

*Kidd & Castano (2013) claimden in Science dat het lezen van literaire fictie het empathisch vermogen van de lezer verbetert. Na publicatie trok het artikel direct veel aandacht. Zo stelde The New York Times Well-blog dat sociale wetenschappers het lezen van Tjechov aanraden om betere sociale vaardigheden te ontwikkelen. Maar is het geen gewaagde stelling dat het lezen van een literair verhaal leidt tot meer empathie? Aan de Erasmus Universiteit werd het onderzoek van Kidd & Castano gerepliceerd.*

## EEN REPLICATIESTUDIE

LEIDT HET LEZEN VAN  
LITERAIRE FICTIE TOT  
MEER EMPATHIE?

## THEORY OF MIND

Hoe kan het lezen van literatuur leiden tot een groter inlevingsvermogen in het dagelijks leven? Een technische term voor inlevingsvermogen is *Theory of Mind* (ToM). ToM is de vaardigheid om een mentale staat, intenties, kennis en verwachtingen toe te kennen aan jezelf en anderen, en je ook in mentale gesteldheden van anderen te verplaatsen (Wellman, Cross & Watson, 2001). ToM omvat cognitieve en affectieve componenten (Shamay-Tsoory & Aharon-Peretz, 2007). Bij *cognitieve* ToM kunnen mensen overtuigingen en intenties van anderen interpreteren, terwijl mensen zich bij *affectieve* ToM een mentaal beeld kunnen vormen van emoties van anderen. Affectieve ToM is gerelateerd aan empathie. Maar anders dan empathie gaat het bij affectieve ToM om de cognitieve representatie van de emotie van een ander en niet om het opwekken van de emotie zelf (Shamay-Tsoory & Aharon-Peretz, 2007).

Deze twee aspecten van ToM zijn belangrijk; beide spelen een rol in zowel de persoonlijke ontwikkeling van een individu als bij diens interacties met anderen. Allereerst is ToM belangrijk in de cognitieve en emotionele ontwikkeling van kinderen (De Villiers & De Villiers, 2014). Peuters beginnen percepties en emoties te begrijpen en gaan zichzelf als separaat van anderen zien. Vanaf de schoolgaande leeftijd ontwikkelen zij het vermogen om kennis, wensen en emoties van anderen te begrijpen. De ontwikkeling van kinderen omvat dus enerzijds de vorming van hun eigen gedachtegoed en het begrijpen van hun eigen emoties, en anderzijds de ontwikkeling van het vermogen om zich emotioneel en cognitief in anderen te verplaatsen. De taalontwikkeling vormt hier een essentieel onderdeel van

(De Villiers & De Villiers, 2014). Kinderen gebruiken taal, bijvoorbeeld woorden zoals 'willen', 'begrijpen' en 'voelen', om hun mentale toestand en wensen uit te drukken. Daarnaast is ToM nodig om empathie te voelen voor iemand anders en is zo dus cruciaal in sociale en interpersoonlijke interacties. Als een persoon goed in staat is mee te leven met een ander en vice versa, zal dit wederzijds begrip de onderlinge band versterken.

De originele achterliggende gedachte van het onderzoek van Kidd & Castano (2013) in *Science* was dat het lezen van literaire fictie – in dit onderzoek verhalen die literaire prijzen hebben behaald of verhalen van auteurs die tot de literaire canon behoren, bijvoorbeeld Anton Tsjechov – in tegenstelling tot populaire fictie – in dit onderzoek gedefinieerd als deel uitmakend van een anthologie van populaire fictie, bijvoorbeeld een verhaal van Dashiell Hammett – van invloed is op ToM. Deze gedachte kwam voort uit eerder onderzoek dat aantoonde dat het lezen van literatuur een empathie-verhogend effect zou hebben op lezers (Djikic & Oatley, 2014). ToM zou namelijk de denkwijze reflecteren die door het lezen van literaire fictie wordt opgeroepen – zoals in verhalen over personages waarin hun intenties, waarachtigheid, emoties en interacties met anderen naar voren komen. Als iemand een verhaal leest, neemt hij of zij dus een andere leeswijze aan dan wanneer iemand non-fictie leest.

Daarnaast is de vraag of de opeenvolging van gebeurtenissen logisch en waar is, geheel ondergeschikt aan de beleving van de verhaalwereld. De lezer identificeert zich met een personage en zal daarom empathie ondervinden ten aanzien van diens gevoelens (Keen, 2006). Bovendien

## LEIDT HET LEZEN VAN LITERAIRE FICTIE TOT MEER EMPATHIE?

KATINKA DIJKSTRA, PETER VERKOEIJEN,  
IRIS VAN KUIJK, SOKE YEE CHOW, ANNELIES BAKKER  
& ROLF ZWAAN

fungeert het lezen van fictie als een simulatie van sociale ervaringen waarin mensen representaties van het gedachtengoed van anderen maken en als het ware het gedrag in een complexe wereld simuleren (Mar, 2011). Een empirisch onderzoek heeft dit verband tussen het lezen van fictie en ToM gelegd door aan te tonen dat de hersengebieden, zoals de mediale prefrontale cortex die met ToM geassocieerd worden, deels overlappen met hersengebieden die geassocieerd worden met het begrijpen van verhalen (Mar, 2011).

Het effect van het lezen van literaire fictie op ToM zou zijn dat het empathie kan verhogen, alsook ons vermogen om ons zelf in anderen te verplaatsen. Zo zou het lezen over Gustave Flauberts personage Madame Bovary, uit de gelijknamige klassieker van 1856, die uit onvrede met haar sociale status tot gedrag wordt gedreven dat haar uiteindelijk fataal wordt, kunnen leiden tot een beter inzicht in hoe statusarme personen in het algemeen tot dergelijke wanhoopsdaden kunnen komen. De simulatie van de wereld en gebeurtenissen beschreven in de roman die door de ogen van de protagonist tot stand komt, zou kunnen bijdragen aan een verhoging van ToM. Onderzoek heeft laten zien dat lezers die in het verhaal opgingen, ook wel *transportation* genoemd, grotere empathie lieten zien met de protagonist van het verhaal (Johnson, 2012).

Het lezen van populaire fictie daarentegen heeft volgens Kidd en Castano (2013) geen effect op ToM – dit omdat in populaire fictie de verhaallijn consistent en voorspelbaar is en daardoor geen beroep doet op ToM (Gerrig & Rapp, 2004). De lezer wordt dus niet uitgedaagd wanneer populaire fictie wordt gelezen zoals dat bij het lezen van literaire fictie het geval is; de verwachtingen van de lezer worden bij populaire fictie echter voortdurend bevestigd. Het doel van populaire fictie is ook niet het uitdagen van de lezer, bijvoorbeeld door middel van stijlmiddele, maar het prettig onderhouden van de lezer op een meer passieve manier dan bij literaire fictie gebeurt (Barthes, 1974).

