Ridgeway, C. L. (2011). Framed by gender: how gender inequality persists in the modern world. *Choice Reviews Online, 49*(03), 49–1789. <https://doi.org/10.5860/choice.49-1789>
- Rimm-Kaufman, S. E., & Pianta, R. C. (2000). An ecological perspective on the transition to kindergarten. *Journal of Applied Developmental Psychology, 21*(5), 491–511. [https://doi.org/10.1016/s0193-3973\(00\)00051-4](https://doi.org/10.1016/s0193-3973(00)00051-4)
- Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective Teacher-Student relationships on students' school engagement and achievement. *Review of Educational Research, 81*(4), 493–529. <https://doi.org/10.3102/0034654311421793>
- Sadker, M., & Sadker, D. M. (1994). *Failing at fairness: How America's Schools Cheat Girls*. Scribner Book Company.

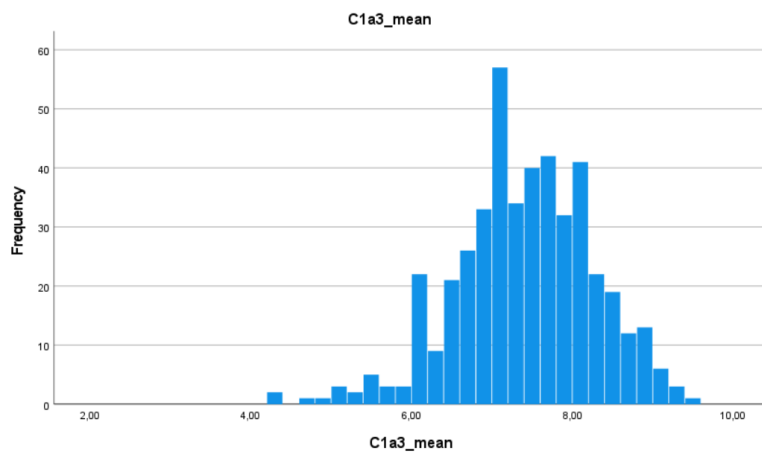
- Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15*(2), 112–120. <https://doi.org/10.1016/j.avb.2009.08.007>
- Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010). Development, Evaluation, and Diffusion of a National Anti-Bullying Program, KiVa. In *Routledge eBooks*. Informa. <https://doi.org/10.4324/9780203866412.ch12>
- Salmivalli, C., & Peets, K. (2009). Bullies, victims, and bully-victim relationships in middle childhood and early adolescence. *Handbook of Peer Interactions, Relationships, and Groups*.
- Sammons, P. (1997). *Key Characteristics of Effective schools: A review of school effectiveness research*.
- Sentse, M., Kretschmer, T., & Salmivalli, C. (2015). The Longitudinal Interplay between Bullying, Victimization, and Social Status: Age-related and Gender Differences. *Social Development, 24*(3), 659–677. <https://doi.org/10.1111/sode.12115>
- Siegler, R. S., DeLoache, J. S., & Eisenberg, N. (2011). *How children develop*. Macmillan.
- Stevens, G. W., Van Dorsselaer, S., Boer, M., De Roos, S., Duinhof, E. L., Ter Bogt, T. F. M., Van Den Eijnden, R., Kuyper, L., Visser, D., Vollebergh, W. a. M., & De Looze, M. (2018). HBSC 2017. Gezondheid en welzijn van jongeren in Nederland. In *Utrecht University eBooks*. Utrecht University. <https://dspace.library.uu.nl/handle/1874/376901>
- Stronge, J. H. (2002). *Qualities of effective teachers*. http://www.hunter.cuny.edu/shp/centers/nycnect/greenteam/docs/qualities_of_effective_teachers.pdf
- Suldo, S. M., & Shaffer, E. J. (2008). Looking Beyond Psychopathology: The Dual-Factor Model of Mental Health in Youth. *School Psychology Review, 37*(1), 52–68. <https://doi.org/10.1080/02796015.2008.12087908>

