

# **Hypothesetoetsing in pedagogisch onderzoek: een vergelijkend replicatieonderzoek naar de bewijsvoering van klassieke en Bayesiaanse statistiek**

Naam: Agnes T. Naber (S5374391)

PAMAE021: Masterthesis Ethics of Education

Faculteit der Gedrags- en Maatschappijwetenschappen, Rijksuniversiteit Groningen

Naam eerste beoordelaar en supervisor: Dr. S. M. Field

Naam tweede beoordelaar: Dr. W. J. Post

Totaal aantal woorden: 7997

31 mei 2024

## Samenvatting

**Inleiding:** Het doel van de wetenschap is waarheidsvinding, waarbij de betrouwbaarheid van onderzoek kan worden getoetst door replicatie. Deze studie repliceert het pedagogisch onderzoek *Reducing bullying through empathy training: the effect of teacher's passive presence?* van Palade en Pascal (2023) om de verschillen tussen klassieke en Bayesiaanse statistiek in hypothese-ondersteuning te analyseren.

**Methode:** De beschrijvende statistieken, correlaties, ANOVA's en regressies uit de oorspronkelijke studie werden gerepliceerd. Daarnaast werden Bayesfactoren berekend om de bewijskracht voor zowel de nul- als alternatieve hypothese te evalueren.

**Resultaten:** De beschrijvende statistieken en correlaties waren reproduceerbaar, de ANOVA's en regressies niet. De Bayesiaanse resultaten waren grotendeels consistent met de klassieke bevindingen. Er werden weinig verschillen gevonden in de conclusies over de interventie-effecten op empathie, verbaal en fysiek pesten, en empathie als voorspeller van pesten. De Bayesiaanse analyse leverde echter een aanvullende conclusie op: er was geen interventie-effect op fysiek pesten.

**Conclusie/discussie:** Deze studie benadrukt het belang van replicatie en het gebruik van Bayesiaanse statistiek naast klassieke statistiek. De bevindingen tonen aan dat de overeenstemming tussen beide methoden het bewijs versterkt. Bovendien biedt de Bayesiaanse analyse extra inzichten door uitspraken mogelijk te maken over zowel de alternatieve hypothese (een interventie-effect) als de nulhypothese (geen interventie-effect). Dit resulteert in een completer beeld dan klassieke statistiek, die geen onderscheid maakt tussen het ontbreken van bewijs en bewijs van geen effect.

**Aanbevelingen:** Een conceptuele replicatie met een grotere steekproef en aanvullende variabelen, zoals de rol van de leerkracht en geslacht, wordt aanbevolen voor beter inzicht in de relatie tussen empathie, verbaal en fysiek pesten, en voor verbeterde generaliseerbaarheid van de bevindingen.

## Summary

**Introduction:** The goal of science is the pursuit of truth, with the reliability of research being testable through replication. This study replicates the pedagogical research of Palade and Pascal (2023), titled *Reducing bullying through empathy training: the effect of teacher's passive presence?*, to examine the differences between classical and Bayesian statistics in hypothesis testing.

**Methods:** Descriptive statistics, correlations, ANOVAs, and regressions from the original study were replicated. Additionally, Bayes factors were calculated to assess the evidence for both the null and alternative hypotheses.

**Results:** Descriptive statistics and correlations were reproducible, while ANOVAs and regressions were not. Bayesian results were largely consistent with classical findings. Few differences were found in the conclusions regarding intervention effects on empathy, verbal and physical bullying, and empathy as a predictor of bullying. However, Bayesian analysis yielded an additional conclusion: no intervention effect on physical bullying.

**Discussion/Conclusion:** This study underscores the value of replication and the use of Bayesian statistics alongside classical methods. The findings demonstrate that concordance between both approaches strengthens the evidence. Moreover, Bayesian analysis provides additional insights by enabling statements about both the alternative hypothesis (an intervention effect) and the null hypothesis (no intervention effect). This results in a more comprehensive picture compared to classical statistics, which cannot distinguish between the absence of evidence and evidence of no effect.

**Recommendations:** A conceptual replication with a larger sample size and additional variables, such as the role of the teacher and gender, is recommended for better insight into the relationship between empathy, verbal and physical bullying, and for improved generalisability of the findings.

## Inhoud

Inleiding .....	4
Betrouwbaar wetenschappelijk onderzoek.....	4
Klassieke statistiek .....	4
Bayesiaanse statistiek.....	6
Huidige onderzoek .....	7
Originele onderzoek.....	8
Dataset.....	10
Data-analyse.....	10
Herhaalde metingen ANOVA's .....	10
Meervoudige regressies.....	11
Bayesiaanse herhaalde metingen ANOVA's en meervoudige regressies.....	12
Replicatie.....	13
Resultaten.....	14
Effecten van de interventie op empathie, verbaal en fysiek pesten .....	14
Empathie.....	16
Verbaal pesten .....	17
Fysiek pesten.....	18
Empathie als voorspeller van verbaal of fysiek pesten .....	18
Overeenkomsten en verschillen onderzoek Palade en Pascal met huidige onderzoek .....	20
Overtuigende Bayesfactoren .....	20
Verschillen en overeenkomsten bij de conclusies.....	21
Conclusie en discussie .....	22
Sterktes en beperkingen van het onderzoek .....	23
Aanbevelingen.....	24
Referenties .....	26
Bijlage A. Beschrijvende statistieken en correlaties .....	31
Bijlage B. Klassieke en Bayesiaanse ANOVA's en assumptiecontroles.....	35
Bijlage C. Lineaire regressies.....	41
Bijlage D. Overzichten van significante en niet-significante resultaten en overtuigende en niet-overtuigende Bayesfactoren.....	52

## Inleiding

### Betrouwbaar wetenschappelijk onderzoek

Waarheidsvinding is het centrale doel van de wetenschap (Ritchie, 2020; Solomon, 2007). Wetenschappers zijn voortdurend op zoek naar nieuwe resultaten, deze vormen de drijvende kracht achter wetenschappelijke vooruitgang (Scheepers et al., 2016). Bij deze zoektocht trachten wetenschappers hun bevindingen te falsifiëren door plausibele alternatieve verklaringen te toetsen, in plaats van alleen informatie te zoeken die bevestiging geeft. Conclusies die bestand zijn tegen falsificatie worden beschouwd als plausibel, totdat overtuigender bewijsmateriaal wordt gevonden (Shadish et al., 2002; Van Dongen et al., 2023).

Toch heeft de zoektocht naar baanbrekende resultaten geleid tot veel onderzoeken die onbetrouwbaar zijn: deze steunen op zwak of niet-reproduceerbaar bewijs (Ioannidis, 2022; Pek & Van Zandt, 2019; Ritchie, 2020). Oorzaken hiervan zijn onder andere publicatiebias, de druk op wetenschappers om zo veel mogelijk te publiceren en *p*-hacking, waarbij in het onderzoek de focus ligt op het vinden van significante resultaten (Field et al., 2019; Ritchie, 2020).

Of een kwantitatief wetenschappelijk onderzoek betrouwbaar is, kan onderzocht worden door replicatie van het onderzoek (Field et al., 2019; Ritchie, 2020; Stroebe & Strack, 2014). Om onderzoek voor replicatie te selecteren, moet volgens Field et al. (2019) onder andere naar de statistische informatie van een gepubliceerd artikel gekeken worden.

### Klassieke statistiek

Bij kwantitatief onderzoek is het gebruik van klassieke statistiek de dominante aanpak (Westera, 2021). Hierbij worden onderzoeksvragen getest met behulp van een steekproef in een populatie: er wordt onderzocht of er een interventie-effect is in de populatie en hoe groot dit effect is (Goulet-Pelletier & Cousineau, 2018; Van de Schoot et al., 2017). Om een effect in de populatie te kunnen vinden, wordt statistisch bewijs verzameld. Hiervoor worden twee elkaar beconcurrerende hypothesen opgesteld om te toetsen: een nulhypothese en een alternatieve hypothese (Pek & Van Zandt, 2019; Van de Schoot et al., 2017). De nulhypothese stelt dat het effect gelijk is aan nul: na een interventie is er geen verschil tussen de experimentele groep en de controlegroep. De alternatieve hypothese stelt dat het effect ongelijk is aan nul: na een interventie is er verschil tussen de experimentele groep en de controlegroep.

Bij klassieke statistiek wordt gebruikgemaakt van statistische modellen gebaseerd op kansberekening (Flohr, 2012). Onder kans wordt hier de limiet van een relatieve frequentie verstaan (long run relative frequency), waarbij de herhaalbaarheid van het kansexperiment centraal staat (Flohr, 2012; Pek & Van Zandt, 2019; Van de Schoot et al., 2017). Aangezien bij inductief statistisch redeneren de gevonden steekproefuitkomst wordt geëxtrapoleerd naar de grotere en onbekende populatie, is er geen zekerheid dat de conclusies op basis van de steekproef juist zijn. De toevalsvariatie zorgt voor onzekerheid: bij herhaalde steekproeftrekking kunnen de steekproefuitkomsten verschillen als gevolg van toeval. Deze onzekerheid wordt uitgedrukt in kansen (Flohr, 2012).

Bij het toetsen van de hypothesen wordt de  $p$ -waarde gebruikt als een maat voor de mate van bewijs, dit is een waarde tussen 0 en 1 (Waterink et al., 2021). Een  $p$ -waarde is de kans op het verkrijgen van het waargenomen effect in de steekproef of een nog extremere uitkomst, bij de veronderstelling dat de nulhypothese waar is (Goulet-Pelletier & Cousineau, 2018; Flohr, 2012). Hoe kleiner deze  $p$ -waarde is, hoe groter de kans dat de nulhypothese wordt verworpen en de alternatieve hypothese wordt aangenomen (Waterink et al., 2021).

De beslissende  $p$ -waarde wordt het significantieniveau genoemd, dit is het vooraf gespecificeerd Type I-fout niveau — een fout-positieve conclusie — aangeduid met  $\alpha$  (Moore & McCabe, 2006; Van Dongen et al., 2023). Hiervoor wordt vaak een grenswaarde van  $\alpha = 0.05$  of lager gehanteerd: dit betekent dat de kans dat een dergelijk effect (of nog extremer) aan het toeval is te wijten, oftewel een vals positief geeft, kleiner is dan 5% (Pashler & Harris, 2012; Waterink et al., 2021). Als de overschrijdingskans kleiner of gelijk is aan  $\alpha$ , dan wordt gesteld dat de data statistisch significant zijn op niveau  $\alpha$ . Hoewel het alfaniveau van 0.05 het meest frequent wordt gebruikt, kunnen er redenen zijn om hogere of lagere niveaus te gebruiken in specifieke gevallen (Shadish et al., 2002).

Tussen significant en niet-significant is geen scherpe scheidslijn; de evidentie neemt geleidelijk toe naarmate de overschrijdingskans kleiner wordt. Het is echter gebruikelijk om een dichotome beslissing te rapporteren (Moore & McCabe, 2006).

Wanneer de conclusie is dat een steekproef geen significant resultaat oplevert, dan betekent dit dat de sterkte van het bewijs de door  $\alpha$  gestelde standaard niet heeft bereikt en er onvoldoende ondersteuning is om de nulhypothese te verwerpen (Moore & McCabe, 2006). Hieruit blijkt echter niet of de nulhypothese wordt ondersteund (Flohr, 2012; Mulder & Wagenmakers, 2016; Pek & Van Zandt, 2019; Waterink et al., 2021).

## Bayesiaanse statistiek

Naast de klassieke statistiek bestaan er verschillende andere manieren van statistisch redeneren, waaronder de Bayesiaanse statistiek (Flohr, 2012; Mulder & Wagenmaker, 2016). Net als in de klassieke statistiek wordt in de Bayesiaanse statistiek gebruikgemaakt van een nulhypothese ( $H_0$ ) en een alternatieve hypothese ( $H_A$ ) (Field et al., 2016; Pek & Van Zandt, 2019). Bij de Bayesiaanse statistiek wordt onderzocht welke van de twee concurrerende hypothesen de data het beste verklaart, dit wordt uitgedrukt via de Bayesfactor (Stevens & Meester, 2021; Waterink et al., 2021). Hierbij wordt gebruikgemaakt van voorwaardelijke kansen: een kans meet de relatieve waarschijnlijkheid van een gebeurtenis op dit moment (Flohr, 2012).

Bij een Bayesiaanse statistische analyse wordt eerst een a-priori kansverdeling (prior) opgesteld op basis van de reeds aanwezige kennis van de onderzoekers, bijvoorbeeld gebaseerd op theoretische inzichten of bekende feiten. Vervolgens worden nieuwe kwantitatieve gegevens gebruikt voor het actualiseren en herzien van de prior, zo wordt deze van meer zekerheid voorzien. De nieuwe onderzoeksgegevens worden gebruikt om de likelihood ratio vast te stellen: hoe elke hypothese, met de bijbehorende parameterwaarde, zich verhoudt tot de gevonden data. Om een statistisch verantwoorde conclusie te trekken worden de producten van de prior en de likelihood berekend, deze zijn evenredig aan de posterior oftewel de Bayesfactor (Flohr, 2012).

De Bayesfactor is een maat voor de sterkte van het bewijs voor een hypothese ten opzichte van de concurrerende hypothesen. In tegenstelling tot klassieke statistiek kan Bayesiaanse statistiek niet alleen bewijs leveren voor de alternatieve hypothese, maar ook voor de nulhypothese, waardoor een completer beeld ontstaat van wat de data suggereren (Waterink et al., 2021; Wetzels et al., 2011). Het kwantificeert het bewijs van de data op een continue schaal en niet als een dichotome beslissing, zoals in de klassieke statistiek (Mulder & Wagenmaker, 2016; Pek & Van Zandt, 2019). De Bayesfactor varieert tussen 0 en  $\infty$ , waarbij 1 aangeeft dat de gegevens geen van beide theorieën meer ondersteunen dan de andere. Wanneer bijvoorbeeld  $H_0$  wordt vergeleken met  $H_A$ , dan wordt de Bayesfactor uitgedrukt in  $BF_{01}$ . Waarden groter dan 1 duiden op toenemend bewijs voor één theorie boven de andere (toenemend bewijs voor de nulhypothese boven een alternatieve hypothese), terwijl waarden kleiner dan 1 duiden op het omgekeerde (toenemend bewijs voor de alternatieve hypothese boven de nulhypothese). Zo maken Bayesfactoren drie verschillende soorten conclusies mogelijk: er is sterk bewijs voor de nulhypothese ( $BF_{01}$  veel groter dan 1); er is sterk bewijs voor de alternatieve hypothese ( $BF_{01}$  dicht bij 0); en het bewijs is

ongevoelig ( $BF_{01}$  dicht bij 1) (Dienes, 2014). Wanneer  $BF_{01} = 10$ , dan impliceert dit dat de data tien keer meer steun vinden bij de nulhypothese dan bij de alternatieve hypothese. Bij de Bayesiaanse statistiek wordt dus de alternatieve hypothese expliciet meegenomen in de statistische conclusie (Mulder & Wagenmaker, 2016; Pek & Van Zandt, 2019; Stevens & Meester, 2021). Dit contrasteert met de klassieke statistiek, waarin het niet verwerpen van de nulhypothese alleen aangeeft dat er onvoldoende steun is voor de alternatieve hypothese, maar de nulhypothese niet wordt gekwantificeerd.

Doordat de Bayesfactor een compleet beeld geeft van wat de data suggereren, zowel van de alternatieve hypothese als van de nulhypothese, kan Bayesiaanse analyse helpen bij de vraag of replicatie van een onderzoek nodig is voor de uitbreiding van wetenschappelijke kennis (Verhagen & Wagenmakers, 2014).

### **Huidige onderzoek**

In de sociale wetenschappen is het reproduceren van resultaten en het gebruik van de  $p$ -waarde voor de bewijsvoering van een onderzoek onderwerp van debat (Field et al., 2016; McShane et al., 2019; Ritchie, 2020). Daarbij wordt gediscussieerd of Bayesiaanse statistiek kan bijdragen aan de bewijsbeslissing van onderzoek (Ritchie, 2020; Stevens & Meester, 2021). Hoewel  $p$ -waarden en Bayesfactoren het eens kunnen zijn over welke hypothese het best wordt ondersteund door de data, kunnen ze verschillen in hun oordeel over de sterkte van dit bewijs. Zo stellen Wetzels et al. (2011) dat de Bayesiaanse benadering vergeleken met de klassieke benadering voorzichtiger is, waardoor overschatting van bewijs voor een effect wordt voorkomen.