Om de rol van literaire fictie op ToM nader te onderzoeken,

voerden Kidd en Castano (2013) een vijftal experimenten uit met behulp van het online platform Mechanical Turk. Dit platform is een internetmarktplaats die onder meer onderzoekers en proefpersonen bij elkaar brengt, waardoor in betrekkelijk korte tijd onderzoekers proefpersonen tegen betaling taken kunnen laten uitvoeren.

In deze experimenten lazen deelnemers voordat ze een ToM-taak uitvoerden 1) literaire teksten, 2) populaire fictie, 3) non-fictie, 4) of geen tekst. Afgezien van de ToM-taken werden ook andere cognitieve en affectieve taken aangeboden. De resultaten van de vijf experimenten lieten stelselmatig zien dat het lezen van literaire teksten tot betere prestaties leidde op cognitieve en affectieve ToM-taken vergeleken met het lezen van populaire fictie of niet lezen.

In Tabel 1 staat een overzicht van de resultaten van Kidd en Castano. Zij vergeleken een literaire fictieconditie (LF) met een controleconditie (C; dit was non-fictie in experiment 1, en populaire fictie in de overige experimenten) op gemiddelde affectieve ToM-score. Een hogere score in kolom 3 voor de affectieve ToM-maten correspondeert met een hogere mate van affectieve ToM (behalve voor de *Diagnostic Analysis of Non-verbal Accuracy 2- Adult Faces* (DANVA2-AF) in experiment 2 waarin het aantal fouten geteld wordt en een lagere score beter is; zie Appendix 1 voor uitleg over deze ToM).

De herhaling van de effecten in meerdere experimenten resulteerde in de conclusie dat het lezen van literaire fictie (in ieder geval tijdelijk) ToM zou vergroten. Gezien de herhaaldelijke replicatie van de bevindingen in het artikel zelf die wijzen op een robuust wetenschappelijk effect en mogelijke implicaties van de bevindingen, lijkt publicatie in een gerenommeerd tijdschrift als *Science* gerechtvaardigd. Men zou zomaar ToM bij willekeurige lezers kunnen verhogen door ze met enige regelmaat aan literatuur bloot te stellen.

Maar is dit echt zo? De resultaten roepen de nodige vragen op. Allereerst: waarom heeft zo'n relatief eenvoudige

TABEL 1. OVERZICHT RESULTATEN KIDD &amp; CASTANO (2013)

	N <sub>LF</sub>	N <sub>C</sub>	M <sub>LF</sub> - M <sub>C</sub>	POOLED SD	COHEN'S D	P (2-ZIJDIG)	GEOBSERVEERDE POWER
Exp. 1 (RMET)	43	43	2.43	4.79	0.51	.021	.63
Exp. 2 (DANVA2-AF)	57	57	-1.15	2.64	0.46	.022	.64
Exp. 3 (RMET)	35	34	2.70	5.21	0.52	.035	.58
Exp. 4 (RMET)	36	36	2.48	5.26	0.47	.049	.52
Exp. 5 (RMET)	119	118	1.25	4.12	0.30	.021	.65

en beperkte manipulatie als het lezen van een kort literair verhaal zulke ingrijpende consequenties? Een literatuurwetenschapper kan beargumenteren dat literaire verhalen en romans onderling nogal verschillen, waaronder de mate waarin een beroep wordt gedaan op de actieve rol van de lezer en de mate waarin simulatie van het vertelde überhaupt mogelijk is. De fragmentarische tekst in James Joyce's *Finnegans Wake*, ook wel omschreven als een kruiswoordpuzzel, zal zich veel minder lenen tot simulatie dan het eerder genoemde *Madame Bovary*. Zijn de door de auteurs gekozen fragmenten wel representatief voor de heterogene genres die ze vertegenwoordigen?

Daarbij is het door de onderzoekers gemaakte onderscheid tussen 'literaire fictie' en 'populaire fictie' in de vijf experimenten niet eenduidig. In hun tweede experiment werden fragmenten van finalisten voor de *National Book Award for Fiction* gebruikt als voorbeelden van literaire fictie, terwijl de voorbeelden van populaire fictie van fragmenten van recente bestsellers op *Amazon* afkomstig waren. In de experimenten drie, vier en vijf daarentegen waren verhalen uit een anthologie van populaire fictie, respectievelijk verhalen uit de collectie winnaars van de PEN/O-prijs geselecteerd (onderscheiding van buitengewoon verdienstelijke korte verhalen), en werden enkele verhalen uit experiment drie in experiment vier en vijf vervangen. De tweedeling in literaire en populaire fictie in deze laat ruimte open voor discussie. Zijn bestsellers op *Amazon* populaire fictiewerken alleen omdat ze goed verkopen? Zijn verdienstelijke korte verhalen juist literaire verhalen omdat ze onderscheiden zijn met de PEN/O prijs? Deze tweedeling lijkt veel minder te berusten op de inhoudelijke genreverschillen en hun effecten op de lezers die door Kidd en Castano worden beargumenteerd.

Als tegenargument kunnen de auteurs aanvoeren dat hun manipulatie erg robuust is aangezien zij herhaaldelijk en met verschillende verhalen een significant effect hebben gevonden. In deze robuustheid kan echter een probleem schuilen. Om dit te verduidelijken kijken we naar de vijfde kolom in tabel 1 met effectgroottes (Cohens  $d$ ). De effecten van de manipulatie zijn middelgroot ( $d=0.5$ ) tot klein ( $d=0.3$ ). Aan de hand van de effectgrootte kan vervolgens bepaald worden hoeveel deelnemers er getest hadden moeten worden voor een vaak gehanteerde *power* (onderscheidingsvermogen) van .80 bij een tweezijdige *t*-toets en een significantieniveau van .05. Bij experiment 1 komt dat neer op minimaal 61 deelnemers per conditie (dus minimaal 122 deelnemers in totaal). Omdat de effectgroottes van de overige

vier experimenten allemaal nagenoeg hetzelfde of kleiner zijn dan in experiment 1, laat deze steekproefgrootteberekening zien dat vier van de vijf experimenten te weinig deelnemers en dus te weinig *power* hadden. Toch werd in alle vier de experimenten een significant effect gevonden (zie de voorlaatste kolom van tabel 1). En dat is opmerkelijk gegeven de waarschijnlijk relatief lage *power* van het merendeel van de experimenten. Hoe opmerkelijk precies kan vastgesteld worden met de *consistency-test* van Francis (2013). Deze methode zullen wij op de vergelijkingen uit tabel 1 toepassen.