- Swearer, S. M., Espelage, D. L., Vaillancourt, T., & Hymel, S. (2010). What Can Be Done About School Bullying? *Educational Researcher*, 39(1), 38–47.
<https://doi.org/10.3102/0013189x09357622>
- Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace Research*, 3(2), 63–73. <https://doi.org/10.1108/17596591111132873>
- Veenstra, R. (2015). Signaleren en tegengaan van pesten: het KiVa antipestprogramma: Eindrapportage voor Onderwijs Bewijs. *Rijksuniversiteit Groningen*.
- Veenstra, R., Lindenberg, S., Huitsing, G., Sainio, M., & Salmivalli, C. (2014). The role of teachers in bullying: The relation between antibullying attitudes, efficacy, and efforts to reduce bullying. *Journal of Educational Psychology*, 106(4), 1135–1143.
<https://doi.org/10.1037/a0036110>
- Vermande, M., Van Der Meulen, M., & Reijntjes, A. H. A. (2015). *Pesten op school: achtergronden en interventies*.
- Wang, M., & Eccles, J. S. (2011). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence*, 22(1), 31–39.
<https://doi.org/10.1111/j.1532-7795.2011.00753.x>
- Wentzel, K. R., & Wigfield, A. (2009). Handbook of Motivation at School. In *Routledge eBooks*. <https://doi.org/10.4324/9780203879498>
- Zych, I., Farrington, D. P., & Ttofi, M. M. (2019). Protective factors against bullying and cyberbullying: A systematic review of meta-analyses. *Aggression and Violent Behavior*, 45, 4–19. <https://doi.org/10.1016/j.avb.2018.06.008>

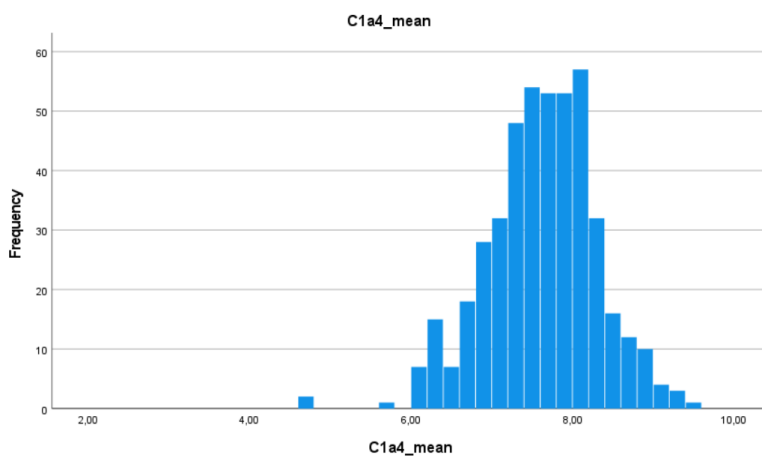
Bijlage I. Grafische weergave variabelen leerkracht-functioneren



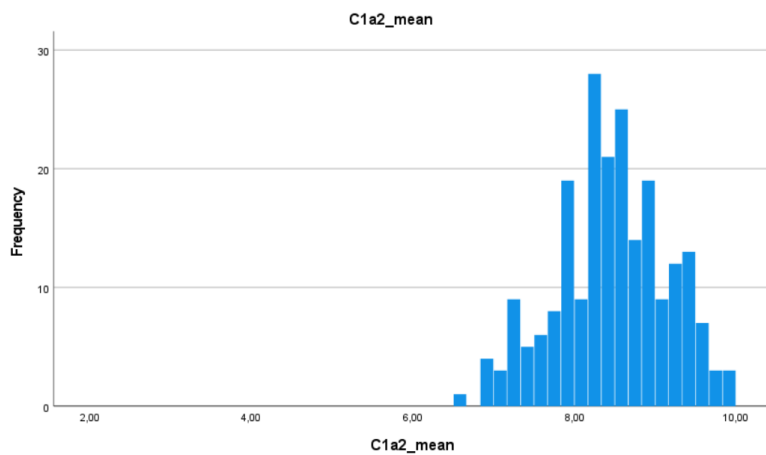
Figuur 8 Verdeling attitude leerkracht 1 tijdens wave 1 (T1)