Dit onderzoek draagt bij aan deze discussie door bij een pedagogisch onderzoek zowel Bayesiaanse als klassieke statistiek toe te passen en de bewijsvoering te vergelijken. Pedagogisch wetenschappelijk onderzoek richt zich op de ontwikkeling van een kind tot volwassene, waarbij de opvoeding, de ontwikkelingsfasen en de omgeving van het kind van invloed zijn (Marton & Booth, 1997). De betrouwbaarheid van pedagogisch onderzoek is belangrijk, omdat conclusies uit deze onderzoeken gebruikt worden door ouders, leraren en anderen voor de optimale ontwikkeling van kinderen (Van IJzendoorn & de Frankrijker, 2005). Daarnaast draagt dit onderzoek bij aan meer kennis over het gebruik van Bayesiaanse statistiek bij de argumentatie van replicatieonderzoek (Field et al., 2016).

Het doel van het huidige onderzoek is om meer inzicht te verwerven in de bewijsvoering van pedagogisch onderzoek. Hiervoor wordt de bewijsvoering van klassieke statistiek van een bestaand onderzoek van Palade en Pascal (2023) vergeleken met de situatie



waarin de onderzoeksvragen worden beantwoord met Bayesiaanse statistiek. Bij het huidige onderzoek wordt bij de Bayesiaanse statistiek als overtuigend bewijs beschouwd, wanneer de gegevens drie keer meer steun geven aan de alternatieve hypothese dan aan de nulhypothese, of omgekeerd — gebaseerd op de geobserveerde data. De onderzoeksvragen zijn:

- 1) Welke resultaten zijn gevonden door Palade en Pascal (2023) in hun pedagogische onderzoek, weergegeven in het artikel *Reducing bullying through empathy training: the effect of teacher's passive presence*?
  - a) Zijn in het huidige onderzoek met klassieke statistiek resultaten te vinden, die vergelijkbaar zijn met de resultaten van Palade en Pascal?
- 2) Welke resultaten worden gevonden bij het huidige onderzoek door herberekening met gebruikmaking van Bayesiaanse statistiek?
  - a) In hoeverre geeft de Bayesiaanse statistiek overtuigende Bayesfactoren voor de alternatieve hypothese ( $BF_{10} > 3.00$ ) voor de significante  $p$ -waarden, die worden gerapporteerd door Palade en Pascal?
  - b) In hoeverre geeft de Bayesiaanse statistiek overtuigende Bayesfactoren voor de nulhypothese ( $BF_{01} > 3.00$ ) voor de niet-significante  $p$ -waarden, die worden gerapporteerd door Palade en Pascal?
- 3) In hoeverre leidt de toepassing van klassieke statistiek versus Bayesiaanse statistiek tot verschillen in de conclusies van het onderzoek van Palade en Pascal over het verband tussen empathie en verbaal of fysiek pesten bij basisschoolleerlingen?
- 4) In hoeverre kan replicatie van het onderzoek van Palade en Pascal bijdragen aan de uitbreiding van de wetenschappelijke kennis van de relatie tussen empathie, verbaal en fysiek pesten bij basisschoolleerlingen?

### Originele onderzoek

De vraagstelling is onderzocht door middel van een reproductie-onderzoek. Hiervoor is gezocht naar toepassingen van klassieke statistiek in wetenschappelijke pedagogische onderzoeken in de gratis online database Open Science Framework (OSF, z.d.). De zoektermen waren 'education' en 'pedagogy'. Uit de gevonden artikelen is een onderzoek van Palade en Pascal (2023) geselecteerd, over de invloed van empathietraining op pestgedrag bij Roemeense scholieren. Dit onderzoek is geselecteerd vanwege de beschikbaarheid van de data in een openbare dataset, de heldere beschrijving van de onderzoeksopzet en de vermelding van zowel significante als niet-significante resultaten. In

Tabel 1 wordt een samenvatting van de inhoud van het onderzoek van Palade en Pascal weergegeven met de hypothesen, de methode en de analyse. Pascal en Palade vermelden geen assumptiecontroles, waardoor het onduidelijk is of deze zijn uitgevoerd.

## Tabel 1

*Samenvatting van de inhoud van het onderzoek van Palade & Pascal (2023), met de hypothesen, de methode en de analyse*

<b>Inhoud</b>	
<b>Hypothesen</b>	<ol style="list-style-type: none"> <li>1. De interventie verhoogt de empathie in de experimentele groepen en dit effect is het sterkst bij de groep waar de leraar aanwezig is.</li> <li>2. Pestgedrag neemt af bij de experimentgroepen en blijft constant bij de controlegroep. Dit effect is het sterkst bij de experimentgroep waar de leraar aanwezig is bij de interventie.</li> <li>3. Het niveau van empathie is een voorspeller van verbaal en fysiek pestgedrag.</li> </ol>
<b>Methode</b> <i>Steekproef</i>	64 Roemeense basisschoolleerlingen in een stedelijke omgeving, $N = 32$ jongens (50%) en $N = 32$ meisjes (50%), gemiddelde leeftijd van de leerlingen was 9.45 ( $SD = 0.50$ ) jaar, met een range van 9 jaar tot en met 10 jaar.
<b>Procedure</b>	<p>De leerlingen zijn willekeurig verdeeld over drie groepen:</p> <ol style="list-style-type: none"> <li>1. Controlegroep (<math>N = 22</math>): geen empathietraining</li> <li>2. Experimentgroep 1 (<math>N = 20</math>): empathietraining waarbij de leraar passief aanwezig was</li> <li>3. Experimentgroep 2 (<math>N = 22</math>): empathietraining zonder aanwezigheid van de leraar</li> </ol> <p>De interventie met een empathietraining duurde vijf aaneengesloten dagen, met sessies van 50 minuten.</p>
<b>Testen</b>	<p>Twee vragenlijsten:</p> <ol style="list-style-type: none"> <li>1. De Multidimensional Peer-Victimization Scale (Mynard &amp; Joseph, 2000) bestaande uit acht vragen is gebruikt om de mate van pesten vast te stellen. De vragen zijn verdeeld over verbaal pesten (vier vragen) en fysiek pesten (vier vragen).</li> <li>2. De Empathy Index for Children and Adolescents (Bryant, 1982) bestaande uit 22 vragen is gebruikt om de mate van empathie te bepalen.</li> </ol> <p>Er waren drie meetmomenten:</p> <ol style="list-style-type: none"> <li>1. Pre-test voor de interventie</li> <li>2. Post-test direct na de laatste interventie sessie</li> <li>3. Follow-up test 21 dagen na de interventie</li> </ol>

Analyse	Beschrijvende statistieken met gemiddelden en standaarddeviaties Correlaties met Pearson's correlatiecoëfficiënten Herhaalde metingen ANOVA's (om de interventie-effecten op de empathie, het fysiek en verbaal pesten van de leerlingen in de drie experimentgroepen op de drie testmomenten te analyseren) Meervoudige lineaire regressies (als controlevariabelen geslacht en leeftijd) om te onderzoeken of empathie een voorspeller is voor verbaal of fysiek pesten
---------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

---

## Dataset

De dataset van het onderzoek van Palade en Pascal bevat twaalf kolommen. Deze omvatten de experimentele conditie (experimentgroep 1: empathie-interventie en aanwezigheid van de leraar; experimentgroep 2: empathie-interventie met afwezigheid van de leraar; en controlegroep: interventie niet gerelateerd aan pesten), het geslacht en de leeftijd van de leerlingen. Verder zijn de totaalscores voor empathie, verbaal pesten en fysiek pesten per leerling en per meetmoment (pre-test, post-test en follow-up test) opgenomen. Alle leerlingen hebben alle vragen van de testen beantwoord.

## Data-analyse

De data van Palade en Pascal zijn opnieuw geanalyseerd met behulp van klassieke statistiek om de reproduceerbaarheid van de resultaten te verifiëren. De analyses zijn uitgevoerd volgens de methoden beschreven door Palade en Pascal, zie Tabel 1. In het huidige onderzoek is onderzocht of herberekeningen met Bayesiaanse statistiek kunnen bijdragen aan de bewijsvoering van het onderzoek van Palade en Pascal. Zowel de significante als de niet-significante resultaten van dat onderzoek zijn herberekend met Bayesiaanse statistiek met behulp van JASP (versie 0.18.1.0), een gratis statistisch softwareprogramma dat zowel klassieke als Bayesiaanse statistiek berekent (Wagenmakers et al., 2017). Om de hypothesen te toetsen is gebruikgemaakt van beschrijvende statistieken, correlaties, ANOVA's en regressieanalyses met zowel klassieke als Bayesiaanse statistiek. Voor alle klassieke testen gold een significantieniveau van 5%.

## Herhaalde metingen ANOVA's

Met herhaalde metingen ANOVA's (RM-ANOVA) is het effect van het meetmoment, van de experimentele conditie en van de interactie tussen het meetmoment met de experimentele conditie op de empathie, het fysiek pesten en het verbaal pesten van de leerlingen getoetst. RM-ANOVA's nemen de correlatie binnen en tussen experimentele

groepen in acht, samen met de meetmomenten (Muhammad, 2023). Deze berekeningen zijn gebruikt om te onderzoeken of er significante verschillen waren tussen de steekproefverwachtingen (Moore & McCabe, 2006).

Bij RM-ANOVA's worden verschillende assumpties gedaan. De afhankelijke variabele wordt op een continu niveau gemeten en de onafhankelijke variabelen zijn categorisch met drie afzonderlijke metingen bij dezelfde deelnemers (Muhammad, 2023). De assumptie dat de data in elke groep normaal is verdeeld, is onderzocht met histogrammen en met de Shapiro-Wilk test. Deze test heeft een meer onderscheidend vermogen dan andere normaaltesten en wordt vaak gebruikt bij kleinere steekproeven ( $N < 50$ ) (Sawyer, 2009). Bij het huidige onderzoek moet rekening gehouden worden met het lage onderscheidend vermogen van normaaltesten bij kleine steekproeven ( $N < 30$ ) (Razali & Wah, 2011). Een normaalverdeling heeft een *skewness* van  $\pm 1.41$  en *kurtosis* van  $\pm 1.10$  bij  $N = 20$  (Sawyer, 2009). De aanname van sphericiteit impliceert dat de variantie van de verschillen tussen groepen metingen ongeveer gelijk aan elkaar is, deze assumptie is onderzocht met Mauchly's W (Van den Bergh et al., 2020). De homogeniteit van varianties, waarbij de residuele variantie hetzelfde is over alle niveaus van de voorspellers, is getest met de Levene's test (Van den Bergh et al., 2020).

De effectgrootte is voor de totale groep aangegeven met Partial Eta squared ( $\eta^2_p$ ) en voor de experimentgroepen met Omega squared ( $\omega^2$ ), deze laatste geeft bij kleine steekproeven ( $N < 30$ ) een onbevooroordeelde maat voor de effectgrootte. Bij deze analyse gold  $\eta^2_p$  of  $\omega^2 > .14$  als een groot effect,  $> .06$  als een matig effect en  $> .01$  als een klein effect (Goss-Sampson, 2020).

### Meervoudige regressies

Om te onderzoeken of empathie een voorspeller is voor verbaal of fysiek pesten, zijn per experimentgroep meervoudige lineaire regressies met vier modellen uitgevoerd. Model 1a bevat de post-test van verbaal pesten als afhankelijke variabele en geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie en post-test empathie als onafhankelijke variabelen. Model 1b bevat de follow-up test van verbaal pesten als afhankelijke variabele en geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie, post-test empathie en follow-up test empathie als onafhankelijke variabelen. Model 2a en Model 2b zijn identiek aan Model 1a en 1b, alleen zijn de testen van verbaal pesten vervangen door testen van fysiek pesten.

Voordat deze analyses konden worden geïnterpreteerd, zijn per model de assumpties gecontroleerd van lineariteit, homoscedastiteit, normaliteit van de residuen en

onafhankelijkheid van de residuen. Lineariteit en homoscedasticiteit zijn onderzocht met behulp van spreidingsdiagrammen van de gestandaardiseerde residuen tegen de gestandaardiseerde voorspelde waarden. De normaliteit van de residuen is gecontroleerd met PP-plotten (Siero et al., 2009). Multicollineariteit is geëvalueerd aan de hand van de Variation Influence Factor (VIF), waarbij een waarde onder de vijf duidt op weinig onderlinge correlatie (Akinwande et al., 2015). Met een Durbin-Watson test is de onafhankelijkheid van de residuen onderzocht, hierbij geeft een waarde kleiner dan 1 of groter dan 3 aan dat niet aan deze assumptie is voldaan (Palupi & Rizki, 2020). Tot slot is gezocht naar uitbijters met Cook's distance, waarbij de vuistregel  $4/N$  is gehanteerd, maar ook door te onderzoeken of er punten zijn die meer dan drie standaarddeviaties vanaf het gemiddelde liggen (Altman & Krzywinski, 2016).

De effectgrootte wordt aangegeven met de  $R^2$ , waarbij  $R^2 > 0.02$  als klein,  $R^2 > 0.15$  als middelgroot en  $R^2 > 0.35$  als groot effect zijn gedefinieerd (Cohen, 1992). De samenhang van de variabelen zijn geanalyseerd door middel van een correlatieanalyse (Pearson's  $r$ ) tussen de variabelen van verbaal pesten, fysiek pesten en empathie. In deze analyse werd een  $r > .70$  beschouwd als een sterke correlatie, een  $r$  tussen  $.50$  en  $.70$  als een matige correlatie, en een  $r < .50$  als een zwakke correlatie (Hinkle et al., 2003).

### **Bayesiaanse herhaalde metingen ANOVA's en meervoudige regressies**

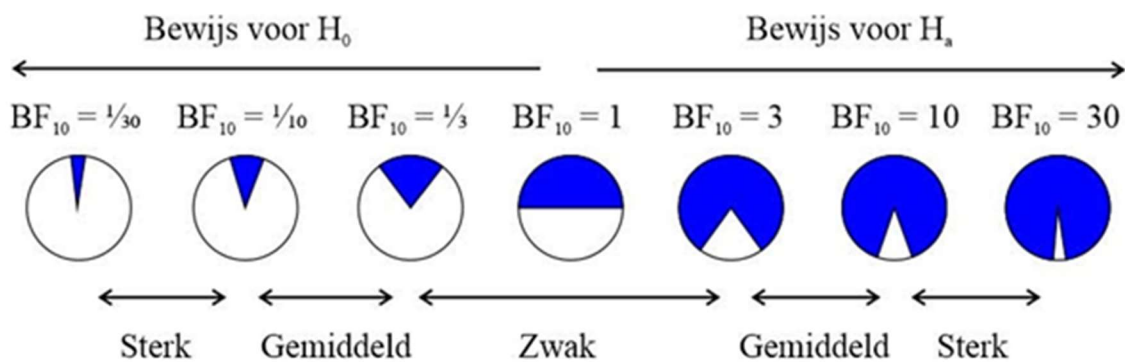
De RM-ANOVA's en meervoudige lineaire regressies zijn ook met Bayesiaanse statistiek berekend, waarbij de priors zijn gekozen volgens de standaardinstellingen in JASP. Deze standaardpriors zijn breed inzetbaar, efficiënt en resulteren in Bayesfactoren met gunstige theoretische eigenschappen (Rouder et al, 2012). De standaardprior in JASP voor regressies is de Jeffreys-Zellner-Siow prior met  $r = 0.354$  (Heo & Van de Schoot, 2020). Voor de RM-ANOVA's gebruikt JASP een standaardprior, waarbij de a priori kans van elk model gelijk is aan één gedeeld door het totale aantal modellen (Van den Bergh, 2023). Door te testen met bredere en smallere priors is de robuustheid van de analyses onderzocht (Van Doorn, 2020). Robuustheid kan gedefinieerd worden als de capaciteit van de vereiste kansberekeningen om resultaten te produceren die ongevoelig zijn voor afwijkingen van ideale assumpties die in de praktijk voorkomen, zoals normaliteit of uitbijters of het gebruik van verschillende priors (Box, 1979; Moore & McCabe, 2006).