Allereerst wordt voor elk experiment de zogenaamde geobserveerde *power* berekend, gebaseerd op de gevonden effectgrootte en het aantal deelnemers per conditie. Hiervoor hebben wij gebruik gemaakt van de rekenmachine op <http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html>. Omdat deze rekenmachine uitgaat van een gelijk aantal deelnemers per conditie, hebben we bij ongelijkheid gebruik gemaakt van het aantal deelnemers in de grootste conditie. De geobserveerde *power* is voor ieder experiment weergegeven in de laatste kolom van tabel 1. De kans op vijf keer een significant resultaat in vijf experimenten wordt dan  $.63 \times .64 \times .58 \times .52 \times .65 = .079$ . Deze kans is de uitkomst van de consistentietest en de gevonden waarde is kleiner dan de grenswaarde van .10 die Francis (2013) voorstelt. Bij een testresultaat onder deze waarde zijn er te veel significante resultaten in een set van experimenten, gegeven de geobserveerde *power* van elk van de experimenten; de resultaten zijn dan *too good to be true*. Francis geeft aan dat zo'n set het gevolg kan zijn van een selectieve rapportage van resultaten (alleen de 'successen' worden vermeld) en/of van incorrect uitgevoerde experimenten.

Dit laatste kan terug te voeren zijn op *questionable research practices* (Simmons, Nelson & Simonsohn, 2011; De Groot (1961) beschrijft ze reeds in *Methodologie*), zoals stoppen met de dataverzameling op het moment dat het verwachte resultaat significant is. Wat in de experimenten van Kidd en Castano (2013) geleid heeft tot de eerder vermelde lage waarde op de consistentietest, blijft giswerk. Het is evenwel aannemelijk dat Kidd en Castano een te optimistisch beeld geven van het effect van het lezen van literaire fictie op affectieve ToM. De vraag die vervolgens rijst is hoe groot dit effect dan werkelijk is, als het effect al bestaat.

Deze twee vragen – waarom heeft zo'n eenvoudige manipulatie zulke verstrekkende consequenties én hoe groot is het effect – vormden ons uitgangspunt voor het uitvoeren van

deze replicatiestudie. Bij een replicatieonderzoek doet een wetenschapper een reeds bestaand onderzoek zo nauwkeurig mogelijk na. Replicatiestudies zijn belangrijk omdat ze kunnen laten zien of een effect betrouwbaar is of niet (Lakens, Haans & Koole, 2012). Geeft de replicatie hetzelfde resultaat, dan beperkt een bevinding zich niet alleen tot de oorspronkelijke specifieke context.

Aangezien de keuzes voor design, proefpersonen en manipulatie door de oorspronkelijke studie zijn bepaald, kan er niet met marges, manipulaties of steekproefgroottes worden geschoven en voegt een replicatie waardevolle informatie toe aan de betrouwbaarheid van een effect (Lakens et al., 2012). We willen wel opmerken dat denken in termen van succesvolle en niet-succesvolle replicaties naar onze mening weinig vruchtbaar is. Beter is het om op basis van een groot aantal replicatiestudies te komen tot een kwantificering van de effectgrootte die gemoeid is met een bepaalde bevinding en een kwantificering van de mate waarin een bevinding varieert over verschillende onderzoeken heen. In het geval van Kidd en Castano (2013) zouden idealiter alle vijf experimenten moeten worden gerepliceerd. Dat is, alleen al vanwege de grote aantallen deelnemers, praktisch niet haalbaar. Voor onze studie is daarom voor de replicatie van alleen experiment 5 gekozen. Dit was de studie met de meeste power, waarin bovendien de meeste relevante condities waren opgenomen.

### DE OORSPRONKELIJKE STUDIE

In experiment vijf van Kidd en Castano lazen deelnemers één van de volgende populaire fictieverhalen *Jane* van Mary Roberts Rinehart, *Too Many Have Lived* van Dashiell Hammett, en *Space Jockey* van Robert Heinlein afkomstig uit de anthologie van populaire fictie (Hoppenstand, 1998); één van de volgende literaire fictie verhalen *The VanderCook* van Alice Mattison; *Corrie* van Alice Munro; en *Uncle Rock* van Dageroberto Gilb, die een Nationale Boekenprijs hadden gewonnen; of geen verhaal. De ToM-taken waren de *Reading the Mind in the Eyes test* (RMET) en de *Yoni-taak* (zie appendix 1).

In het oorspronkelijke experiment werden 356 deelnemers (na exclusie van 100 deelnemers) in de leeftijd van 18 tot 75 ( $M=34.42$ ,  $sd=11.59$ ) willekeurig aan de literaire fictie-, populaire fictie-, of niet-leesconditie toegewezen. De deelnemers vulden naast de gebruikelijke tests de *Transportation Scale* (Green & Brock, 2000) in, waardeerden de tekst in de mate waarin zij de tekst een goed voorbeeld vonden van een literaire tekst en voerden de *Author Recognition Test* (ART; zie appendix 1). ToM wordt namelijk gerelateerd aan het

bekend zijn met auteurs van fictie (Mar, Oatley & Peterson, 2009).

De resultaten lieten zien dat de literaire teksten minder gewaardeerd werden dan de populaire fictieteksten, maar ze werden wel als betere voorbeelden van literatuur gezien. De mate van *transportation* verschilde niet tussen de twee typen teksten – een verrassend resultaat omdat Kidd en Castano veronderstellen dat mensen die literaire fictie lezen *sterker* aangezet worden tot verplaatsing in de hoofdpersonen dan mensen die populaire fictie lezen. De resultaten op de verschillende ToM-maten (RMET en Yoni) waren evenwel in lijn met de voorspellingen van Kidd en Castano. Door de bank genomen lieten de resultaten in alle experimenten zien dat deelnemers hogere ToM-scores behaalden na het lezen van literaire fictie dan na het lezen van populaire fictie (zie tabel 1).

### DE REPLICATIE

In de oorspronkelijke studie was naast een literaire- en populaire-fictieconditie een niet-lezenconditie opgenomen. Deze conditie is in onze replicatiestudie buiten beschouwing gelaten, omdat we de claim over het effect van het lezen van literaire fictie maar niet van populaire fictie op het verhogen van ToM als meest boude claim zagen.

Voor de replicatiestudie zijn dezelfde exclusiecriteria gehanteerd als in het originele experiment: deelnemers die al aan een eerder experiment hadden meegedaan en voor wie Engels niet de moedertaal was, werden uitgesloten van deelname. Teksten waren gereduceerd tot een standaard-lengte per pagina waardoor extreem korte leestijden (minder dan dertig seconden per pagina) uitgesloten konden worden. Deelnemers met leestijden boven 3.5 standaarddeviaties van het gemiddelde werden eveneens uitgesloten. Andere exclusiecriteria waren uitbijters (3.5  $sd$  van het gemiddelde) op de *Author Recognition Test* of afhankelijke variabelen.