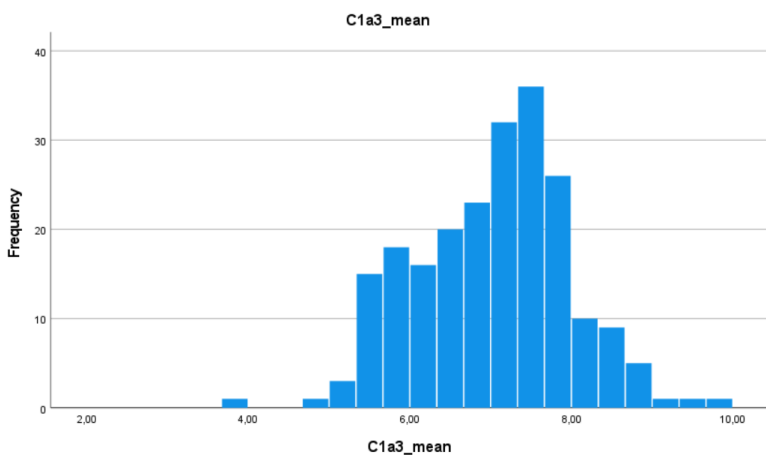
Figuur 9 Verdeling bekwaamheid leerkracht 1 tijdens wave 1 (T1)



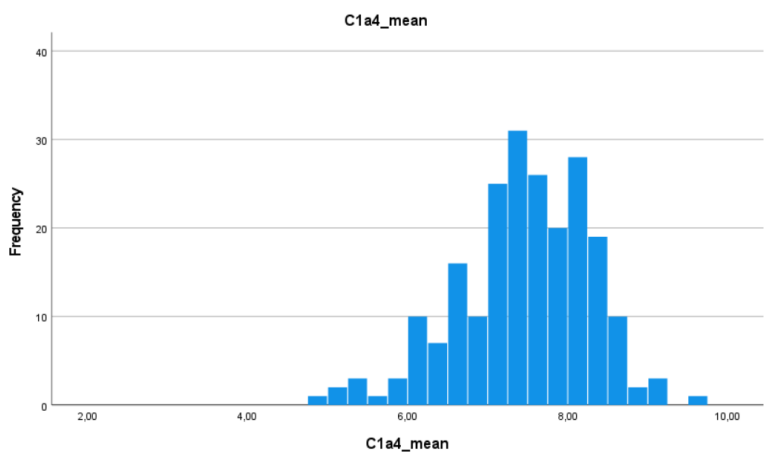
Figuur 10 Verdeling inspanning leerkracht 1 tijdens wave 1 (T1)



Figuur 11 Verdeling attitude leerkracht 2 tijdens wave 1 (T1)

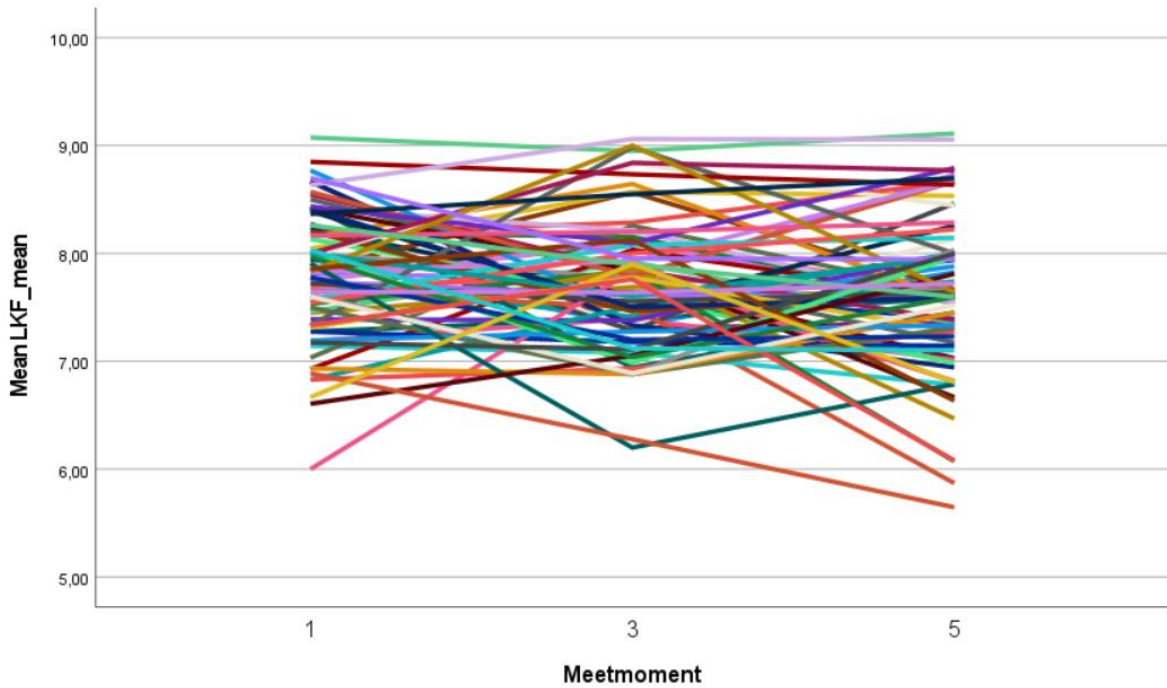


Figuur 12 Verdeling bekwaamheid leerkracht 2 tijdens wave 1 (T1)