Voor zowel de RM-ANOVA's als de lineaire regressies zijn inclusie en exclusie Bayesfactoren gebruikt. Met inclusie Bayesfactoren (optie  $BF_{10}$  in JASP) kunnen alle modellen worden vergeleken die een variabele bevatten (zoals bijvoorbeeld het effect van de

conditie) met modellen die dat niet doen. Op basis van deze gegevens kan de relatieve sterkte van het bewijs voor elke factor ten opzichte van de afhankelijke variabele worden bepaald. De exclusie Bayesfactor (optie  $BF_{01}$  in JASP) doet het tegenovergestelde: die vertelt hoeveel waarschijnlijker de data zijn, gegeven dat een bepaalde variabele niet in het model is opgenomen (Linde & Van Ravenzwaaij, 2021). Figuur 1 geeft de Bayesfactor en de classificering van de ondersteuning voor de hypothesen weer.

### Figuur 1

*De classificering van de Bayesfactor als ondersteuning voor de nulhypothese en de alternatieve hypothese*



Opmerking. Aangepast overgenomen van Van Doorn et al. (2020, p. 821).

Na het vaststellen van welke voorspellers relevant zijn, zijn de plots van de gemiddelde posteriorverdelingen bekeken om de omvang van de relaties te onderzoeken. Hierbij is een betrouwbaarheidsinterval van 95% gehanteerd. Vervolgens is geanalyseerd welke Bayesfactor overeenkomt met de significante en niet-significante resultaten van de klassieke statistiek en wat dit betekent voor de bewijsvoering en de conclusies van het onderzoek van Palade en Pascal.

### Replicatie

Op basis van de klassieke en Bayesiaanse berekeningen is tot slot geanalyseerd of het onderzoek van Palade en Pascal mogelijk geschikt is voor replicatie, waarbij is gebruikgemaakt van de volgende verdeling in drie klassen (Field et al., 2019):

1. Resultaten waarvan het statistische bewijs overtuigend ondersteuning geeft aan de alternatieve hypothese (geen replicatie nodig).
2. Resultaten waarvan het statistische bewijs overtuigend ondersteuning geeft aan de nulhypothese (geen replicatie nodig).

3. Resultaten waarvan het statistische bewijs geen duidelijkheid geeft voor de ondersteuning van een hypothese (replicatie kan nodig zijn afhankelijk van theoretische en methodologische overwegingen). Hierbij ligt  $BF_{10}$  tussen 0.33 en 3. Als een onderzoek in klasse 3 valt, kan overwogen worden om het te repliceren. Hierbij worden drie selectiecriteria in aanmerking genomen: het theoretisch belang, de relevantie en de volledigheid van het onderzoek (Mackey, 2012).

## Resultaten

Niet alle bevindingen van Palade en Pascal konden worden gereproduceerd door de onderzoeker en de scriptiebegeleider, e-mails naar de auteurs om mogelijke misverstanden te controleren werden niet beantwoord. In Bijlage A worden de beschrijvende statistieken en correlaties van het huidige onderzoek besproken, deze resultaten komen overeen met Palade en Pascal. Bij leerlingen die een interventie kregen (met en zonder leraar), nam het gemiddelde van empathie toe bij de drie meetmomenten, bij de controlegroep nam dit af. Bij de correlaties komen de significante  $p$ -waarden met een  $\alpha < .05$  niet consistent overeen met overtuigende Bayesfactoren, met een  $\alpha < .01$  is dit wel consistent. Bij de resultaten van de herhaalde metingen ANOVA's en de meervoudige regressieanalyses werden verschillen met Palade en Pascal gevonden.

### Effecten van de interventie op empathie, verbaal en fysiek pesten

Tabel 2 geeft een samenvatting van de resultaten van zowel klassieke als Bayesiaanse RM-ANOVA's, deze resultaten omvatten de hoofdeffecten en interactie-effecten per groep voor empathie, verbaal pesten en fysiek pesten, zoals gevonden in het onderzoek van Palade en Pascal en in het huidige onderzoek. De significante resultaten waren niet reproduceerbaar. Palade en Pascal geven ongespecificeerde niet-significante uitkomsten, waardoor vergelijking van hun resultaten met die van het huidige onderzoek niet mogelijk is. Het vergelijken van exacte testwaarden zou nuttig zijn geweest, gezien de inconsistentie in de reproductie van hun bevindingen.

**Tabel 2**

*Samenvatting van de resultaten van klassieke RM-ANOVA's van het onderzoek van Palade en Pascal (2023) en huidig onderzoek en van Bayesiaanse RM-ANOVA's per hoofdeffect en interactie-effect en per experimentgroep voor empathie, verbaal en fysiek pesten*

	<b>RM-ANOVA, Palade &amp; Pascal</b>	<b>RM-ANOVA Huidige onderzoek</b>	<b>Bayesiaanse RM- ANOVA</b>
<b>Empathie</b>			
Hoofdeffect van meetmoment (binnen-subjecten variabele)	$F_{(1,61)} = 9.86,$ $p = 0.003, \eta^2_p =$ 0.139	$F_{(2,122)} = 5.24,$ $p = 0.01, \eta^2_p = 0.08$	$BF_{incl} = 3.68$
Hoofdeffect van de conditie (tussen-subjecten variabele)	$p = n.s.$	$F_{(2,61)} = 2.96,$ $p = 0.06, \eta^2_p = 0.09$	$BF_{excl} = 1.41$
Interactie tussen meetmoment en conditie	$F_{(2,61)} = 3.76,$ $p = 0.02, \eta^2_p = 0.11$	$F_{(4,122)} = 2.16,$ $p = 0.08, \eta^2_p = 0.07$	$BF_{incl} = 1.02$
Experimentgroep 1	$F_{(2,38)} = 6.76,$ $p = 0.003, \eta^2_p =$ 0.26	$F_{(2,38)} = 6.76,$ $p = 0.003, \omega^2 = 0.14$	$BF_{incl} = 25.84$
Experimentgroep 2	$p = n.s.$	$F_{(2,42)} = 2.89,$ $p = 0.07, \omega^2 = 0.06$	$BF_{incl} = 1.64$
Controlegroep	$p = n.s.$	$F_{(2,42)} = 0.27,$ $p = 0.76, \omega^2 = 0.00$	$BF_{excl} = 6.43$
<b>Verbaal pesten</b>			
Hoofdeffect van meetmoment (binnen-subjecten variabele)	$F_{(1,61)} = 29.70,$ $p < 0.001, \eta^2_p = 0.33$	$F_{(2,122)} = 12.10,$ $p < 0.001, \eta^2_p =$ 0.09	$BF_{incl} = 123240.17$
Hoofdeffect van de conditie (tussen-subjecten variabele)	$F_{(2,61)} = 7.59,$ $p = 0.001, \eta^2_p =$ 0.20	$F_{(2,61)} = 7.96,$ $p = 0.001, \eta^2_p =$ 0.08	$BF_{incl} = 1923.19$
Interactie tussen meetmoment en conditie	$F_{(2,61)} = 13.30,$ $p < 0.001, \eta^2_p =$ 0.30	$F_{(4,122)} = 5.96,$ $p < 0.001, \eta^2_p =$ 0.09	$BF_{incl} = 913.00$
Experimentgroep 1	$F_{(2,38)} = 16.79,$ $p < 0.001, \eta^2_p =$ 0.46	$F_{(2,38)} = 16.79,$ $p < 0.001, \omega^2 = 0.31$	$BF_{incl} = 16671.27$
Experimentgroep 2	$p = n.s.$	$F_{(2,42)} = 2.16,$ $p = 0.13, \omega^2 = 0.03$	$BF_{excl} = 1.41$
Controlegroep	$p = n.s.$	$F_{(2,42)} = 0.08,$ $p = 0.93, \omega^2 = 0.00$	$BF_{excl} = 7.36$
<b>Fysiek pesten</b>			
Hoofdeffect van meetmoment (binnen-subjecten variabele)	$p = n.s.$	$F_{(2,122)} = 1.18,$ $p = 0.31, \eta^2_p = 0.02$	$BF_{excl} = 8.68$
Hoofdeffect van de conditie (tussen-subjecten variabele)	$p = n.s.$	$F_{(2,61)} = 1.11,$ $p = 0.34, \eta^2_p = 0.04$	$BF_{excl} = 9.98$
Interactie tussen meetmoment en conditie	$p = n.s.$	$F_{(4,122)} = 0.19,$ $p = 0.94, \eta^2_p = 0.01$	$BF_{excl} = 294.17$
Experimentgroep 1	$p = n.s.$	$F_{(2,38)} = 0.59,$ $p = 0.56, \omega^2 = 0.00$	$BF_{excl} = 4.19$
Experimentgroep 2	$p = n.s.$	$F_{(2,42)} = 0.95,$ $p = 0.40, \omega^2 = 0.00$	$BF_{excl} = 4.01$
Controlegroep	$p = n.s.$	$F_{(2,42)} = 0.03,$ $p = 0.97, \omega^2 = 0.00$	$BF_{excl} = 7.91$



In Bijlage B worden de assumptiecontroles en de resultaten van de klassieke en Bayesiaanse ANOVA's weergegeven. Niet alle testen voldeden aan de normaalverdeling, zoals de pre-test voor empathie van experimentgroep 2 ( $W = 0.89, p = 0.02$ ), zie Tabel B2. Aan de assumptie van homogeniteit van varianties werd voldaan, behalve door de post-test van verbaal pesten ( $F_{(2,61)} = 8.41, p < 0.001$ ), zie Tabel B5. Alle testen voldeden echter aan de assumptie van sphericiteit, zie Tabel B6. Hoewel de F-toets van ANOVA robuust is tegen matige afwijkingen van normaliteit, kunnen lichte tot sterke afwijkingen leiden tot onjuiste  $p$ -waarden en mogelijk verkeerde conclusies (Moore & McCabe, 2006).

Ook zijn robuustheidcontroles uitgevoerd bij de Bayesiaanse modellen met bredere ( $r = 1$  voor vaste effecten) en smallere priors ( $r = 0.2$  voor vaste effecten) (Van Doorn et al., 2020). Deze controles tonen aan dat tien modellen niet consistent waren met de oorspronkelijke analyse.

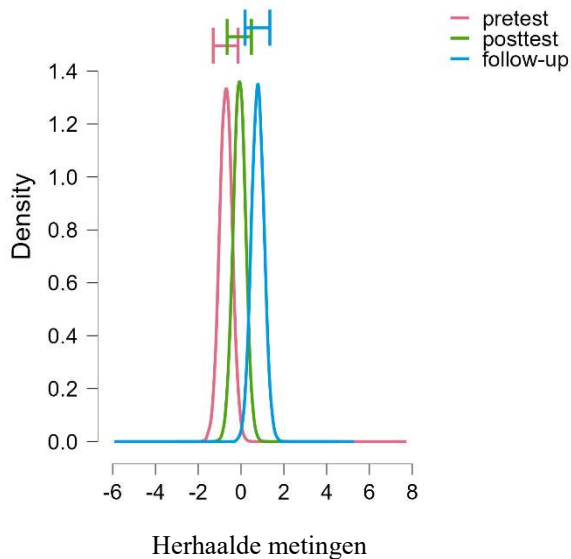
## Empathie

De klassieke RM-ANOVA toont dat het meetmoment een effect had op empathie met een medium effectgrootte ( $F_{(2,122)} = 5.24, p = 0.01, \eta^2_p = 0.08$ ), zie Tabel 2. Er is geen duidelijk bewijs voor een effect bij de experimentele conditie ( $F_{(2,61)} = 2.96, p = 0.06, \eta^2_p = 0.09$ ) of de interactie ( $F_{(4,122)} = 2.16, p = 0.08, \eta^2_p = 0.07$ ). Bij de leerlingen van experimentgroep 1 nam de empathie significant toe door de interventie ( $F_{(2,38)} = 6.76, p = 0.003, \omega^2 = 0.14$ ). Voor de andere twee experimentgroepen was er geen significant effect: experimentgroep 2 ( $F_{(2,42)} = 2.89, p = 0.07, \omega^2 = 0.06$ ) en de controlegroep ( $F_{(2,42)} = 0.27, p = 0.76, \omega^2 = 0.00$ ). Pascal en Palade rapporteren echter wel significantie bij het meetmoment, de interactie en experimentgroep 1.

De Bayesiaanse RM-ANOVA toont overtuigend bewijs voor een effect van het meetmoment ( $BF_{incl} = 3.68$ ). Figuur 2 laat de gemiddelde posteriorverdeling van empathie per meetmoment zien van de totale groep leerlingen. Bij de follow-up test ( $M = 0.79, SD = 0.29$ ) is een hogere empathiescore gemeten dan bij de andere meetmomenten. Er is sterk bewijs dat de meetmomenten een positief effect hebben op de empathie van experimentgroep 1 ( $BF_{incl} = 25.84$ ), de posteriorgemiddelden lopen op van -1.50 tot 1.48, zie Tabel B15. Voor de controlegroep is er overtuigend bewijs voor het ontbreken van een effect bij de meetmomenten ( $BF_{excl} = 6.43$ ). De overige resultaten bieden onvoldoende informatie om sterke conclusies te trekken over de effecten.

## Figuur 2

Plot van de gemiddelde posteriorverdeling van empathie van de totale groep per meetmoment met een 95% betrouwbaarheidsinterval



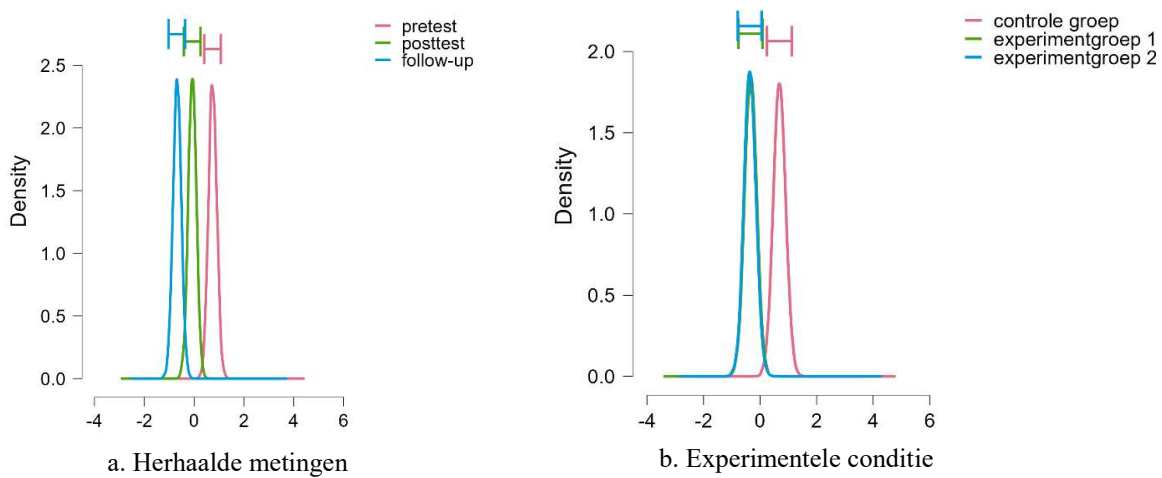
## Verbaal pesten

De klassieke RM-ANOVA toont een significant interventie-effect met een medium effectgrootte bij het meetmoment ( $F_{(2,122)} = 12.10, p < 0.00, \eta^2_p = 0.09$ ), de conditie ( $F_{(2,61)} = 7.96, p = 0.00, \eta^2_p = 0.08$ ) en de interactie ( $F_{(4,122)} = 5.96, p < 0.00, \eta^2_p = 0.09$ ), zie Tabel 2. Bij experimentgroep 1 had de interventie een significant effect met een grote effectgrootte op verbaal pesten ( $F_{(2,38)} = 16.79, p < 0.00, \omega^2 = 0.31$ ), bij de andere twee experimentgroepen niet: experimentgroep 1 ( $F_{(2,42)} = 2.16, p < 0.13, \omega^2 = 0.03$ ) en de controlegroep ( $F_{(2,42)} = 0.08, p < 0.93, \omega^2 = 0.00$ ).