### METHODE

**DEELNEMERS** Net als in de oorspronkelijke studie van Kidd en Castano deden in onze replicatiestudie 528 Amerikaanse deelnemers mee aan het onderzoek dat met het eerder beschreven Mechanical Turk-platform werd uitgevoerd. Deelnemers ontvingen een vergoeding van \$1.75 voor een taak die naar verwachting rond de vijftig minuten in beslag zou nemen. Deelnemers werden willekeurig toegewezen aan de twee condities. Deelnemers die in meerdere versies van het

experiment meededen, vielen af ( $n = 64$ ), evenals deelnemers die het experiment niet afmaakten ( $n = 48$ ) en *non-native English speakers* ( $n = 10$ ). Verder vielen uitbijters (3.5 sd boven of onder het gemiddelde) af voor de ART ( $n = 1$ ), RMET ( $n = 4$ ), en trials van de Yoni-taak ( $n = 10$ ). Ten slotte werden de data van deelnemers verwijderd die met leestijden meer dan 3.5 sd onder het gemiddelde zaten ( $n = 90$ ) en een proefpersoon met een extreem lange leestijd ( $n = 1$ ).

De uiteindelijke steekproef bestond uit 300 deelnemers: 150 in de populaire-fictieconditie en 150 in de literaire-fictieconditie. Het percentage vrouwen bedroeg 65% in de populaire-fictieconditie en 69% in de literaire-fictieconditie. De gemiddelde leeftijd in de populaire-fictieconditie was 39.84 jaar (sd = 13.47) en 37.08 (sd = 12.80; gebaseerd op 149 deelnemers) in de literaire-fictieconditie. De deelnemers van de twee condities verschilden niet qua etniciteit en hoogst genoten opleiding. De twee condities waren dus vergelijkbaar voor aanvang van de experimentele procedure.

**TAKEN** Vergelijkbaar met de oorspronkelijke studie was de verwachting dat de scores op de volgende drie taken beïnvloed worden door het lezen van literaire versus populaire teksten: de RMET, de Yoni-test en de ART. Voor een verdere uitleg over deze taken verwijzen we naar Appendix 1.

Verder werden gegevens verzameld ten aanzien van positief en negatief affect die op dit moment ervaren werden (Watson, Clark & Tellegen, 1988), of deelnemers blijdschap dan wel verdriet voelden, en de *Transportation Scale* die bevroegde in hoeverre de lezer opging in het verhaal. Hierna gaven de deelnemers aan in hoeverre ze het verhaal leuk vonden en in hoeverre ze dachten dat het gelezen verhaal een voorbeeld was van goede literatuur. Dit konden zij aangeven op een 7-puntsschaal (van 1 = *not at all* tot 7 = *extremely*).

Tenslotte werd naar demografische gegevens (leeftijd, geslacht, opleidingsniveau) gevraagd en naar wat zij dachten waarover het onderzoek ging met de *Perceived Awareness of the Research Hypothesis* (PARH; Rubin, Paolini, & Crisp, 2010).

**PROCEDURE** Net als in de oorspronkelijke studie waren er zes verschillende versies van het experiment. Elke versie bestond uit één verhaal uit de selectie van zes verhalen (populaire of literaire fictie) en de testen die elke proefpersoon maakte. Elke versie bevatte een ander verhaal. Deelnemers mochten uitsluitend aan een van de zes versies meedoen (hun data werden verwijderd

van hun tweede deelname als ze toch aan een tweede versie deelnamen). De volgorde van taken was voor alle deelnemers hetzelfde. Allereerst moesten zij de tekst lezen die voor hun versie geselecteerd was. Vervolgens maakten zij de taken in de volgorde zoals hieronder is beschreven.

**DESIGN** Het design was een between-subjectsdesign met conditie (literaire of populaire fictie) als between-subjects variabele. Dit houdt in dat een groep deelnemers een literair verhaal te lezen kreeg en een andere groep deelnemers een populair fictieverhaal (zie kader Verhalen). De verschillende maten van de ToM-taken waren de afhankelijke variabelen. De rol van het type verhaal (literair of populair) op de scores van de ToM-taken was wat werd onderzocht. Andere maten werden als covariaat (een variabele die het effect kan beïnvloeden van de relatie tussen een onafhankelijke en afhankelijke variabele) meegenomen, zoals de ART en de score op de controletasks van de Yoni-taak.

## RESULTATEN

In Tabel 2 staan de gemiddelden en standaarddeviaties op relevante oordelen over de verhalen. Op de vraag 'hoeveel plezier heb je beleefd aan het lezen van de tekst?' werd in de literaire-fictieconditie gemiddeld iets hoger gescoord dan in de populaire-fictieconditie ( $F(1,298) = 6.076, p = .014, \eta^2 = .019$ ). Daarnaast werd op de vraag 'in hoeverre is de gelezen tekst een goed voorbeeld van literatuur?' in de literaire-fictieconditie gemiddeld genomen iets hoger gescoord dan in de populaire fictie conditie ( $F(1,298) = 7.728, p = .006, \eta^2 = .025$ ). Voor beide vragen geldt dat er weliswaar een significant verschil is tussen beide groepen, maar dat het effect van het type literatuur erg klein is.

Verder was de gemiddelde gerapporteerde mate van

**TABEL 2. GEMIDDELDEN EN STANDAARDDEVIATIES VOOR OORDELEN VAN STUDENTEN OVER DE VERHALEN EN TESTSCORES**

VRAAG	LITERAIRE FICTIE	POPULAIRE FICTIE
	M (sd)	M (sd)
Hoeveel plezier beleefd (0-7)	4.63 (1.19)	4.10 (1.84)
Goed voorbeeld van literatuur (0-7)	4.00 (1.69)	3.41 (1.73)
Transportatie	4.01 (0.69)	3.68 (0.73)
Author Recognition Test	21.63 (13.27)	22.59 (13.91)
Yoni-controletaak	.97 (.06)	.97 (.07)



transportatiescore hoger in de literaire-fictieconditie dan in de populaire-fictieconditie ( $F(1,298)=15.257, p<.01, \eta^2=.048$ ); ook dit effect is erg klein. De gemiddelde score op de ART verschilde niet tussen de literaire- en de populaire-fictieconditie ( $F < 1, \eta^2<.01$ ). De gemiddelde score (proportie correct beantwoorde items) op de Yoni-controletaak was vergelijkbaar voor de literaire-fictieconditie ( $M=.97, sd=.06$ ) en de populaire conditie ( $M=.97, sd=.07$ ); in beide condities lijkt er bovendien sprake te zijn van een plafondeffect.