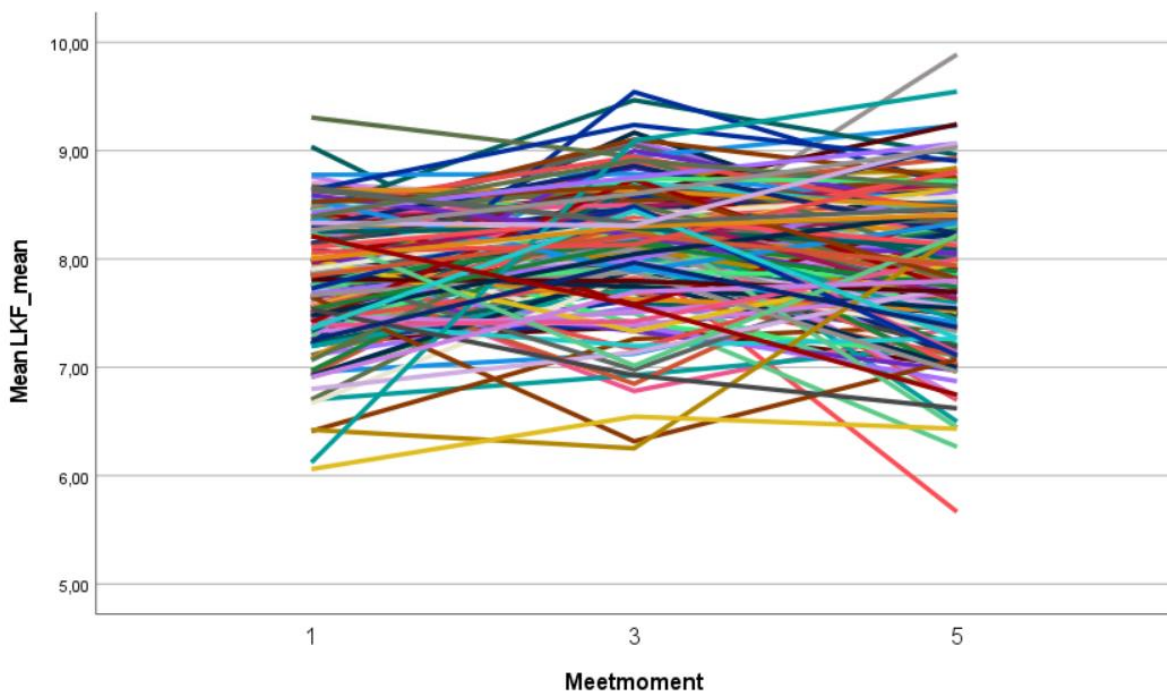


Figuur 13 Verdeling inspanning leerkracht 2 tijdens wave 1 (T1)

Bijlage II. Grafische weergave ontwikkeling leerkracht-functioneren



Figuur 14 Visualisatie ontwikkeling leerkrachten uit controlegroep



Figuur 15 Visualisatie ontwikkeling KiVa-leerkrachten

Bijlage III. ANOVA-analyse en regressieanalyse

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
C5wb_mean	Between Groups	,670	4	,168	4,337	,002
	Within Groups	9,154	237	,039		
	Total	9,824	241			
C5ss_pct	Between Groups	230,865	4	57,716	,397	,811
	Within Groups	34472,187	237	145,452		
	Total	34703,052	241			

Figuur 16 ANOVA-tabel welbevinden en slachtofferschap (T5)