De Bayesiaanse RM-ANOVA's tonen beslissend bewijs voor een effect bij het meetmoment ( $BF_{incl} = 123240.17$ ), de conditie ( $BF_{incl} = 1923.19$ ), de interactie tussen meetmoment en conditie ( $BF_{incl} = 913.00$ ) en experimentgroep 1 ( $BF_{incl} = 16671.27$ ). Het verbaal pesten van experimentgroep 1 nam af, de posteriorgemiddelden verminderen van 1.71 tot -1.57, zie Tabel B16. Figuur 3 geeft de gemiddelde posteriorverdeling van verbaal pesten per meetmoment en per conditie weer: de hoogste scores op verbaal pesten werden gehaald bij de pre-test en bij de controlegroep. Voor de controlegroep is gemiddeld bewijs gevonden voor de nulhypothese ( $BF_{excl} = 7.36$ ). Voor experimentgroep 2 geven de data onvoldoende informatie om een sterke conclusie te kunnen trekken over een effect.

**Figuur 3**

*Plots van de gemiddelde posteriorverdeling van verbaal pesten van de totale groep per meetmoment (a) en per experimentele conditie (b) met een 95% betrouwbaarheidsinterval*



### Fysiek pesten

Met de klassieke RM-ANOVA's is geen significantie gevonden bij de testen van fysiek pesten. De Bayesiaanse RM-ANOVA's geven bij alle testen van fysiek pesten gemiddeld tot beslissend bewijs voor het ontbreken van een interventie-effect, zoals bij het meetmoment ( $BF_{excl} = 8.68$ ) en bij de interactie ( $BF_{excl} = 294.17$ ), zie Tabel 2.

### Empathie als voorspeller van verbaal of fysiek pesten

Tabel 3 geeft een samenvatting van de resultaten van klassieke en Bayesiaanse meervoudige lineaire regressies, deze resultaten tonen de rol van empathie als voorspeller van verbaal en fysiek pesten per experimentgroep, zoals gevonden in het onderzoek van Palade en Pascal en in het huidige onderzoek. In Bijlage C worden de assumptiecontroles en de resultaten van de klassieke en Bayesiaanse regressies weergegeven. Er werden lichte afwijkingen gevonden bij de assumpties van lineariteit, homoscedasticiteit en normaalverdeling van de residuen, wat kan leiden tot onjuiste  $p$ -waarden of Bayesfactoren. Er was geen sprake van multicollineariteit, wel werden er uitbijters gevonden. De lineaire regressies werden berekend met en zonder uitbijters (volgens Cook's distance) om de consistentie van de resultaten te controleren. Het uitsluiten van uitbijters veranderde de pasvorm en inferentie van de meeste modellen niet significant, behalve bij twee modellen, zie Tabel C14 en C24 (Altman & Krzywinski, 2016). Deze leerlingen zijn niet uit de dataset verwijderd, omdat dit de steekproef te veel zou verkleinen en de representativiteit zou

beïnvloeden. Er werden geen uitbijters gevonden met drie standaardafwijkingen van het gemiddelde (Moore & McCabe, 2006). De robuustheid van de Bayesiaanse modellen werd getest door de standaardprior te halveren en te verdubbelen: acht modellen vertoonden geen consistente bewijslast.

**Tabel 3**

*Samenvatting van resultaten van klassieke meervoudige lineaire regressies van het onderzoek van Palade en Pascal (2023) en het huidig onderzoek en van Bayesiaanse meervoudige lineaire regressies met empathie als voorspeller van verbaal en fysiek pesten per experimentgroep*

	Klassieke regressie Palade & Pascal	Klassieke regressie Huidige onderzoek	Bayesiaanse regressie
<b>Verbaal pesten</b>			
Model 1a*	$p = n.s$	$R^2 = 0.07,$	$BF_{incl} = 0.47$
Experimentgroep 1		$p = 0.16$	
Model 1b**	$\Delta R^2 = 0.28^{***},$	$R^2 = 0.28,$	$BF_{incl} = 0.61$
Experimentgroep 1	$p = 0.02$	$p = 0.06$	
Model 1a*	$p = n.s$	$R^2 = 0.09,$	$BF_{incl} = 0.51$
Experimentgroep 2		$p = 0.45$	
Model 1b**	$p = n.s$	$R^2 = 0.32,$	$BF_{incl} = 1.10$
Experimentgroep 2		$p = 0.23$	
Model 1a*	$p = n.s.$	$R^2 = 0.12,$	$BF_{incl} = 2.07$
Controlegroep		$p = 0.28$	
Model 1b**	$\Delta R^2 = 0.28^{***},$	$R^2 = 0.29,$	$BF_{incl} = 5.16$
Controlegroep	$p = 0.01$	$p = 0.29$	
<b>Fysiek pesten</b>			
Model 2a*	Post-test: $p = n.s$	$R^2 = 0.14,$	$BF_{incl} = 0.48$
Experimentgroep 1		$p = 0.65$	
Model 2b**	Follow up test:	$R^2 = -0.01,$	$BF_{incl} = 0.41$
Experimentgroep 1	$p = n.s$	$p = 0.12$	
Model 2a*	Post-test: $p = n.s$	$R^2 = 0.09,$	$BF_{incl} = 0.48$
Experimentgroep 2		$p = 0.16$	
Model 2b**	Follow up test:	$R^2 = 0.32,$	$BF_{incl} = 1.20$
Experimentgroep 2	$p = n.s$	$p = 0.44$	
Model 2a*	Post-test: $p = n.s$	$R^2 = 0.22,$	$BF_{incl} = 1.62$
Controlegroep		$p = 0.29$	
Model 2b**	Follow up test:	$R^2 = 0.45,$	$BF_{incl} = 0.99$
Controlegroep	$p = n.s$	$p = 0.87$	

\* Model voor regressieanalyse met als afhankelijke variabele de post-test van verbaal/fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal/fysiek pesten, pre-test empathie en post-test empathie

\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van verbaal/fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal/fysiek pesten, pre-test empathie, post-test empathie en follow-up test empathie

\*\*\* Palade en Pascal rapporteren in hun artikel  $\Delta R^2$ , het huidige onderzoek gaat uit van de adjusted  $R$ .

In het huidige onderzoek werden geen significante resultaten gevonden voor empathie als voorspeller van verbaal en fysiek pesten. Bij verbaal pesten vinden Palade en Pascal wel significant bewijs bij de follow-up test van experimentgroep 1 ( $\Delta R^2 = 0.28, p = 0.02$ ) en bij de follow-up test van de controlegroep ( $\Delta R^2 = 0.28, p = 0.01$ ). Bij alle niet-significante resultaten van de lineaire regressies in het huidige onderzoek rapporteren Palade en Pascal ook niet-significante resultaten. De Bayesiaanse regressies tonen zwak bewijs voor de nul- of alternatieve hypothese, alleen bij de controlegroep werd gemiddeld bewijs gevonden dat empathie een voorspeller is van verbaal pesten, bij de follow-up test ( $BF_{incl} = 5.16$ ).

### **Overeenkomsten en verschillen onderzoek Palade en Pascal met huidige onderzoek**

Bijlage D geeft overzichten van de significante en niet-significante resultaten van het onderzoek van Palade en Pascal en het huidige onderzoek, en van de overtuigende en niet-overtuigende Bayesfactoren. Dit overzicht verschaft inzicht in de overeenkomsten en verschillen in de bewijslast tussen de verschillende onderzoeken.

Uit de vergelijking van het aantal significante en niet-significante resultaten van de RM-ANOVA's en regressies tussen het onderzoek van Palade en Pascal en het huidige onderzoek blijkt dat 10% van de testen niet overeenkomt, zie Tabel D1. Voor de RM-ANOVA's werd een klein verschil in significante resultaten gevonden: 23% bij Palade en Pascal en 20% in het huidige onderzoek. Bij de lineaire regressieresultaten was het verschil groter: 7% bij Palade en Pascal tegenover 0% in het huidige onderzoek.

### **Overtuigende Bayesfactoren**

Uit de vergelijking van de overtuigende Bayesfactoren voor de alternatieve hypothese met de significante resultaten van Palade en Pascal en het huidige onderzoek, komt een klein verschil naar voren bij de ANOVA's: zes overtuigende Bayesfactoren, zeven significante resultaten van Palade en Pascal het huidige onderzoek zes, zie Tabel D2. Bij de regressies is één overtuigende Bayesfactor voor de alternatieve hypothese gevonden, bij Palade en Pascal zijn twee significante resultaten gevonden, bij het huidige onderzoek nul.

Wanneer het aantal overtuigende Bayesfactoren voor de nulhypothese worden vergeleken met de niet-significante resultaten, dan zijn de verschillen bij de ANOVA's: acht overtuigende Bayesfactoren, elf niet-significante resultaten bij Palade en Pascal en twaalf bij het huidige onderzoek, zie Tabel D3. Bij de regressies zijn geen overtuigende Bayesfactoren voor de nulhypothese gevonden, bij Palade en Pascal zijn tien niet-significante resultaten gevonden, bij het huidige onderzoek twaalf.

Het huidige onderzoek bestaat uit dertig testmodellen: achttien RM-ANOVA modellen en twaalf regressiemodellen. Bij 14% van de RM-ANOVA's zijn niet-overtuigende Bayesfactor gevonden (7% voor de alternatieve hypothese, 7% voor de nulhypothese), bij de regressies 36% (3% voor de alternatieve hypothese, 33% voor de nulhypothese), zie Tabel D4. Voor deze testen gaven de data onvoldoende informatie om een sterke conclusie te kunnen trekken over een effect.

### Verschillen en overeenkomsten bij de conclusies

In Tabel 4 zijn de overeenkomsten en verschillen in de conclusies tussen Palade en Pascal en het huidige onderzoek weergegeven. Palade en Pascal hebben klassieke statistiek toegepast en het huidige onderzoek zowel klassieke als Bayesiaanse statistiek.

**Tabel 4**

*Overeenkomsten en verschillen in de conclusies van het onderzoek van Palade en Pascal (2023) en het huidige onderzoek*

<b>Palade en Pascal</b>	<b>Huidige onderzoek</b>
Wanneer de leraar aanwezig is bij de interventie dan neemt de empathie van de leerlingen toe en het verbaal pesten af. In de andere experimentele groepen werd geen significant effect van de interventie waargenomen.	Klassieke en Bayesiaanse conclusie: wanneer de leraar aanwezig is bij de interventie dan neemt de empathie van de leerlingen toe en het verbaal pesten af. In de andere experimentele groepen werd geen significant effect van de interventie waargenomen.
Bij fysiek pesten is geen significant effect van de interventie waargenomen.	Klassieke conclusie: Bij fysiek pesten is geen significant effect van de interventie waargenomen. Bayesiaanse conclusie: er is bewijs gevonden dat de interventie geen effect had op het fysiek pesten door leerlingen.
Er is geen eenduidig bewijs voor empathie als voorspeller van verbaal of fysiek pesten. Empathie voorspelt verbaal pesten niet, behalve op twee meetmomenten en alleen voor experimentegroep 1 en de controlegroep. Voor empathie als voorspeller van fysiek pesten is geen bewijs gevonden.	Klassieke conclusie: er is geen significant bewijs gevonden dat empathie een voorspeller is van verbaal of fysiek pesten. Bayesiaanse conclusie: bij de controlegroep is empathie een voorspeller van verbaal pesten op de langere termijn (follow-up test). Bij de andere testen geven de data onvoldoende informatie om sterke conclusies te kunnen trekken.

## Conclusie en discussie

In dit onderzoek is onderzocht of er verschil is in de bewijsvoering voor een hypothese, wanneer een bestaand pedagogisch onderzoek wordt gereproduceerd met zowel klassieke als Bayesiaanse statistiek. Hiervoor is het onderzoek *Reducing bullying through empathy training: the effect of teacher's passive presence?* van Palade en Pascal (2023) geselecteerd. Zij hebben onderzocht of een empathietraining met en zonder aanwezigheid van de leraar invloed heeft op verbaal of fysiek pesten van basisschoolleerlingen, op de korte en langere termijn. Hiervoor hebben zij herhaalde metingen ANOVA's (RM-ANOVA's) en meervoudige regressies met klassieke statistiek uitgevoerd.

Eerst is onderzocht of de resultaten van Palade en Pascal konden worden gerepliceerd, hieruit bleek dat de beschrijvende statistieken en correlaties vergelijkbaar waren. Dit gold echter niet voor de significante resultaten van de RM-ANOVA's en regressies, die geen enkele overeenkomst vertoonden. Bovendien werd bij een tiende van de testen geen consensus bereikt over de significantie van de resultaten.

Vervolgens zijn de RM-ANOVA's en regressies opnieuw uitgevoerd met Bayesiaanse statistiek om de ondersteuning voor de hypothesen te vergelijken tussen de klassieke en Bayesiaanse benaderingen. Er is onderzocht of de Bayesiaanse statistiek overtuigend bewijs levert voor een interventie-effect ( $BF_{10} > 3$ ) bij significante  $p$ -waarden en voor het ontbreken van een interventie-effect ( $BF_{01} > 3$ ) bij niet-significante  $p$ -waarden, zoals gerapporteerd door Palade en Pascal en het huidige onderzoek. Het overtuigende bewijs van de Bayesiaanse statistiek kwam grotendeels overeen met de significante bevindingen van de klassieke statistiek van Palade en Pascal en het huidige onderzoek. Daarnaast ondersteunen overtuigende Bayesfactoren de nulhypothese — geen interventie-effect — voor fysiek pesten, terwijl de klassieke statistiek alleen aangeeft dat de nulhypothese niet kan worden verworpen. Dit bevestigt de conclusies van Wetzels et al. (2011), waarin wordt vastgesteld dat  $p$ -waarden en Bayesfactoren bijna altijd overeenkomen in welke hypothese beter wordt ondersteund, maar verschillen in de mate van ondersteuning.

Hierna is onderzocht in hoeverre de conclusies van het huidige onderzoek, met klassieke en Bayesiaanse statistiek, verschillen met de conclusies van Palade en Pascal die ook klassieke statistiek hebben toegepast. Er zijn geen verschillen gevonden in de conclusies van een interventie-effect tussen en binnen de experimentgroepen voor empathie, verbaal en fysiek pesten. Het huidige onderzoek voegt toe dat, op basis van overtuigende Bayesfactoren, geconcludeerd kan worden dat er bij fysiek pesten geen interventie-effect is. Er zijn geen

verschillen gevonden tussen de onderzoeken bij de conclusies over empathie als voorspeller van fysiek pesten, wel als voorspeller van verbaal pesten.

Tot slot kan worden afgewogen of replicatie van het onderzoek van Palade en Pascal de wetenschappelijke kennis uitbreidt van de relatie tussen empathie, verbaal en fysiek pesten. Aangezien de Bayesfactoren bij de regressies voornamelijk niet-overtuigend waren — de data bieden onvoldoende informatie om sterke conclusies te trekken over empathie als voorspeller van verbaal of fysiek pesten — kan replicatie overwogen worden, waarbij de steekproefomvang wordt vergroot om het onderscheidingsvermogen van de testen te verbeteren en het effect nader te onderzoeken (Moore & McCabe, 2006; Button et al., 2013).

Voor replicatie is de relevantie van het onderzoeksdoel belangrijk. De regressies zijn uitgevoerd om de derde hypothese van Palade en Pascal te testen: ‘Het niveau van empathie is een voorspeller van verbaal en fysiek pestgedrag’. Dit is een relevant onderwerp, omdat pestgedrag op de basisschool een wereldwijd probleem is. Het trainen van de empathie van leerlingen om pestgedrag te verminderen wordt door meerdere artikelen ondersteund, deze tonen een positieve relatie aan (Stavrinos et al., 2010). Het verder onderzoeken van de rol van empathie is praktisch relevant voor basisscholen. Meer kennis over de relatie tussen empathie en pestgedrag van leerlingen kan leiden tot het tijdig anticiperen op het empathisch vermogen. Daarnaast kunnen interventies gericht op de relaties tussen kinderen en hun klasgenoten mogelijk worden aangepast en verbeterd. Ook is het van maatschappelijk belang om onderzoek te doen naar de rol van empathie bij pestgedrag, omdat problemen met klasgenoten kinderen veel stress geven (Núñez-Regueiro & Núñez-Regueiro, 2021).