In Tabel 3 staan de gemiddelden en standaarddeviaties op de belangrijkste uitkomstmaten per verhaal en per conditie. Bekijken we de resultaten uit die tabel per kolom, dan wordt duidelijk dat de zes verschillende verhalen elkaar nauwelijks ontlopen in termen van gemiddelde scores op de uitkomstmaat en in standaarddeviaties. Daarnaast blijkt uit de 'totaal'scores per uitkomstmaat dat de deelnemers in de literaire conditie en in de populaire-fictieconditie zeer vergelijkbaar presteren. Ten slotte laten de scores op de twee eerste orde Yoni-taken – en in mindere mate de scores op de tweede orde affectieve Yoni-taak – per verhaal en in de totalen per conditie een plafondeffect zien. Alles overziend, is in de beschrijvende statistieken van onze steekproef geen evidentie te vinden voor de hypothese dat mensen beter scoren op ToM-taken na het lezen van literaire fictie dan na het lezen van populaire fictie.

In het oorspronkelijke onderzoek werden toetsende analyses uitgevoerd op de rMET-scores en de Yoni-scores. Wij

hebben voor wat betreft de rMET-replicatiedata dezelfde analyses uitgevoerd als in de originele studie. De analyse van de Yoni-scores verschilde echter enigszins van de analyse uit de originele studie omdat een belangrijke assumptie geschonden was. Dezelfde assumptie was overigens ook geschonden in de originele studie, maar hier hebben Kidd en Castano geen rekening mee gehouden.

Onderstaand volgen de belangrijkste uitkomsten van de toetsende statistieken (zie ook Appendix 2 voor meer details). Het effect van type tekst op rMET-scores was extreem klein met een partiële correlatie van .071. Het gemiddelde was 27.233 [95% Betrouwbaarheidsinterval 26.581; 27.885] in de populaire-fictieconditie en 27.808 [95% Betrouwbaarheidsinterval 27.156; 28.460] in de literaire-fictieconditie. De eerste toetsende analyse laat dus – in tegenstelling tot de bevindingen uit de originele studie – slechts een heel klein en statistisch niet-significant effect zien van type tekst op rMET-scores.

Een vergelijkbaar resultaat werd gevonden voor de Yoni-taken: het gemiddelde in de populaire-fictieconditie ( $M=.927, 95\%$  Betrouwbaarheidsinterval .918; .936) was vergelijkbaar met dat in de literaire-fictieconditie ( $M=.933, 95\%$  Betrouwbaarheidsinterval .924; .942],  $F < 1, p=.329$ , partial  $\eta^2=.003$ ). De resultaten van deze analyse lijken dus – evenals de resultaten van de rMET-analyses – niet in lijn met de bevindingen van de originele studie.

## REFLECTIE

**M**aakt het lezen van literatuur ons meer empathisch? De meesten van ons zouden waarschijnlijk willen dat dit zo was. Niet alleen heeft het lezen van literatuur (vergeleken met andere fictie) dan ineens een praktisch nut, we hebben ook een gemakkelijke manier gevonden om een betere samenleving te creëren. We schotelen de mensen gewoon een dieet van *Der Zauberberg*, *The Sound and the Fury* en *Terug tot Ina Damman* voor, en hopla, we vallen elkaar zingend in de armen. Misschien heeft deze roze wolk invloed gehad op het besluit van

de reviewers en redacteurs bij *Science* om het artikel van Kidd en Castano te accepteren. Het artikel heeft het nodige stof doen opwaaien in de pers en het vooruitzicht hierop heeft waarschijnlijk ook een rol gespeeld bij die acceptatie van het manuscript.

Want laten we wel wezen: als normale reviewcriteria waren gehanteerd, dan was het artikel niet geaccepteerd. Er kleven nogal wat zwakke punten aan het onderzoek, waarvan er een paar prima worden beschreven in de blog *Language Log* van de Universiteit van Pennsylvania:

<http://bit.ly/1MnQTRe>. Het is duidelijk dat de steekproeven van literaire en andere teksten veel te klein en niet-representatief waren. Zoals de blogpost suggereert, hebben de auteurs teksten gekozen waarmee ze verwachtten de voorspelde effecten te vinden. Zou bijvoorbeeld een fragment uit een experimentele literaire roman, zoals James Joyce's *Finnegan's Wake* of W.F. Hermans' *De God Denkbaar, Denkbaar de God* leiden tot meer empathie?

Daarnaast wordt empathie onmiddellijk na het lezen van de passages gemeten, terwijl de claim is dat het lezen

## DISCUSSIE

Over het algemeen is het beschreven patroon in ons replicatie-experiment anders dan in de oorspronkelijke studie. In de oorspronkelijke studie werd in vijf experimenten aangetoond dat het lezen van literaire fictie de ToM

beïnvloedt. De replicatiestudie liet uitsluitend verschillen zien ten aanzien van de waardering van populaire fictie versus literaire fictie, in welke mate de literaire verhalen goede voorbeelden waren van literatuur en de mate van opgenomen worden in het verhaal die hoger was bij de literaire verhalen.

**TABEL 3. GEMIDDELDEN EN STANDAARDDEVIATIES VOOR DE SCORES OP TWEE YONI-TAKEN (UITGESPLITST NAAR EERSTE ORDE EN TWEEDE ORDE EN VOOR DE SCORES OP DE YONI-TAAK. 'L' VERWIJST NAAR DE LITERAIRE-FICTIE-CONDITIE EN 'P' NAAR DE POPULAIRE-FICTIECONDITIE. DE GEMIDDELDEN EN STANDAARDDEVIATIES VOOR DE VIJF UITKOMSTMATEN WORDEN EVENEENS PER CONDITIE WEERGEGEVEN BIJ RESPECTIEVELIJK TOTAAL L EN TOTAAL P.**