Multiple Comparisons							
Tukey HSD							
Dependent Variable	(I) Q_groups	(J) Q_groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
C5wb_mean	1,00	2,00	-,00421	,04349	1,000	-,1237	,1153
		3,00	-,05256	,03650	,602	-,1529	,0478
		4,00	-,14956*	,03903	,002	-,2568	-,0423
		5,00	-,07595	,03650	,232	-,1763	,0244
	2,00	1,00	,00421	,04349	1,000	-,1153	,1237
		3,00	-,04835	,04490	,818	-,1718	,0751
		4,00	-,14535*	,04698	,019	-,2745	-,0162
		5,00	-,07174	,04490	,500	-,1952	,0517
	3,00	1,00	,05256	,03650	,602	-,0478	,1529
		2,00	,04835	,04490	,818	-,0751	,1718
		4,00	-,09700	,04060	,122	-,2086	,0146
		5,00	-,02339	,03818	,973	-,1283	,0816
	4,00	1,00	,14956*	,03903	,002	,0423	,2568
		2,00	,14535*	,04698	,019	,0162	,2745
		3,00	,09700	,04060	,122	-,0146	,2086
		5,00	,07361	,04060	,368	-,0380	,1852
	5,00	1,00	,07595	,03650	,232	-,0244	,1763
		2,00	,07174	,04490	,500	-,0517	,1952
		3,00	,02339	,03818	,973	-,0816	,1283
		4,00	-,07361	,04060	,368	-,1852	,0380
C5ss_pct	1,00	2,00	-,5086	2,6685	1,000	-7,844	6,827
		3,00	2,3436	2,2399	,833	-3,814	8,501
		4,00	,2053	2,3950	1,000	-6,378	6,789
		5,00	,9077	2,2399	,994	-5,249	7,065
	2,00	1,00	,5086	2,6685	1,000	-6,827	7,844
		3,00	2,8522	2,7555	,839	-4,722	10,427
		4,00	,7138	2,8830	,999	-7,211	8,639
		5,00	1,4163	2,7555	,986	-6,158	8,991
	3,00	1,00	-2,3436	2,2399	,833	-8,501	3,814
		2,00	-2,8522	2,7555	,839	-10,427	4,722
		4,00	-2,1383	2,4915	,912	-8,987	4,711
		5,00	-1,4359	2,3428	,973	-7,876	5,004
	4,00	1,00	-,2053	2,3950	1,000	-6,789	6,378
		2,00	-,7138	2,8830	,999	-8,639	7,211
		3,00	2,1383	2,4915	,912	-4,711	8,987
		5,00	,7024	2,4915	,999	-6,146	7,551
	5,00	1,00	-,9077	2,2399	,994	-7,065	5,249
		2,00	-1,4163	2,7555	,986	-8,991	6,158
		3,00	1,4359	2,3428	,973	-5,004	7,876
		4,00	-,7024	2,4915	,999	-7,551	6,146

*. The mean difference is significant at the 0.05 level.

Figuur 17 Post-hoc analyse welbevinden en slachtofferschap

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,423 ^a	,179	,172	,63259

a. Predictors: (Constant), KiVa condition, C5wb_mean

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	20,852	2	10,426	26,054	<,001 ^b
	Residual	95,640	239	,400		
	Total	116,493	241			

a. Dependent Variable: C5_mean_LKF

b. Predictors: (Constant), KiVa condition, C5wb_mean

Figuur 18 Modelschatting (ANOVA) overzicht leerkracht-functioneren met welbevinden (T5)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3,986	,638		6,251	<,001
	C5wb_mean	1,150	,202	,334	5,694	<,001
	KiVa condition	,367	,087	,247	4,210	<,001

a. Dependent Variable: C5_mean_LKF

Figuur 19 Modelschatting coëfficiënten leerkracht-functioneren met welbevinden (T5)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,260 ^a	,068	,060	,67407

a. Predictors: (Constant), KiVa condition, C5ss_pct

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7,897	2	3,949	8,690	<,001 ^b
	Residual	108,595	239	,454		
	Total	116,493	241			

a. Dependent Variable: C5_mean_LKF

b. Predictors: (Constant), KiVa condition, C5ss_pct

Figuur 20 Modelschatting (ANOVA) overzicht leerkracht-functioneren met slachtofferschap (T5)

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7,607	,099		76,783	<,001
	C5ss_pct	-,001	,004	-,014	-,215	,830
	KIVa condition	,383	,094	,258	4,073	<,001

a. Dependent Variable: C5_mean_LKF

Figuur 21 Modelschatting coëfficiënten leerkracht-functioneren met slachtofferschap (T5)