Het huidige onderzoek suggereert dat de aanwezigheid van de leraar empathie bevordert en verbaal pesten vermindert. Aangezien de leraar een cruciale rol speelt in de sociaal-emotionele ontwikkeling van kinderen, is het interessant om te onderzoeken of de onderzoeksresultaten verschillen bij de aanwezigheid van een passieve of actieve leraar (Brown et al., 2010). Daarnaast is het geslacht van de leerlingen interessant, omdat empathietraining volgens Kerivel (2016) helpt verschillen tussen jongens en meisjes te verkleinen, terwijl hun pestgedrag zich anders manifesteert.

### **Sterktes en beperkingen van het onderzoek**

Sterk aan dit onderzoek is dat er nog weinig pedagogisch onderzoek is gedaan met Bayesiaanse statistiek, dit onderzoek draagt daaraan bij. Het gebruik van Bayesiaanse statistiek zorgde ervoor, dat er niet alleen uitspraken konden worden gedaan over de ondersteuning van de alternatieve hypothese (er is een interventie-effect), maar ook voor de



nulhypothese (er is geen interventie-effect). Hierdoor biedt het huidige onderzoek een completer beeld van wat de data suggereren dan wanneer alleen klassieke statistiek was gebruikt, aangezien  $p$ -waarden geen onderscheid kunnen maken tussen het ontbreken van bewijs en het bewijs dat er geen effect is (Wagenmakers et al., 2017).

Een beperking van het huidige onderzoek is dat er geen verklaring is gevonden voor de verschillen in de resultaten van de klassieke statistiek uitgevoerd door Palade en Pascal en het huidige onderzoek. Een aantal herberekeningen zijn daarom zowel door de onderzoeker als de scriptiebegeleider uitgevoerd, deze kwamen wel met elkaar overeen.

Daarnaast zijn er beperkingen met betrekking tot de statistische conclusies, doordat niet aan de assumpties van normaalverdeling bij de RM-ANOVA's en uitbijters bij de regressies is voldaan. Dit kan leiden tot onjuiste  $p$ -waarden of Bayesfactoren en daardoor mogelijk tot verkeerde conclusies. De analyses met en zonder uitbijters bij de lineaire regressies waren nuttig, omdat daarmee duidelijk werd welke invloed de uitbijters op de resultaten hadden. De invloed van uitbijters had echter ook nog anders onderzocht kunnen worden — gezien het grote aantal uitbijters — door leverage toe te passen. Leverage geeft de invloed van een individu op de voorspelde score aan. Dit is niet alleen afhankelijk van de scores op de predictoren, maar ook van de score op de afhankelijke variabele (Osborne, 2015).

Een andere beperking is dat bij de robuustheidscontroles van de Bayesiaanse RM-ANOVA's en regressies bleek dat conclusies niet consistent waren. Dit sluit aan bij bevindingen van Van Doorn et al. (2020): de keuze van de prior distributie kan de conclusies beïnvloeden, bijvoorbeeld bij kleine steekproeven.

## **Aanbevelingen**

De resultaten van het huidige onderzoek lijken te suggereren dat Bayesiaanse statistiek op zijn minst een goede aanvulling kan zijn op klassieke statistiek voor de bewijsvoering, doordat het een completer beeld geeft van de informatie die in de data is opgenomen. Daarom wordt aanbevolen om bij pedagogisch onderzoek, naast de klassieke statistiek, vaker Bayesiaanse statistiek toe te passen, aangezien beide benaderingen waarde toevoegen aan statistische conclusies. Vanwege de onbekendheid van veel onderzoekers met Bayesiaanse statistiek, wordt het aanbevolen om gebruik te maken van praktijkgerichte wetenschappelijke artikelen. Een voorbeeld hiervan is de handleiding voor het uitvoeren en rapporteren van een Bayesiaanse analyse in JASP van Van Doorn et al. (2020).

Aanbevolen wordt om een conceptuele replicatie van het onderzoek van Palade en Pascal uit te voeren om meer inzicht te verkrijgen over het verband tussen empathie, verbaal en fysiek pesten. Deze replicatie moet voldoende power hebben om een effect te detecteren, daarom wordt aanbevolen de steekproefgrootte te vergroten. Door extra variabelen, zoals de rol van de leerkracht en geslacht, te onderzoeken worden de gerapporteerde resultaten beter generaliseerbaar naar de populatie.

## Referenties

- Akinwande, M. O., Dikko, H. G., & Agboola, S. (2015). Variance Inflation Factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, *05*(07), 754–767. <https://doi.org/10.4236/ojs.2015.57075>
- Altman, N., & Krzywinski, M. (2016). Analyzing outliers: influential or nuisance? *Nature Methods*, *13*(4), 281–2. <https://doi-org.proxy-ub.rug.nl/10.1038/nmeth.3812>
- Box, G. E. P. (1979). *Robustness in the strategy of scientific model building*. In Elsevier eBooks (pp. 201–236). <https://doi.org/10.1016/b978-0-12-438150-6.50018-2>
- Brown, J. L., Jones, S. M., LaRusso, M. D., & Aber, J. L. (2010). Improving classroom quality: Teacher influences and experimental impacts of the 4rs program. *Journal of Educational Psychology*, *102*(1), 153–167. <https://doi.org/10.1037/a0018160>
- Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, *53*(2), 413. <https://doi.org/10.2307/1128984>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159. Geraadpleegd op 11 mei 2024, van <https://www2.psych.ubc.ca/~schaller/528Readings/Cohen1992.pdf>
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. <https://doi.org/10.3389/fpsyg.2014.00781>
- Field, S. M., Hoekstra, R., Bringmann, L. F., & Van Ravenzwaaij, D. (2019). When and why to replicate: as easy as 1, 2, 3? *Collabra*, *5*(1). <https://doi.org/10.1525/collabra.218>
- Field, S. M., Wagenmakers, E. J., Newell, B. R., Zeelenberg, R., & Van Ravenzwaaij, D. (2016). Two Bayesian tests of the GLOMOsys Model. *Journal of Experimental Psychology: General*, *145*(12), e81–e95. <https://doi.org/10.1037/xge0000067>
- Flohr, R. (2012). *De Bayesiaanse benadering. Basisprincipes en -technieken van de Bayesiaanse statistiek*. Academic Service.
- Goss-Sampson, M. (2020). *Statistical Analysis in JASP-A Students Guide. JASP v0, 14*. Geraadpleegd op 8 mei 2024, van <https://jasp-stats.org/wp-content/uploads/2020/11/Statistical-Analysis-in-JASP-A-Students-Guide-v14-Nov2020.pdf>

- Goulet-Pelletier, J. C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's *d* family. *The Quantitative Methods for Psychology*, 14(4), 242–265. <https://www.tqmp.org/RegularArticles/vol14-4/p242/p242.pdf>
- Heo, I., & Van de Schoot, R. (2020, September). Tutorial: Advanced Bayesian regression in JASP. Zenodo. <https://doi.org/10.5281/zenodo.3991326>
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioural sciences*. (Vol.663). Houghton Mifflin college division.
- Ioannidis, J. P. A. (2022). Correction: Why most published research findings are false. *PLOS Medicine*, 19(8), e1004085. <https://doi.org/10.1371/journal.pmed.1004085>
- Kerivel, A. (2016, July). *Does emotion have a gender? Violence, bullying and empathy at school. Girls 'and boys 'experiences*. Geraadpleegd op 21 mei 2024, van <https://orbilu.uni.lu/handle/10993/28008>
- Lammers, J., Pelzer, B. J., Hendrickx, J., & Eisinga, R. (2007). *Categorische data analyse met SPSS. Inleiding in loglineaire analysetechnieken*. Van Gorcum.
- Linde, M., & Van Ravenzwaaij, D. (2021). Bayes factor model comparisons across parameter values for mixed models. *Computational Brain & Behavior/Computational Brain & Behavior*, 6(1), 14–27. <https://doi.org/10.1007/s42113-021-00117-y>
- Mackey, A. (2012). Why (or why not), when and how to replicate research. In: G. Porte (Ed.), *Replication research in applied linguistics* (pp. 34–69). Cambridge University Press.
- Marton, F., & Booth, S. (1997). *Learning and awareness*, Lawrence Erlbaum.
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, 73(sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>
- Moore, D. S., & McCabe, G. P. (2006). *Statistiek in de praktijk. Theorieboek*. Academic Service.
- Muhammad, L. N. (2023). Guidelines for repeated measures statistical analysis approaches with basic science research considerations. *the Journal of Clinical Investigation/the Journal of Clinical Investigation*, 133(11). <https://doi.org/10.1172/jci171058>
- Mulder, J., & Wagenmakers, E.-J. (2016). Editor's introduction to the special issue on "Bayes factors for testing hypotheses in psychological research: Practical relevance and new developments." *Journal of Mathematical Psychology*, 72, 1–5. <http://dx.doi.org/10.1016/j.jmp.2016.01.002>

- Mynard, H., & Joseph, S. (2000). Development of the multidimensional peer-victimization scale. *Aggressive Behavior*, 26(2), 169. [https://doi.org/10.1002/\(sici\)1098-2337\(2000\)26:2](https://doi.org/10.1002/(sici)1098-2337(2000)26:2)
- Núñez-Regueiro, F. & Núñez-Regueiro, S. (2021). Identifying salient stressors of adolescence: A systematic review and content analysis. *Journal of Youth and Adolescence*, 50(12), 2533–2556. <https://doi.org/10.1007/s10964-021-01492-2>
- Osborne, J. (2015). Best practices in logistic regression. (Vols. 1-0). SAGE Publications, Ltd, <https://doi.org/10.4135/9781483399041>
- OSF. (z.d.). *Discover public research*. Geraadpleegd op 12 december 2023, van <https://osf.io>
- Palade, T., & Pascal, E. (2023). Reducing bullying through empathy training: the effect of teacher's passive presence. *Behavioral Sciences*, 13(3), 216. <https://doi.org/10.3390/bs13030216>
- Palupi, W. R., & Rizki, L. K. (2020). Analysis of factor affecting infant mortality rate (IMR) in East Java using multiple linear regression. *Jurnal Biometrika Dan Kependudukan*, 9(1), 69. <https://doi.org/10.20473/jbk.v9i1.2020.69-76>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7, 531–536. <http://dx.doi.org/10.1177/1745691612463401>
- Pek, J., & Van Zandt, T. (2019). Frequentist and Bayesian approaches to data analysis: Evaluation and estimation. *Psychology Learning & Teaching*, 19(1), 21–35. <https://doi.org/10.1177/1475725719874542>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1), 21–33. Geraadpleegd op 15 mei 2024, van [https://www.nbi.dk/~petersen/Teaching/Stat2017/Power\\_Comparisons\\_of\\_Shapiro-Wilk\\_Kolmogorov-Smirn.pdf](https://www.nbi.dk/~petersen/Teaching/Stat2017/Power_Comparisons_of_Shapiro-Wilk_Kolmogorov-Smirn.pdf)
- Ritchie, S. (2020). *Science fictions: Exposing fraud, bias, negligence and hype in science*. Random House.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374. <https://doi.org/10.1016/j.jmp.2012.08.001>
- Sawyer, S. F. (2009). Analysis of variance: the fundamental concepts. *Journal of Manual & Manipulative Therapy*, 17(2), 27E–38E. <https://doi.org/10.1179/jmt.2009.17.2.27e>
- Scheepers, P., Tobi, H. & Boeije, H. (2016). *Onderzoeksmethoden*. Boom.

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Siero, F., Huisman, M., & Kiers, H. (2009). *Voortgezette regressie- en variantieanalyse*. Bohn Stafleu van Loghum.
- Stavrinos, P., Georgiou, S., & Theofanous, V. (2010). Bullying and empathy: a short-term longitudinal investigation. *Educational Psychology, 30*(7), 793–802.  
<https://doi.org/10.1080/01443410.2010.506004>
- Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*(1), 59–71.  
<https://doi.org/10.1177/1745691613514450>
- Solomon, M. (2007). *Social empiricism*. MIT press.
- Stevens, L., & Meester, R. (2021). Correct redeneren: wat een Bayesiaanse analyse zegt over vier recente strafrechtelijke uitspraken. *Expertise en Recht, 2021*(3), 112–121.  
<https://research.vu.nl/ws/portalfiles/portal/131620353/EeR.pdf>
- Van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods, 22*(2), 217–239. <https://doi.org/10.1037/met0000100>
- Van den Bergh, D., Van Doorn, J., Marsman, M., Draws, T., Van Kesteren, E. J., Derks, K., Dablander, F., Gronau, Q. F., Kucharsky, S., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2020). A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année psychologique, 120*(1), 73–96.  
<https://doi.org/10.3917/anpsy1.201.0073>
- Van den Bergh, D., Wagenmakers, E., & Aust, F. (2023). Bayesian Repeated-Measures Analysis of Variance: An updated methodology implemented in JASP. *Advances in Methods and Practices in Psychological Science, 6*(2).  
<https://doi.org/10.1177/25152459231168024>
- Van Dongen, N., Wagenmakers, E.-J., & Sprenger, J. (2023). A Bayesian perspective on severity: Risky predictions and specific hypotheses. *Psychonomic Bulletin & Review, 30*, 516–533. <https://doi.org/10.3758/s13423-022-02069-1>
- Van Doorn, J., Van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E. (2020). The JASP guidelines for conducting and reporting a

- Bayesian analysis. *Psychonomic Bulletin & Review*, 28(3), 813–826.  
<https://doi.org/10.3758/s13423-020-01798-5>
- Van IJendoorn, M. H., & de Frankrijker, H. (2005). *Pedagogiek in beeld: Een inleiding in de pedagogische studie van opvoeding, onderwijs en hulpverlener*. Bohn Stafleu van Loghum.
- Verhagen, J., & Wagenmakers, E. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology. General*, 143(4), 1457–1475.  
<https://doi.org/10.1037/a0036731>
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2017). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57.  
<https://doi.org/10.3758/s13423-017-1343-3>
- Waterink, W., De Witte, M., & Van Hooren, S. A. H. (2021). Introductie van Bayesiaanse analyses binnen de vaktherapie: een eerste verkenning. *Tijdschrift voor Vaktherapie*, 1, 32–39. Geraadpleegd op 10 december 2023, van <https://kenvak.nl/wp-content/uploads/2021/01/Waterink-W.-De-Witte-M.-en-Van-Hooren-S.-2021.-Introductie-van-Bayesiaanse-analyses-binnen-de-vaktherapie.pdf>
- Westera, W. (2021). Comparing Bayesian statistics and frequentist statistics in serious games research. *International Journal of Serious Games*, 8(1), 27–44.  
<https://doi.org/10.17083/ijsg.v8i1.403>
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: an empirical comparison using 855 t Tests. *Perspectives on Psychological Science*, 6(3), 291–298.  
<https://doi.org/10.1177/1745691611406923>

## Bijlage A. Beschrijvende statistieken en correlaties

### Beschrijvende statistieken

Tabel 1 geeft de beschrijvende statistieken weer van de variabelen empathie, fysiek pesten en verbaal pesten: gemiddelden, standaarddeviaties en minimum- en maximumwaarde per conditie en meetmoment. Figuur 2, 3 en 4 geven de veranderingen in empathie, verbaal en fysiek pesten als effect van de experimentele conditie en het meetmoment weer. Bij leerlingen die een interventie kregen (met en zonder leraar), nam het gemiddelde van empathie toe bij de drie meetmomenten, bij de controlegroep nam dit af (van  $M_{pre} = 12.23$  naar  $M_{follow-up} = 12.05$ ). Bij experimentgroep 1 (met leraar) werd een significant verschil gevonden tussen de pre-test en de follow-up test ( $M_{dif} = -3.50, p = 0.01$ ) van empathie. Het gemiddelde van fysiek en verbaal pesten van het totaal aantal leerlingen nam af bij de drie meetmomenten. Bij experimentgroep 1 werden significante verschillen gevonden voor verbaal pesten tussen de pre-test met de post-test ( $M_{dif} = 2.00, p = 0.01$ ) en met de follow-up test ( $M_{dif} = 3.55, p < 0.001$ ).