		YONI COGNITIEF		YONI AFFECTIEF	RMET	
		EERSTE ORDE	TWEEDE ORDE	EERSTE ORDE	TWEEDE ORDE	
Verhaal	n	M (sd)	M (sd)	M (sd)	M (sd)	M (sd)
Corrie (L)	50	.99 (.02)	.85 (.16)	.97 (.05)	.92 (.07)	28.54 (4.11)
Uncle Rock (L)	53	.99 (.02)	.85 (.18)	.97 (.06)	.91 (.07)	27.47 (3.76)
The_Vandercook (L)	47	.99 (.04)	.87 (.15)	.96 (.06)	.91 (.08)	27.28 (3.39)
<b>Totaal L</b>	<b>150</b>	<b>.99 (.03)</b>	<b>.86 (.17)</b>	<b>.97 (.06)</b>	<b>.92 (.07)</b>	<b>27.78 (3.79)</b>
Jane (P)	46	.98 (.05)	.83 (.14)	.97 (.06)	.89 (.09)	26.98 (4.38)
Space Jockey (P)	59	.99 (.03)	.85 (.14)	.98 (.04)	.91 (.08)	27.15 (4.04)
Too Many Have Lived (P)	45	.98 (.03)	.84 (.18)	.97 (.04)	.92 (.07)	27.67 (4.10)
<b>Totaal P</b>	<b>150</b>	<b>.99 (.04)</b>	<b>.84 (.15)</b>	<b>.97 (.05)</b>	<b>.91 (.08)</b>	<b>27.23 (4.15)</b>

van literatuur lange termijneffecten heeft. Verder kan men zich afvragen of het überhaupt wel zo plausibel is dat het lezen van literatuur tot andere effecten op empathie leidt dan het lezen van populaire literatuur. In populaire literatuur worden immers ook vaak gedachten en gevoelens beschreven vanuit verschillende perspectieven (denk aan de populaire serie *Het Lied van Ijs en Vuur*, waarop de tv-serie *Game of Thrones* is gebaseerd). En is het onderscheid tussen literatuur en niet-literatuur wel te maken? Hebben generaties literatuurwetenschap-

pers zich hier niet al het hoofd over gebroken? Sommigen hebben zelfs geopperd dat 'literariteit' niet zozeer een kwaliteit van de tekst is, maar een manier van lezen.

De huidige nonreplicatie van de resultaten werpt een verdere schaduw op de conclusies uit het *Science*-artikel. Nu zegt één experiment op zich niet veel. Dat geldt dus zowel voor de originele studie als voor onze non-replicatie. Niettemin laten onze resultaten wel de volgende conclusie toe. Zelfs als men precies dezelfde selectieve steekproeven van teksten

gebruikt als Kidd en Castano en zelfs als men dezelfde methoden gebruikt en dezelfde populatie, is het niet gegeven dat men dezelfde resultaten vindt. Reden temeer dus te concluderen dat *Science* voorbarig heeft gehandeld door deze studie te publiceren.

Sterke claims zoals 'het lezen van literatuur maakt ons betere mensen' vragen om sterk bewijs. En het is inmiddels wel duidelijk dat dit sterke bewijs volkomen afwezig is in het onderzoek van Kidd en Castano. (*Rolf Zwaan*)



Hoe is deze discrepantie mogelijk? Het kan niet aan de grootte van de steekproef hebben gelegen, want er waren meer deelnemers in deze replicatiestudie dan in de oorspronkelijke studie. Het kan evenmin aan de aard van de populatie hebben gelegen omdat wij uit dezelfde populatie (Mturk-workers) hebben geput als Kidd en Castano. Ook de testmaterialen waren identiek aan die van de oorspronkelijke studie. Over de demografische gegevens van de groep waren geen gegevens beschikbaar, dus die vergelijking kan hier niet worden gemaakt. Maar verwacht mag worden dat deze vergelijkbaar is als deze uit dezelfde populatie wordt geworven. Er was wel een verschil in vergoeding voor de deelnemers (\$3 voor de oorspronkelijke studie en \$1.75 voor de replicatiestudie); gezien de overeenkomsten in populatie en gemiddelde scores vormt dit echter geen verklaring voor het nul-effect van de replicatiestudie.

Waren er verschillen tussen de twee studies wat betreft de scores op de tests? Als we de gemiddelden vergelijken (voor zover mogelijk) met die van de oorspronkelijke studie, dan valt op dat de scores voor de *Reading the Mind in the Eyes Test* hoog liggen in de replicatiestudie, hoger (tussen de 27.15 en 28.54) dan die van de oorspronkelijke studie (tussen de 24.18 en 26.97) en duidelijk hoger voor de populaire-fictieconditie dan in de oorspronkelijke studie. De scores voor de Yoni-taak lijken wat hoger te liggen in de replicatiestudie, met uitzondering van de eerste orde maten. Ook al is uit dezelfde populatie van Mturk-workers geput, scores verschillen op belangrijke maten.

Op basis van deze replicatiepoging kan een aantal kanttekeningen bij de oorspronkelijke studie worden gezet. Ten eerste: de hoofdbevindingen worden niet gerepliceerd. Sterker, uit zowel de beschrijvende als toetsende statistieken blijken vrijwel geen verschillen tussen de literaire- en populaire-fictieconditie in de replicatie. Dit kan niet aan verschillen in procedure, materialen of populatie te wijten zijn en dit is dan ook de meest problematische uitkomst met betrekking tot de oorspronkelijke studie. Een effect dat door de auteurs zelf maar liefst vijf keer werd gerepliceerd, was daarentegen in de replicatiepoging geheel afwezig.

Ten tweede: andere effecten uit de oorspronkelijke studie worden niet gerepliceerd. Er was sprake van een plafond-effect voor de Yoni-taak in de replicatiestudie; die was een van de belangrijkste taken van de studie maar kan blijkbaar weinig differentiëren in een steekproef van een algemene populatie. Dit roept de vraag op waarom niet een betere, meer onderscheidende ToM-taak was gebruikt.

Ten derde: het blijft onwaarschijnlijk dat het simpelweg

## Zijn verdienstelijke korte verhalen juist literaire verhalen omdat ze onderscheiden zijn met de PEN/O prijs?

lezen van een literair verhaal in vergelijking met een populair fictieverhaal zulke vergaande consequenties heeft voor het zich kunnen verplaatsen in de emoties en mentale staat van anderen, terwijl het gehanteerde onderscheid tussen deze twee genres zo weinig op eenduidige en inhoudelijke kenmerken is gebaseerd.

Wij concluderen dat onze replicatie van Kidd en Castano (2013) niet in lijn was met hun resultaten en stellen de betrouwbaarheid van het door hen gerapporteerde effect dan ook ter discussie. De resultaten van deze replicatiestudie benadrukken tevens het belang van het uitvoeren van replicatiestudies. Hoe robuust de resultaten van Kidd en Castano in eerste instantie ook leken, uit deze replicatiestudie en de effectanalyses blijkt dat het lezen van literaire fictie niet zonder meer en zelfs niet tijdelijk bijdraagt aan een verbetering van *Theory of Mind*. Verdere replicatiestudies van experimenten uit de oorspronkelijke studie moeten uitwijzen of het effect überhaupt repliceerbaar is.

Hoe dan ook lijkt het nogal voorbarig om mensen aan te zetten tot het lezen van literaire fictie met als doel het inlevingsvermogen in anderen te verhogen. Hoe mooi dit idee in principe ook klinkt, het blijkt in een onafhankelijke meting empirisch niet robuust te zijn.

### OVER DE AUTEURS

Dr. Katinka Dijkstra en dr. Peter Verkoeijen zijn als universitair hoofddocent verbonden aan de Capaciteitsgroep Psychologie, Faculteit Sociale Wetenschappen, Erasmus Universiteit Rotterdam. Annelies Bakker en Soke Yee Chow zijn masterstudent Klinische Psychologie aan de Erasmus Universiteit. Iris van Kuijk volgt de Advanced Research Master bij de Capaciteitsgroep Psychologie van de Erasmus Universiteit Rotterdam. Prof. dr. Rolf A. Zwaan werkt als hoogleraar bij de Capaciteitsgroep Psychologie van de Erasmus Universiteit Rotterdam. Correspondentie aangaande dit artikel: Katinka Dijkstra, Capaciteitsgroep Psychologie, Faculteit Sociale Wetenschappen, Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam. E-mail: k.dijkstra@fsw.eur.nl.

## Summary

DOES READING LITERATURE RESULT IN HIGHER EMPATHY? A REPLICATION STUDY  
K. DIJKSTRA, P. VERKOEIJEN, I. VAN KUIJK, S. YEE CHOW, A. BAKKER & R. ZWAAN

A recently published paper in *Science* (Kidd & Castano, 2013) showed in five experiments that reading literary fiction led to an increase in performance on tests of affective and cognitive Theory of Mind when compared with reading popular fiction, nonfiction or no reading. Although the authors only claim that this is a temporary effect, the claim itself is remarkable and warrants further examination. The current study is a replication study of one the experiments from the *Science* paper to examine how robust these findings are. The results of this study demonstrate that the findings of this original study do not replicate in an independent study with a similar population and the exact same materials and tasks.

## Literatuur

Acheson, D.J., Wells, J.B. & MacDonald, M.C. (2008). New and updated tests of print exposure and reading abilities in college students. *Behavior Research Methods*, 40, 278-289.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry*, 42, 241-251.

Barthes, R. (1974). *S/Z*. New York, NY: Farrar, Straus and Giroux, Inc.

Dijkic, M. & Oatley, K. (2014). The art in fiction: From indirect communication to changes of the self. *Psychology of Aesthetics, Creativity, and the Arts*, 8, 498-505.

Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, 57, 153-169.

Gerrig, R.J. & Rapp, D.N. (2004). Psychological processes underlying literary impact. *Poetics Today*, 25, 265-281.

Green, M.C. & Brock, T.C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79, 701-721.

Groot, A.D. de (1961). *Methodologie. Grondslagen van onderzoek en denken in de gedragswetenschappen*. Assen: Van Gorcum.

Hoppenstand, G. (Ed.; 1998). *Popular fiction: an anthology*. New York, NY: Longman.

Johnson, D.R. (2012). Transportation into a story increases empathy, prosocial behavior, and perceptual bias toward fearful expressions. *Personality and Individual Differences*, 52, 150-155.

Keen, S. (2006). A theory of narrative empathy. *Narrative*, 14, 207-236.

Kidd, D.C. & Castano, E. (2013). Reading literary fiction improves theory of mind. *Science*, 342, 377-380.

Lakens, D., Haans, A. & Koole, S.L. (2012). Het

belang van replicaties voor de psychologische wetenschap. Eén onderzoek is geen onderzoek. *De Psycholoog*, 2012, 10-18.

Mar, R.A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, 62, 103-134.

Mar, R.A., Oatley, K. & Peterson, J.B. (2009). Exploring the link between fiction and empathy: Ruling out individual differences and examining outcomes. *Communications*, 34, 407-428.

Nowicki, S. & Carton, J. (1993). The measurement of emotional intensity from facial expressions. *Journal of Social Psychology*, 133, 749-750.

Rubin, M., Paolini, S. & Crisp, R.J. (2010). A processing fluency explanation of bias against migrants. *Journal of Experimental Social Psychology*, 46, 21-28.

Shamay-Tsoory, S.G. & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: A lesion study. *Neuropsychologia*, 45, 3054-3067.

Simmons, J.P., Nelson, L.F. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.

Villiers, J.G. de & Villiers, P.A. de (2014). The role of language in Theory of Mind development. *Topics in Language Disorders*, 34, 313-328.

Watson, D., Clark, L.A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063-1070.

Wellman, H.M., Cross, D. & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655-684.

### APPENDIX 1: DE TAKEN

*De Reading the Mind in the Eyes Test* (RMET;  $\alpha=.63$ ) is een test waarbij de affectieve component van ToM wordt gemeten (Baron-Cohen et al., 2001). De test bestaat uit 36 trials, met plaatjes van alleen de ogen van een acteur. Bedoeling is dat de deelnemers een van de vier emoties kiezen die bij het plaatje zou horen.



*De Diagnostic Analysis of Non-verbal Accuracy 2 – Adult Faces Test* (DANVA2-AF;  $\alpha=.78$ ) is een van de acht subtests van de DANVA2, een test die meet in hoeverre men in staat is om non-verbale sociale informatie te interpreteren en uit te drukken (Nowicki & Carton, 1993). De DANVA2-AF bestaat uit 24 foto's van gezichten, waarvan de deelnemer de uitdrukking moet benoemen. Steeds kan worden gekozen uit de emoties blijheid, verdriet, angst of boosheid. *De Yoni Task* ( $\alpha=.67$ ) bestaat uit 61 trials (Shamay-Tsoory & Aharon-Peretz, 2007). Bij elke trial staat er een smiley in het midden, genaamd Yoni (in ons onderzoek genaamd John), met daaromheen in elke hoek een gekleurd plaatje (zie onderstaande afbeelding). Over elke trial wordt een vraag gesteld die de proefper-

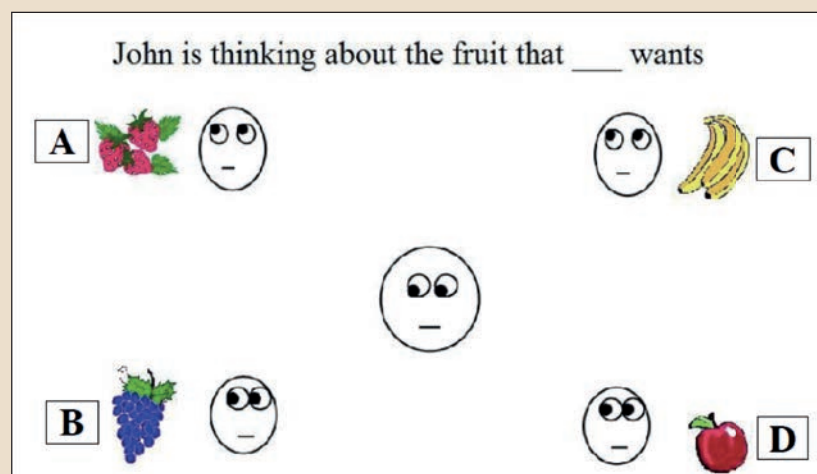
## VERVOLG APPENDIX 1: DE TAKEN

soon moet beantwoorden met de opties A, B, C of D. Het antwoord op de vraag kan worden bepaald door verschillende factoren, waaronder de emotie van Yoni, de kijkrichting van Yoni en de karakteristieken en emoties van de overige vier plaatjes.