### Correlaties tussen empathie, verbaal en fysiek pesten

Tabel 2 geeft de correlaties tussen de drie meetmomenten van empathie, verbaal en fysiek pesten weer met klassieke en Bayesiaanse statistiek van het totaal aantal leerlingen en de controlegroep, Tabel 3 van de experimentgroepen 1 en 2. Bij de totale groep leerlingen en bij experimentgroep 1 zijn geen significante correlaties gevonden. Bij de controlegroep en bij experimentgroep 2 komen de significante  $p$ -waarden met een  $\alpha < .05$  niet consistent overeen met overtuigende Bayesfactoren, met een  $\alpha < .01$  is dit wel consistent. Bijvoorbeeld correleren bij de controlegroep de post-test van verbaal pesten en de pre-test van fysiek pesten significant negatief ( $r = -0.41, p < 0.05$ ), maar geeft de Bayesfactor zwak bewijs voor de alternatieve hypothese ( $BF_{01} = 0.38$ ). Bij experimentgroep 2 correleren de pre-test van verbaal pesten en de pre-test van fysiek pesten significant ( $r = 0.61, p < 0.01$ ) en geeft de Bayesfactor sterk bewijs voor de alternatieve hypothese ( $BF_{01} = 0.06$ ).



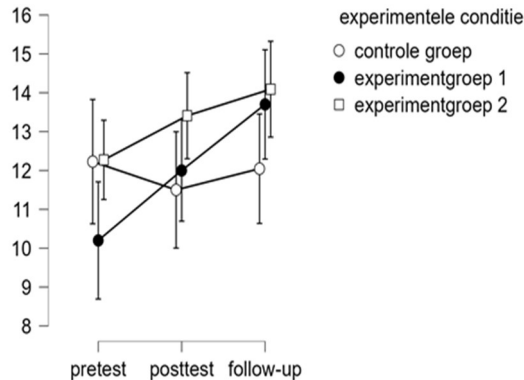
**Tabel 1**

Beschrijvingen van de in de analyse opgenomen variabelen: gemiddelden, standaarddeviaties en minimum- en maximumwaarde voor empathie, fysiek pesten en verbaal pesten per onderzoeksconditie en meetmoment van de test

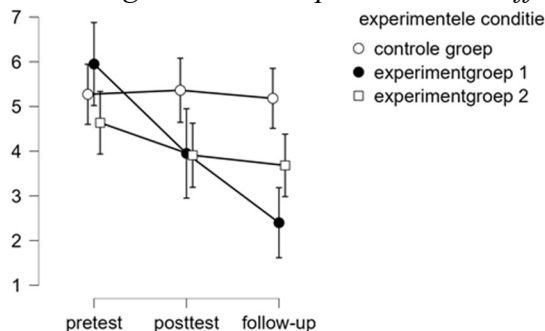
Conditie van de groep	Meet-moment	Empathie				Fysiek pesten				Verbaal pesten			
		Min	Max	M	SD	Min	Max	M	SD	Min	Max	M	SD
Experiment groep 1	Pre-test	4.00	19.00	10.20	3.79	0.00	8.00	5.45	2.35	1.00	8.00	5.95	1.96
	Post-test	6.00	16.00	12.00	2.96	0.00	8.00	4.85	2.32	0.00	8.00	3.95	2.67
	Follow-up	8.00	18.00	13.70	2.83	0.00	8.00	4.55	2.50	0.00	5.00	2.40	1.57
Experiment groep 2	Pre-test	9.00	15.00	12.27	1.91	0.00	8.00	5.00	2.47	1.00	8.00	4.64	1.84
	Post-test	8.00	19.00	13.41	2.56	0.00	8.00	4.59	2.11	0.00	7.00	3.91	1.41
	Follow-up	9.00	19.00	14.09	2.88	0.00	7.00	4.32	1.84	0.00	7.00	3.68	1.94
Controle groep	Pre-test	3.00	18.00	12.23	3.85	2.00	7.00	4.46	1.44	3.00	8.00	5.27	1.45
	Post-test	4.00	18.00	11.50	3.42	1.00	8.00	4.41	1.99	3.00	8.00	5.36	1.59
	Follow-up	5.00	20.00	12.05	3.97	0.00	7.00	4.32	1.86	2.00	8.00	5.18	1.56

**Figuur 2**

Veranderingen in empathie als een effect van de experimentele conditie en meetmoment

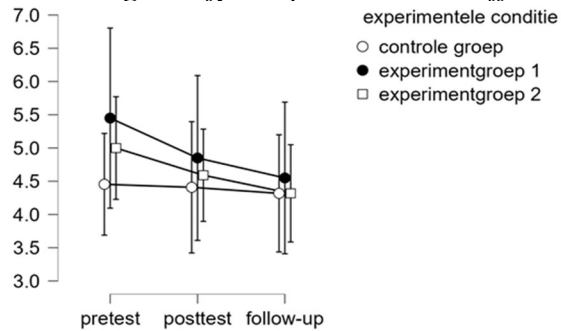

**Figuur 3**

Veranderingen in verbaal pesten als een effect van de experimentele conditie en meetmoment



**Figuur 4**

*Veranderingen in fysiek pesten als een effect van de experimentele conditie en meetmoment*


**Tabel 2**

*Correlaties van de totale groep en de controlegroep met klassieke en Bayesiaanse statistiek*

Variabele		1	2	3	4	5	6	7	8	9
1. Verbaal pesten pre-test	Pearson's $r$	—	0.09	0.04	0.44*	0.09	-0.28	0.16	-0.21	0.01
	BF <sub>01</sub>	—	3.54	3.73	0.53	3.51	1.79	3.00	2.49	3.78
2. Verbaal pesten post-test	Pearson's $r$	0.06	—	-0.01	-0.41*	0.21	0.04	-0.03	0.53*	0.15
	BF <sub>01</sub>	5.84	—	3.78	0.38	2.54	3.73	3.76	0.17	3.09
3. Verbaal pesten follow up	Pearson's $r$	0.05	0.28	—	-0.02	-0.18	0.24	0.41	-0.01	0.61**
	BF <sub>01</sub>	5.95	0.59	—	3.78	2.82	2.18	0.69	3.78	0.06
4. Fysiek pesten pre-test	Pearson's $r$	0.40	-0.18	0.05	—	-0.33	-0.07	0.14	0.41	0.15
	BF <sub>01</sub>	0.03	2.32	6.02	—	1.28	3.60	3.20	0.72	3.10
5. Fysiek pesten post-test	Pearson's $r$	0.08	0.23	0.06	-0.06	—	-0.33	-0.22	0.44*	-0.13
	BF <sub>01</sub>	5.24	1.22	5.77	5.81	—	1.29	2.43	0.54	3.25
6. Fysiek pesten follow up	Pearson's $r$	-0.15	-0.08	0.12	0.03	0.02	—	0.23	-0.03	-0.08
	BF <sub>01</sub>	3.28	5.22	4.05	6.20	6.32	—	2.31	3.76	2.17
7. Empathie pre-test	Pearson's $r$	0.18	0.17	0.27	0.01	-0.12	-0.14	—	0.03	0.24
	BF <sub>01</sub>	2.50	2.78	0.67	6.40	4.16	3.51	—	2.76	2.17
8. Empathie post-test	Pearson's $r$	0.05	0.20	0.13	0.08	0.22	0.11	0.10	—	0.28
	BF <sub>01</sub>	5.93	1.81	3.90	5.41	1.49	4.46	4.62	—	1.84
9. Empathie follow up	Pearson's $r$	0.12	0.03	0.13	0.11	0.12	-0.04	0.09	0.20	—
	BF <sub>01</sub>	4.26	6.26	3.78	4.50	4.14	6.16	5.04	1.98	—

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Coëfficiënten onder de diagonaal representeren correlaties tussen het totaal aantal leerlingen ( $N = 64$ ); coëfficiënten boven de diagonaal representeren de correlaties bij de controlegroep ( $N = 22$ )

**Tabel 3**
*Correlaties van de experimentgroep 1 en experimentgroep 2 met klassieke en Bayesiaanse statistiek*

Variabele		1	2	3	4	5	6	7	8	9
1. Verbaal pesten pre-test	Pearson's $r$	—	0.04	0.31	0.61**	0.19	0.06	0.34	0.05	0.20
	$BF_{01}$	—	3.73	1.47	0.06	2.67	3.65	1.21	3.69	2.65
2. Verbaal pesten post-test	Pearson's $r$	0.14	—	0.11	0.26	0.39	0.36	-0.29	0.25	-0.12
	$BF_{01}$	3.07	—	3.38	1.99	0.85	1.05	1.68	2.11	3.35
3. Verbaal pesten follow up	Pearson's $r$	0.06	0.28	—	0.37	0.62**	0.39	-0.35	0.17	0.34
	$BF_{01}$	3.51	1.83	—	0.99	0.05	0.82	1.15	2.88	1.24
4. Fysiek pesten pre-test	Pearson's $r$	0.17	-0.26	0.08	—	0.49*	0.38	0.06	0.39	0.21
	$BF_{01}$	2.88	2.07	3.44	—	0.32	0.92	3.66	0.87	2.52
5. Fysiek pesten post-test	Pearson's $r$	-0.06	0.26	-0.19	-0.52	—	0.38	-0.09	0.25	0.34
	$BF_{01}$	3.51	2.01	2.71	0.28	—	0.91	3.52	2.06	1.26
6. Fysiek pesten follow up	Pearson's $r$	-0.28	-0.33	-0.07	-0.22	-0.00	—	-0.34	0.47*	-0.20
	$BF_{01}$	1.81	1.43	3.46	2.38	3.61	—	1.01	0.39	2.56
7. Empathie pre-test	Pearson's $r$	0.37	0.38	0.30	0.01	-0.01	-0.35	—	-0.00	-0.12
	$BF_{01}$	1.08	1.00	1.63	3.61	3.61	1.28	—	3.78	3.33
8. Empathie post-test	Pearson's $r$	0.38	2.81	0.11	3.25	3.59	3.61	2.40	—	3.77
	$BF_{01}$	0.38	2.81	0.11	3.25	3.59	3.61	2.40	—	3.77
9. Empathie follow up	Pearson's $r$	0.26	0.28	-0.13	-0.15	0.20	0.13	0.08	0.08	—
	$BF_{01}$	2.01	1.81	3.17	3.04	2.58	3.15	3.42	3.45	—

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ . Coëfficiënten onder de diagonaal representeren correlaties van experimentgroep 1 ( $N = 20$ ); coëfficiënten boven de diagonaal representeren de correlaties van experimentgroep 2 ( $N = 22$ ).

## Bijlage B. Klassieke en Bayesiaanse ANOVA's en assumptiecontroles

### Assumptiecontroles

**Tabel 1**

*Normaal verdeling van de testen van empathie met Shapiro-Wilk toets, skewness en kurtosis voor de variabelen empathie, verbaal en fysiek pesten per meetmoment van de totale groep leerlingen*

Variabele	Meetmoment	W	Skewness	Kurtosis
Empathie	Pre-test	0.98 ( $p = 0.25$ )	-0.20	-0.01
	Post-test	0.98 ( $p = 0.24$ )	-0.35	0.21
	Follow up test	0.98 ( $p = 0.54$ )	-0.20	-0.22
Verbaal pesten	Pre-test	0.95 ( $p = 0.01$ )*	-0.26	-0.42
	Post-test	0.95 ( $p = 0.01$ )*	-0.30	-0.23
	Follow up test	0.95 ( $p = 0.01$ )*	-0.22	-0.41
Fysiek pesten	Pre-test	0.92 ( $p < 0.001$ )*	-0.69	-0.27
	Post-test	0.95 ( $p = 0.02$ )*	-0.15	-0.80
	Follow up test	0.96 ( $p = 0.02$ )*	-0.37	-0.40

\* Voldoet niet aan de normaalverdeling

**Tabel 2**

*Resultaten met betrekking tot de normaalverdeling van de testen van empathie met de Shapiro-Wilk toets, skewness en kurtosis per conditie en meetmoment*

Conditie	Meetmoment	W	Skewness	Kurtosis
Experiment-groep 1	Pre-test	0.94 ( $p = 0.26$ )	0.77	0.20
	Post-test	0.93 ( $p = 0.14$ )	-0.60	-0.29
	Follow up test	0.96 ( $p = 0.59$ )	-0.20	-0.69
Experiment-groep 2	Pre-test	0.89 ( $p = 0.02$ )*	-0.20	-1.46*
	Post-test	0.97 ( $p = 0.68$ )	0.29	0.43
	Follow up test	0.96 ( $p = 0.45$ )	0.16	-0.51
Controlegroep	Pre-test	0.93 ( $p = 0.14$ )	-0.64	0.50
	Post-test	0.98 ( $p = 0.93$ )	-0.20	0.02
	Follow up test	0.98 ( $p = 0.83$ )	0.12	-0.38

\* Voldoet niet aan de normaalverdeling

**Tabel 3**

*Resultaten met betrekking tot de normaalverdeling van de testen van verbaal pesten met de Shapiro-Wilk toets, skewness en kurtosis per conditie en meetmoment*

Conditie	Meetmoment	W	Skewness	Kurtosis
Experiment-groep 1	Pre-test	0.89 ( $p = 0.03$ )*	-0.95	0.46
	Post-test	0.92 ( $p = 0.08$ )	-0.13	-1.37*
	Follow up test	0.91 ( $p = 0.054$ )	-0.29	-0.97
Experiment-groep 2	Pre-test	0.96 ( $p = 0.48$ )	-0.01	-0.21
	Post-test	0.91 ( $p = 0.053$ )	-0.50	2.38
	Follow up test	0.92 ( $p = 0.08$ )	-0.54	0.02
Controlegroep	Pre-test	0.94 ( $p = 0.17$ )	0.19	-0.39
	Post-test	0.92 ( $p = 0.08$ )	0.12	-1.10
	Follow up test	0.96 ( $p = 0.42$ )	-0.17	-0.65

\* Voldoet niet aan de normaalverdeling

**Tabel 4**

*Resultaten met betrekking tot de normaalverdeling van de testen van fysiek pesten met de Shapiro-Wilk toets, skewness en kurtosis per conditie en meetmoment*

Conditie	Meetmoment	W	Skewness	Kurtosis
Experiment-groep 1	Pre-test	0.84 ( $p = 0.00$ )*	-1.23	0.79
	Post-test	0.94 ( $p = 0.20$ )	-0.22	-0.80
	Follow up test	0.94 ( $p = 0.19$ )	-0.26	-0.82
Experiment-groep 2	Pre-test	0.88 ( $p = 0.01$ )*	-0.77	-0.54
	Post-test	0.95 ( $p = 0.29$ )	-0.37	-0.40
	Follow up test	0.93 ( $p = 0.13$ )	-0.72	0.24
Controlegroep	Pre-test	0.92 ( $p = 0.06$ )	-0.38	-0.70
	Post-test	0.95 ( $p = 0.35$ )	0.05	-1.05
	Follow up test	0.94 ( $p = 0.16$ )	-0.51	-0.36

\* Voldoet niet aan de normaalverdeling

**Tabel 5**

*Resultaten van een Levene's test van de homogeniteit van de varianties voor de variabelen empathie, verbaal en fysiek pesten, per meetmoment van de totale groep*

<b>Variabele</b>	<b>Meetmoment</b>	<b>F</b>	<b>df1</b>	<b>df2</b>	<b>p</b>
Empathie	Pre-test	2.84	2	61	0.07
	Post-test	1.06	2	61	0.35
	Follow-up test	1.15	2	61	0.32
Verbaal pesten	Pre-test	0.93	2	61	0.40
	Post-test	8.41	2	61	< .001*
	Follow-up test	0.22	2	61	0.80
Fysiek pesten	Pre-test	2.43	2	61	0.10
	Post-test	0.34	2	61	0.71
	Follow-up test	2.21	2	61	0.12

\* Voldoet niet aan de assumptie van homogeniteit van de varianties

**Tabel 6**

*Resultaten van een Mauchly's W test van sphericiteit van de totale groep en per conditie voor de variabelen empathie, verbaal en fysiek pesten*

<b>Variabele</b>	<b>Conditie</b>	<b>Mauchly's W</b>	<b>df</b>	<b>p</b>
Empathie	Totale groep	0.99	2	0.87
	Experimentgroep 1	0.97	2	0.78
	Experimentgroep 2	0.95	2	0.61
	Controlegroep	0.98	2	0.80
Verbaal pesten	Totale groep	0.98	2	0.54
	Experimentgroep 1	0.93	2	0.50
	Experimentgroep 2	0.99	2	0.99
	Controlegroep	0.99	2	0.92
Fysiek pesten	Totale groep	0.99	2	0.81
	Experimentgroep 1	0.96	2	0.70
	Experimentgroep 2	0.99	2	0.86
	Controlegroep	0.92	2	0.42

## Klassieke ANOVA's

**Tabel 7**

*Binnen-groeps-effecten van een RM-ANOVA van de variabelen empathie, verbaal en fysiek pesten van de totale groep*

Variabele	Cases	Sum of Squares	df	F	p	$\eta^2_p$
Empathie	Herhaalde metingen	94.22	2.00	5.24	0.01	0.08
	Herhaalde metingen * experimentele conditie	77.51	4.00	2.16	0.08	0.07
Verbaal pesten	Herhaalde metingen	69.54	2.00	12.10	< .001	0.17
	Herhaalde metingen * experimentele conditie	68.46	1.96	5.96	< .001	0.16
Fysiek pesten	Herhaalde metingen	10.66	2.00	1.18	0.31	0.02
	Herhaalde metingen * experimentele conditie	3.50	4.00	0.19	0.94	0.01

**Tabel 8**

*Tussen-groepen-effecten van een RM-ANOVA van de variabelen empathie, verbaal en fysiek pesten van de totale groep*

Variabele	Sum of Squares	df	F	p	$\eta^2_p$
Empathie	74.74	2	2.96	0.06	0.09
Verbaal pesten	60.88	2	7.60	0.001	0.20
Fysiek pesten	9.74	2	1.11	0.34	0.04

**Tabel 9**

*Binnen-groeps-effecten van een RM-ANOVA van de variabelen empathie, verbaal en fysiek pesten van groep 1*

Variabele	Sum of Squares	df	F	p	$\omega^2$
Empathie	122.53	2	6.76	0.003	0.14
Verbaal pesten	126.70	2	16.79	< .001	0.31
Fysiek pesten	8.40	2	0.59	0.56	0.00

**Tabel 10**

*Binnen-groeps-effecten van een RM-ANOVA van de variabelen empathie, verbaal en fysiek pesten van groep 2*

Variabele	Sum of Squares	df	F	p	$\omega^2$
Empathie	37.12	2	2.89	0.07	0.06
Verbaal pesten	10.94	2	2.16	0.13	0.03
Fysiek pesten	5.18	2	0.95	0.40	0.00

**Tabel 11**

*Binnen-groeps-effecten van een RM-ANOVA van de variabelen empathie, verbaal en fysiek pesten van de controlegroep*

Variabele	Sum of Squares	df	F	p	$\omega^2$
Empathie	6.30	2	0.27	0.76	0.00
Verbaal pesten	0.36	2	0.08	0.93	0.00
Fysiek pesten	0.21	2	0.03	0.97	0.00

**Bayesiaanse ANOVA's**
**Tabel 12**

*Effectanalyse van Bayesiaanse herhaalde metingen ANOVA's voor empathie*

Effect		P (incl)	P (excl)	P (incl  data)	P (excl  data)	BF <sub>incl</sub>	BF <sub>excl</sub>	BF <sub>incl*</sub>	BF <sub>incl</sub> **
Totale groep	Herhaalde metingen	0.60	0.40	0.40	0.15	3.68	0.27	7.17	1.20
	Experimentele conditie	0.60	0.40	0.40	0.49	0.71	1.41	1.73	0.20
	Herhaalde metingen * experimentele conditie	0.20	0.80	0.80	0.80	1.02	0.98	2.59	0.12
Experiment- groep 1	Herhaalde metingen	0.50	0.50	0.50	0.04	25.84	0.04	18.82	19.84
Experiment- groep 2	Herhaalde metingen	0.50	0.50	0.50	0.38	1.64	0.61	2.02	0.86
Controlegroep	Herhaalde metingen	0.50	0.50	0.50	0.87	0.16	6.43	0.42	0.05

\* Robuustheidscontrole met  $r = 1$

\*\* Robuustheidscontrole met  $r = 0.2$

**Tabel 13**

*Effectanalyse van Bayesiaanse herhaalde metingen ANOVA's voor verbaal pesten*

Effect		P (incl)	P (excl)	P (incl  data)	P (excl  data)	BF <sub>incl</sub>	BF <sub>excl</sub>	BF <sub>incl*</sub>	BF <sub>incl</sub> **
Totale groep	Herhaalde metingen	0.60	0.40	1.00	$5.41 \times 10^{-6}$	123240.17	$8.11 \times 10^{-6}$	61919.44	47892.87
	Experimentele conditie	0.60	0.40	1.00	$3.47 \times 10^{-4}$	1923.19	$5.20 \times 10^{-4}$	1229.59	741.04
	Herhaalde metingen * experimentele conditie	0.20	0.80	1.00	0.004	913.00	0.001	546.95	601.27
Experiment- groep 1	Herhaalde metingen	0.50	0.50	1.00	$6.00 \times 10^{-5}$	16671.27	$6.00 \times 10^{-5}$	8168.74	19394.83
Experiment- groep 2	Herhaalde metingen	0.50	0.50	0.41	0.59	0.71	1.41	1.12	0.32
Controle- groep	Herhaalde metingen	0.50	0.50	0.12	0.88	0.14	7.36	0.38	0.04

\* Robuustheidscontrole met  $r = 1$

\*\* Robuustheidscontrole met  $r = 0.2$



**Tabel 14**
*Effectanalyse van Bayesiaanse herhaalde metingen ANOVA's voor fysiek pesten*

	Effect	P (incl)	P (excl)	P (incl  data)	P (excl  data)	BF <sub>incl</sub>	BF <sub>excl</sub>	BF <sub>incl</sub> *	BF <sub>incl</sub> **
Totale groep	Herhaalde metingen	0.60	0.40	1.15	0.85	0.12	8.68	0.34	0.04
	Experimentele conditie	0.60	0.40	1.13	0.87	0.10	9.98	0.31	0.03
	Herhaalde metingen * experimentele conditie	0.20	0.80	8.49× 10 <sup>-4</sup>	1.00	0.00	294.17	0.12	5.86× 10 <sup>-5</sup>
Experiment- groep 1	Herhaalde metingen	0.50	0.50	0.19	0.81	0.24	4.19	0.54	0.09
Experiment- groep 2	Herhaalde metingen	0.50	0.50	0.20	0.80	0.25	4.01	0.57	0.10
Controle- groep	Herhaalde metingen	0.50	0.50	0.11	0.89	0.13	7.91	0.37	0.04

 \* Robuustheidscontrole met  $r = 1$ 

 \*\* Robuustheidscontrole met  $r = 0.2$ 
**Tabel 15**
*Samenvatting van de gemiddelde posteriorverdeling van de toetsen van empathie van groep 1 met het gemiddelde, de standaarddeviatie en het betrouwbaarheidsinterval*

Variabele	Toets	Gemiddelde	SD	95% Betrouwbaarheidsinterval	
				Laag	Hoog
Intercept		11.97	0.55	10.86	13.06
Herhaalde metingen	Pre-test	-1.50	0.53	-2.60	-0.45
	Post-test	0.02	0.49	-0.98	0.97
	Follow-up test	1.48	0.52	0.44	2.53

**Tabel 16**
*Samenvatting van de gemiddelde posteriorverdeling van de toetsen van verbaal pesten van groep 1 met het gemiddelde, de standaarddeviatie en het betrouwbaarheidsinterval*

Variabele	Toets	Gemiddelde	SD	95% Betrouwbaarheidsinterval	
				Laag	Hoog
Intercept		4.10	0.36	3.38	4.83
Herhaalde metingen	Pre-test	1.71	0.36	0.97	2.40
	Post-test	-0.14	0.34	-0.83	0.52
	Follow-up	-1.57	0.35	-2.28	-0.87

## Bijlage C. Lineaire regressies

### Multicollineariteit en uitbijters

Er is sprake van multicollineariteit als twee of meer onafhankelijke variabelen sterk met elkaar correleren. Gevolgen van (te) grote samenhang zijn dat het belang van afzonderlijke variabelen moeilijk te bepalen zijn en er grotere standaardfouten (SE) zijn. Hierbij duidde een VIF-waarde hoger dan 4 op hoge onderlinge correlatie. In Tabel 1, 2 en 3 zijn de VIF-waarden weergegeven en het aantal uitbijters berekend met Cook's Distance en is onderzocht of er data waren met meer dan drie standaarddeviaties vanaf het gemiddelde.

Bij alle modellen is geen sprake van multicollineariteit, de waarden zijn onder de 4. De uitbijters zijn bepaald met Cook's Distance. Cook's Distance is een maat voor de invloed van het individu op de eigen fit en die van andere individuen en daarmee op de schatting van alle parameters tezamen in het model (Lammers et al., 2007).

**Tabel 1**

*Multicollineariteit van de vier regressiemodellen en het aantal uitbijters, van experimentgroep 1*

Variabele	Model 1a*	Model 1b**	Model 2a ***	Model 2b ****
	VIF	VIF	VIF	VIF
Constante				
Geslacht	1.08	1.09	1.13	1.19
Leeftijd	1.17	1.21	1.19	1.21
Verbaal pesten pre-test	1.56	1.63		
Fysiek pesten pre-test			1.19	1.21
Empathie pre-test	1.23	1.23	1.09	1.10
Empathie post-test	1.34	1.35	1.12	1.12
Empathie follow-up test		1.13		1.10
Aantal uitbijters met Cook's Distance	2.00	3.00	1.00	3.00
Aantal uitbijters met 3 SD of meer	0.00	0.00	0.00	0.00

*Noot.* Cook's Distance:  $4/20 = 0.2$  ( $N=20$ )

\* Model voor regressieanalyse met als afhankelijke variabele de post-test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie en post-test empathie

\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie, post-test empathie en follow-up test empathie

\*\*\* Model voor regressieanalyse met als afhankelijke variabele de post-test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie en post-test empathie

\*\*\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie, post-test empathie en follow-up test empathie

**Tabel 2**

*Multicollineariteit van de vier regressiemodellen en het aantal uitbijters, van experimentgroep 2*

Variabele	Model 1a*	Model 1b**	Model 2a ***	Model 2b ****
	VIF	VIF	VIF	VIF
Constante				
Geslacht	1.03	1.03	1.07	1.07
Leeftijd	1.81	1.82	1.84	1.85
Verbaal pesten pre-test	1.16	1.24		
Fysiek pesten pre-test			1.26	1.35
Empathie pre-test	1.13	1.18	1.01	1.02
Empathie post-test	1.82	1.83	2.15	2.19
Empathie follow-up test		1.09		1.09
Aantal uitbijters met Cook's Distance	3.00	2.00	1.00	2.00
Aantal uitbijters met 3 SD of meer	0.00	0.00	0.00	0.00

*Noot.* Cook's Distance:  $4/22 = 0.182$  ( $N = 22$ )

\* Model voor regressieanalyse met als afhankelijke variabele de post-test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie en post-test empathie

\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie, post-test empathie en follow-up test empathie

\*\*\* Model voor regressieanalyse met als afhankelijke variabele de post-test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie en post-test empathie

\*\*\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie, post-test empathie en follow-up test empathie

**Tabel 3**

*Multicollineariteit van de vier regressiemodellen en het aantal uitbijters, van de controlegroep*

Variabele	Model 1a*	Model 1b**	Model 2a ***	Model 2b ****
	VIF	VIF	VIF	VIF
Constante				
Geslacht	1.71	1.78	1.71	1.85
Leeftijd	1.29	1.30	1.19	1.20
Verbaal pesten pre-test	1.31	1.32		
Fysiek pesten pre-test			1.35	1.51
Empathie pre-test	1.29	1.41	1.29	1.42
Empathie post-test	1.46	1.65	1.72	2.13
Empathie follow-up test		1.20		1.34
Aantal uitbijters met Cook's Distance	4.00	2.00	3.00	4.00
Aantal uitbijters met 3 SD of meer	0.00	0.00	0.00	0.00

*Noot.* Cook's Distance:  $4/22 = 0.182$  ( $N = 22$ )

\* Model voor regressieanalyse met als afhankelijke variabele de post-test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie en post-test empathie

\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van verbaal pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test verbaal pesten, pre-test empathie, post-test empathie en follow-up test empathie

\*\*\* Model voor regressieanalyse met als afhankelijke variabele de post-test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie en post-test empathie

\*\*\*\* Model voor regressieanalyse met als afhankelijke variabele de follow-up test van fysiek pesten, als onafhankelijke variabelen geslacht, leeftijd, pre-test fysiek pesten, pre-test empathie, post-test empathie en follow-up test empathie

## Klassieke regressies

**Tabel 4**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten post-test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van experimentgroep 1, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	0.12	1.73	0.54	0.00	0.84
H <sub>1</sub>	0.07	0.10	1.56	0.16	0.36	0.93

\* Twee uitbijters zijn verwijderd

**Tabel 5**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten follow-up test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test, empathie post-test en empathie follow-up test als onafhankelijke variabelen, van experimentgroep 1, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.22	2.29	0.51	0.00	0.97
H <sub>1</sub>	0.28	-0.50	2.96	0.06	0.58	0.07

\* Drie uitbijters zijn verwijderd

**Tabel 6**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten post-test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van experimentgroep 1, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.07	2.02	0.97	0.00	0.58
H <sub>1</sub>	0.14	-0.17	2.30	0.65	0.05	0.38

\* Een uitbijter is verwijderd

**Tabel 7**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten follow-up test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test, empathie post-test en empathie follow-up test als onafhankelijke variabelen, van experimentgroep 1, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	0.18	1.53	0.28	0.00	0.26
H <sub>1</sub>	-0.01	0.21	1.41	0.12	0.04	0.67

\* Drie uitbijters zijn verwijderd

**Tabel 8**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten post-test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van experimentgroep 2, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.03	2.01	0.98	0.00	0.37
H <sub>1</sub>	0.09	-0.23	2.44	0.45	0.04	0.57

\* Drie uitbijters zijn verwijderd

**Tabel 9**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten follow-up test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test, empathie post-test en empathie follow-up test als onafhankelijke variabelen, van experimentgroep 2, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	0.32	1.36	0.12	0.00	0.34
H <sub>1</sub>	0.32	0.16	1.58	0.23	0.17	0.75

\* Twee uitbijters zijn verwijderd

**Tabel 10**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten post-test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van experimentgroep 2, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.23	2.39	0.35	0.00	0.38
H <sub>1</sub>	0.09	-0.44	2.72	0.16	0.15	0.23

\* Een uitbijter is verwijderd

**Tabel 11**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten follow-up test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test, empathie post-test en empathie follow-up test als onafhankelijke variabelen, van experimentgroep 2, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	0.04	1.88	0.77	0.00	0.65
H <sub>1</sub>	0.32	0.11	1.77	0.44	0.55	0.38

\* Twee uitbijters zijn verwijderd

**Tabel 12**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten post-test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van controlegroep, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	0.02	1.79	0.62	0.00	0.52
H <sub>1</sub>	0.12	0.11	1.64	0.28	0.30	0.88

\* Vier uitbijters zijn verwijderd

**Tabel 13**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele verbaal pesten follow-up test, en geslacht, leeftijd, verbaal pesten pre-test, empathie pre-test, empathie post-test en empathie follow-up test als onafhankelijke variabelen, van de controlegroep, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.44	2.69	0.09	0.00	0.81
H <sub>1</sub>	0.29	-0.28	2.48	0.29	0.57	0.76

\* Twee uitbijters zijn verwijderd

**Tabel 14**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten post-test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van de controlegroep, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.35	2.65	0.11	0.00	0.02**
H <sub>1</sub>	0.22	-0.29	2.51	0.29	0.40	0.02**

\* Drie uitbijters zijn verwijderd, \*\*  $p < .05$

**Tabel 15**

*Modelsamenvatting van een lineaire regressie met als afhankelijke variabele fysiek pesten post-test, en geslacht, leeftijd, fysiek pesten pre-test, empathie pre-test en empathie post-test als onafhankelijke variabelen, van de controlegroep, met en zonder uitbijters*

Model	Adjusted R <sup>2</sup>	Durbin-Watson			Zonder uitbijters*	
		Autocorrelatie	Statistiek	p	Adjusted R <sup>2</sup>	p
H <sub>0</sub>	0.00	-0.04	1.98	0.96	0.96	0.89
H <sub>1</sub>	0.45	-0.09	1.99	0.87	0.87	0.47

\* Vier uitbijters zijn verwijderd

## Bayesiaanse regressies

**Tabel 16**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten post-test als afhankelijke variabele, groep 1*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	3.95	0.58	2.58	5.05
Geslacht	0.50	0.50	0.31	0.69	0.45	-0.06	0.53	-1.32	1.44
Leeftijd	0.50	0.50	0.43	0.57	0.77	0.54	0.88	-0.35	3.08
Verbaal pesten pre-test	0.50	0.50	0.32	0.68	0.47	-0.02	0.17	-0.44	0.43
Empathie pre-test	0.50	0.50	0.47	0.53	0.89	0.09	0.13	-0.04	0.42
Empathie post-test	0.50	0.50	0.32	0.68	0.47* <sup>1</sup>	0.02	0.10	-0.17	0.34

\* Bij prior = 0.177,  $BF_{incl} = 0.69$ ; bij prior = 0.708,  $BF_{incl} = 0.23$

1. Bij verwijdering van 2 uitbijters  $BF_{incl} = 0.90$

**Tabel 17**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten follow-up test als afhankelijke variabele, groep 1*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	2.40	0.31	1.77	2.99
Geslacht	0.50	0.50	0.36	0.64	0.56	0.05	0.31	-0.51	1.11
Leeftijd	0.50	0.50	0.36	0.64	0.56	-0.01	0.32	-0.74	0.71
Verbaal pesten pre-test	0.50	0.50	0.46	0.54	0.86	-0.09	0.14	-0.44	0.09
Empathie pre-test	0.50	0.50	0.44	0.56	0.78	0.03	0.06	-0.03	0.19
Empathie post-test	0.50	0.50	0.80	0.21	3.88	0.20	0.13	0.00	0.42
Empathie follow-up test	0.50	0.50	0.38	0.62	0.61* <sup>1</sup>	-0.09	0.06	-0.18	0.11

\* bij prior = 0.177,  $BF_{incl} = 0.79$ ; bij prior = 0.708,  $BF_{incl} = 0.37$

1. Bij verwijdering van 3 uitbijters  $BF_{incl} = 1.24$

**Tabel 18**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten post-test als afhankelijke variabele, groep 1*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.850	0.487	3.862	5.944
Geslacht	0.50	0.50	0.42	0.58	0.73	0.388	0.716	-0.617	2.124
Leeftijd	0.50	0.50	0.33	0.67	0.49	0.044	0.472	-1.107	1.212
Fysiek pesten pre-test	0.50	0.50	0.66	0.34	1.97	-0.26	0.24	-0.69	0.00
Empathie pre-test	0.50	0.50	0.33	0.67	0.49	-0.004	0.06	-0.18	0.14
Empathie post-test	0.50	0.50	0.33	0.67	0.48* <sup>1</sup>	-1.315×10 <sup>-4</sup>	0.08	-0.19	0.22

\* bij prior = 0.177,  $BF_{incl} = 0.69$ ; bij prior = 0.708,  $BF_{incl} = 0.26$   
1. Bij verwijdering van 1 uitbijter  $BF_{incl} = 0.44$

**Tabel 19**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten follow-up test als afhankelijke variabele, groep 1*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.55	0.55	3.43	5.67
Geslacht	0.50	0.50	0.33	0.68	0.48	-0.21	0.60	-1.91	0.80
Leeftijd	0.50	0.50	0.41	0.59	0.7	0.46	0.81	-0.25	2.58
Fysiek pesten pre-test	0.50	0.50	0.30	0.700	0.43	-0.03	0.12	-0.39	0.22
Empathie pre-test	0.50	0.50	0.38	0.62	0.61	-0.05	0.10	-0.31	0.07
Empathie post-test	0.50	0.50	0.28	0.72	0.39	0.01	0.08	-0.17	0.26
Empathie follow-up test	0.50	0.50	0.29	0.71	0.41* <sup>1</sup>	0.01	0.09	-0.17	0.31

\* bij prior = 0.177,  $BF_{incl} = 0.63$ ; bij prior = 0.708,  $BF_{incl} = 0.18$   
1. Bij verwijdering van 3 uitbijters  $BF_{incl} = 0.46$



**Tabel 20**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten post-test als afhankelijke variabele, groep 2*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	3.91	0.29	3.38	4.59
Geslacht	0.50	0.50	0.36	0.64	0.57	0.14	0.34	-0.29	1.02
Leeftijd	0.50	0.50	0.46	0.54	0.86	0.36	0.55	-0.11	1.73
Verbaal pesten pre-test	0.50	0.50	0.33	0.67	0.49	0.02	0.09	-0.12	0.25
Empathie pre-test	0.50	0.50	0.42	0.58	0.71	-0.06	0.11	-0.35	0.04
Empathie post-test	0.50	0.50	0.34	0.66	0.51* <sup>1</sup>	0.01	0.07	-0.16	0.20

\* bij prior = 0.177,  $BF_{incl} = 0.73$ ; bij prior = 0.708,  $BF_{incl} = 0.25$   
 1. Bij verwijdering van 3 uitbijters  $BF_{incl} = 0.59$

**Tabel 21**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten follow-up test als afhankelijke variabele, groep 2*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	3.68	0.37	2.96	4.48
Geslacht	0.50	0.50	0.62	0.38	1.64	0.58	0.65	-0.19	1.95
Leeftijd	0.50	0.50	0.45	0.55	0.81	0.08	0.52	-1.05	1.41
Verbaal pesten pre-test	0.50	0.50	0.67	0.33	2.02	0.22	0.22	-0.01	0.67
Empathie pre-test	0.50	0.50	0.68	0.32	2.17	-0.23	0.21	-0.64	0.00
Empathie post-test	0.50	0.50	0.46	0.54	0.84	0.03	0.10	-0.18	0.30
Empathie follow-up test	0.50	0.50	0.52	0.48	1.10* <sup>1</sup>	0.05	0.09	-0.07	0.28

\* bij prior = 0.177,  $BF_{incl} = 1.23$ ; bij prior = 0.708,  $BF_{incl} = 0.73$   
 1. Bij verwijdering van 2 uitbijters  $BF_{incl} = 0.92$

**Tabel 22**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten post-test als afhankelijke variabele, groep 2*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.59	0.43	3.65	5.41
geslacht	0.50	0.50	0.35	0.65	0.53	0.18	0.48	-0.42	1.54
Leeftijd	0.50	0.50	0.32	0.68	0.48	0.13	0.54	-0.93	1.61
Fysiek pesten pre-test	0.50	0.50	0.60	0.40	1.48	0.18	0.19	-0.00	0.52
Empathie pre-test	0.50	0.50	0.32	0.68	0.47	-0.03	0.11	-0.30	0.20
Empathie post-test	0.50	0.50	0.32	0.68	0.48* <sup>1</sup>	0.01	0.11	-0.23	0.26

\* bij prior = 0.177,  $BF_{incl} = 0.69$ ; bij prior = 0.708,  $BF_{incl} = 0.25$   
 1. Bij verwijdering van 1 uitbijter  $BF_{incl} = 0.54$

**Tabel 23**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten follow-up test als afhankelijke variabele, groep 2*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddeld e	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.32	0.35	3.57	5.03
Geslacht	0.50	0.50	0.43	0.57	0.75	-0.08	0.38	-0.92	0.89
Leeftijd	0.50	0.50	0.47	0.53	0.90	0.21	0.58	-0.64	1.70
Fysiek pesten pre-test	0.50	0.50	0.58	0.42	1.40	0.11	0.13	-0.01	0.41
Empathie pre-test	0.50	0.50	0.65	0.35	1.87	-0.18	0.18	-0.52	1.819×10 <sup>-5</sup>
Empathie post-test	0.50	0.50	0.60	0.40	1.48	0.11	0.14	-0.07	0.42
Empathie follow-up test	0.50	0.50	0.55	0.46	1.20* <sup>1</sup>	-0.07	0.10	-0.29	0.05

\* bij prior = 0.177,  $BF_{incl} = 1.29$ ; bij prior = 0.708,  $BF_{incl} = 0.81$   
 1. Bij verwijdering van 2 uitbijters  $BF_{incl} = 0.92$

**Tabel 24**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten post-test als afhankelijke variabele, controlegroep*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	5.36	0.32	4.77	6.05
Geslacht	0.50	0.50	0.34	0.66	0.52	0.12	0.41	-0.50	1.31
Leeftijd	0.50	0.50	0.33	0.67	0.49	0.14	0.51	-0.91	1.55
Verbaal pesten pre-test	0.50	0.50	0.31	0.69	0.44	-0.01	0.11	-0.26	0.32
Empathie pre-test	0.50	0.50	0.31	0.69	0.45	-0.01	0.04	-0.14	0.07
Empathie post-test	0.50	0.50	0.67	0.33	2.07* <sup>1</sup>	0.12	0.11	0.00	0.32

\* bij prior = 0.177,  $BF_{incl} = 1.92$ ; bij prior = 0.708,  $BF_{incl} = 1.66$

1. Bij verwijdering van 4 uitbijters  $BF_{incl} = 5.78$

**Tabel 25**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met verbaal pesten follow-up test als afhankelijke variabele, controlegroep*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	5.18	0.28	4.58	5.77
Geslacht	0.50	0.50	0.35	0.66	0.53	0.04	0.35	-0.81	1.07
Leeftijd	0.50	0.50	0.35	0.65	0.53	-0.12	0.47	-1.57	0.73
Verbaal pesten pre-test	0.50	0.50	0.33	0.67	0.50	-0.00	0.11	-0.28	0.26
Empathie pre-test	0.50	0.50	0.48	0.53	0.91	0.04	0.06	-0.04	0.20
Empathie post-test	0.50	0.50	0.39	0.61	0.65	-0.03	0.06	-0.21	0.07
Empathie follow-up test	0.50	0.50	0.84	0.16	5.16* <sup>1</sup>	0.15	0.09	0.00	0.29

\* bij prior = 0.177,  $BF_{incl} = 3.91$ ; bij prior = 0.708,  $BF_{incl} = 4.88$

1. Bij verwijdering van 2 uitbijters  $BF_{incl} = 26.72$

**Tabel 26**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten post-test als afhankelijke variabele, controlegroep*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.41	0.40	3.51	5.15
Geslacht	0.50	0.50	0.40	0.61	0.65	0.04	0.51	-1.07	1.43
Leeftijd	0.50	0.50	0.52	0.48	1.10	0.73	1.00	-0.18	3.36
Fysiek pesten pre-test	0.50	0.50	0.46	0.54	0.86	-0.11	0.21	-0.62	0.15
Empathie pre-test	0.50	0.50	0.45	0.55	0.83	-0.04	0.07	-0.21	0.07
Empathie post-test	0.50	0.50	0.62	0.38	1.62* <sup>1</sup>	0.12	0.13	-0.01	0.40

\* bij prior = 0.177,  $BF_{incl} = 1.64$ ; bij prior = 0.708,  $BF_{incl} = 1.11$   
1. Bij verwijdering van 3 uitbijters  $BF_{incl} = 1.19$

**Tabel 27**

*Posterior samenvatting van de coëfficiënten van de Bayesiaanse lineaire regressie met fysiek pesten follow-up test als afhankelijke variabele, controlegroep*

Coëfficiënt	P(incl)	P(excl)	P(incl data)	P(excl data)	BF <sub>inclusief</sub>	Gemiddelde	SD	95% Geloofwaardigheids Interval	
								Lager	Hoger
Intercept	1.00	0.00	1.00	0.00	1.00	4.32	0.31	3.60	4.94
Geslacht	0.50	0.50	0.81	0.19	4.36	1.31	0.88	0.00	2.77
Leeftijd	0.50	0.50	0.90	0.10	8.64	-2.42	1.17	-4.14	0.00
Fysiek pesten pre-test	0.50	0.50	0.55	0.45	1.24	-0.10	0.19	-0.62	0.20
Empathie pre-test	0.50	0.50	0.51	0.49	1.05	0.01	0.06	-0.14	0.16
Empathie post-test	0.50	0.50	0.74	0.26	2.91	-0.16	0.13	-0.41	0.01
Empathie follow-up test	0.50	0.50	0.50	0.50	0.99* <sup>1</sup>	0.02	0.05	-0.13	0.13

\* bij prior = 0.177,  $BF_{incl} = 1.11$ ; bij prior = 0.708,  $BF_{incl} = 0.72$   
1. Bij verwijdering van 4 uitbijters  $BF_{incl} = 0.86$

## Bijlage D. Overzichten van significante en niet-significante resultaten en overtuigende en niet-overtuigende Bayesfactoren

**Tabel 1**

*Aantal (percentage) significante en niet-significante resultaten van de RM-ANOVA's en lineaire regressies met klassieke statistiek van het onderzoek van Palade en Pascal (2023) en van het huidige onderzoek*

	Test	Aantal significante resultaten	Aantal niet- significante resultaten	Totaal
Palade en Pascal	RM-ANOVA	7 (23%)	11 (37%)	18 (60%)
	Regressies	2 (7%)	10 (33%)	12 (40%)
	Totaal	9 (30%)	21 (70%)	30 (100%)
Huidige onderzoek	RM-ANOVA	6 (20%)	12 (40%)	18 (60%)
	Regressies	0 (0%)	12 (40%)	12 (40%)
	Totaal	6 (20%)	24 (80%)	30 (100%)

**Tabel 2**

*Aantal resultaten van de RM-ANOVA's en regressies van het onderzoek van Palade en Pascal (2023) en het huidige onderzoek met significantie en met overtuigende Bayesfactor voor de alternatieve hypothese*

	Test	Aantal significante resultaten	Aantal overtuigende Bayesfactoren voor alternatieve hypothese
Palade en Pascal	RM-ANOVA	7	6
	Regressie	2	1
Huidig onderzoek	RM-ANOVA	6	6
	Regressie	0	1

**Tabel 3**

*Aantal resultaten van de RM-ANOVA's en regressies van het onderzoek van Palade en Pascal (2023) en het huidige onderzoek met niet-significantie en met overtuigende Bayesfactor voor de alternatieve hypothese*

	Test	Aantal niet- significante resultaten	Aantal overtuigende Bayesfactoren voor nulhypothese
Palade en Pascal	RM-ANOVA	11	8
	Regressie	10	0
Huidig onderzoek	RM-ANOVA	12	8
	Regressie	12	0

**Tabel 4**

*Aantal (percentage) resultaten met een overtuigende en niet-overtuigende Bayesfactor ten gunste van de alternatieve en de nulhypothese van de RM-ANOVA's en lineaire regressies van het huidige onderzoek*

		<b>Aantal resultaten pro alternatieve hypothese</b>	<b>Aantal resultaten pro nulhypothese</b>	<b>Totaal</b>
Overtuigende	RM-ANOVA	6 (20%)	8 (27%)	14 (47%)
Bayesfactor	Regressie	1 (3%)	0 (0%)	1 (3%)
$BF > 3.00$	Subtotaal	7 (23%)	8 (27%)	15 (50%)
Niet-	RM-ANOVA	2 (7%)	2 (7%)	4 (14%)
overtuigende	Regressie	1 (3%)	10 (33%)	11 (36%)
Bayesfactor	Subtotaal	3 (10%)	12 (40%)	15 (50%)
$BF = 0.33-3.00$				
	Totaal	12 (40%)	18 (60%)	30 (100%)