Deze test kent twee hoofdcondities: de eerste-ordeconditie en de tweede-ordeconditie. In de eerste-ordeconditie gaat de vraag over Yoni en de weergegeven objecten in de vier hoeken van het beeldscherm. Bij de tweede-ordeconditie staan er bij de vier objecten/plaatjes in de hoeken ook andere smileys. Bij deze conditie gaat het om de verhouding tussen Yoni en de smileys ten opzichte van elkaar en de bijbehorende objecten. Beide hoofdcondities maken onderscheid tussen fysieke, affectieve en cognitieve trials. Bij de fysieke en cognitieve conditie geven Yoni en de

andere plaatjes neutrale informatie, maar bij de affectieve conditie geven Yoni en de andere plaatjes affectieve informatie. Ook zit er affectieve informatie in de vraag verwerkt. De fysieke conditie geldt als controleconditie. Voorafgaand aan de 61 trials werden er vier voorbeeldvragen aangeboden waar ook feedback op werd gegeven.

De *Author Recognition Task* (ART) bestaat uit een lijst met 130 namen, waarvan er 65 van bestaande, populaire auteurs zijn (Acheson, Wells & MacDonald, 2008). De andere 65 namen zijn fictief. Bedoeling is dat deelnemers de auteurs aanvinken die zij herkennen. Gokken wordt afgeraden, want voor elke fout gaat er een punt van de totaalscore af. Zo kan goed gemeten worden in welke mate de deelnemers bekend zijn met literatuur.



## APPENDIX 2- GEDETAILEERDE TOETS

## RMET-scores

De scores op de RMET-test werden voorspeld op basis van een lineair regressiemodel met leeftijd, geslacht, hoogst genoten opleiding, totale tijd besteed aan de RMET-test en type gelezen tekst (populaire fictie versus literaire fictie) als prediktoren. De totale RMET-tijd is gebaseerd op *second clicks*: een *second click* geeft voor elk item in de RMET-test hoeveel tijd een deelnemer een item in beeld heeft gezien voordat hij doorging naar het volgende scherm. De totale RMET is voor iedere deelnemer vervolgens bepaald door alle *second click*-tijden bij elkaar op te tellen.

De regressieanalyse ( $n=294$ ) liet zien dat slechts 1,7% van de variantie in RMET-scores voorspeld kon worden op basis van het eerder genoemde lineaire model ( $F(5,288) = 1.004, p=.416$ ). Voor de individuele prediktoren gold dat de relatie met RMET-scores in alle gevallen zeer zwak was: de maximale (niet signi-

## VERHALEN

**C**orrie van Alice Munro (fragment uit een literair fictieverhaal)

Corrie was right across the table, looking their guest in the eye. She seemed to think this was funny. "Who's she going to marry?" her father continued. "She's twenty-five." Corrie raised her eyebrows, made a face. "You missed a year," she said. "Twenty-six." "Go ahead," her father said. "Laugh all you like." She laughed out loud, and, indeed, what else could she do? the guest

## ENDE STATISTIEKEN

ficante) partiële correlatie was .085 en deze gold voor de voorspeller 'hoogst genoten opleiding'. Het effect van conditie op RMET-scores was extreem klein met een partiële correlatie van .071. Het gecorrigeerde gemiddelde was 27.233 [95% betrouwbaarheidsinterval 26.581; 27.885] in de populaire-fictieconditie en 27.808 [95% betrouwbaarheidsinterval 27.156; 28.460] in de literaire-fictieconditie.

**Yoni-scores**

De scores op de Yoni-taken werden in de tweede analyse, net als in de originele studie, geanalyseerd met een 2 Type Yoni (cognitief versus affectief) x 2 Moeilijkheidsgraad (eerste orde versus tweede orde) x 2 Type tekst (populaire fictie versus literaire fictie) repeated measures ANCOVA met ART-scores en de scores op een Yoni-controletaak (de scores op de fysieke Yoni-taak) als covariaten. Uit de analyse kwam echter naar voren dat de ART-covariaat interacteerde met de moeilijkheidsgraadfactor. Dit betekent dat de hellingscoëfficiënt van

de lineaire relatie tussen ART-scores en de geaggregeerde Yoni-scores verschilt voor eerste orde- en tweede ordetaken. Hiermee is een belangrijke assumptie voor het uitvoeren van de ANCOVA geschonden (een vergelijkbare schending kwam overigens ook voor in de originele studie) en daarom zullen wij hier niet de resultaten van de eerder genoemde ANCOVA maar van de 2 Type Yoni (cognitief versus affectief) x 2 Moeilijkheidsgraad (eerste orde eerste orde versus tweede orde) x 2 Type tekst (populaire fictie versus literaire fictie) mixed ANOVA rapporteren. Voor het hoofdeffect van type tekst – het cruciale effect in de replicatiepoging – is het resultaat voor beide analyses in termen van effectgrootte en statistische significantie overigens hetzelfde.

De mixed ANOVA liet zien dat het marginale gemiddelde voor de cognitieve Yoni-taken ( $M=.919$ , 95% betrouwbaarheidsinterval .909; .928] lager was dan voor de affectieve Yoni-taken ( $M=.941$ , 95% betrouwbaarheidsinterval .936; .947],  $F(1,298)=21.550$ ,  $MSE=.007$ ,

$p<.001$ , partial  $\eta^2=.067$ . Daarnaast was het marginale gemiddelde voor de eerste orde Yoni-taken ( $M=.981$ , 95% Betrouwbaarheidsinterval .977; .985] hoger dan voor de tweede orde Yoni-taken ( $M=.879$ , 95% Betrouwbaarheidsinterval .869; .890],  $F(1,298)=363.410$ ,  $MSE=.008$ ,  $p<.001$ , partial  $\eta^2=.549$ . Het effect van moeilijkheidsgraad was echter groter voor de cognitieve Yoni-taken dan voor de affectieve Yoni-taken,  $F(1,298)=72.222$ ,  $MSE=.008$ ,  $p<.001$ , partial  $\eta^2=.195$ . De overige tweewegsinteractie-effecten, het driewegsinteractie-effect en het hoofdeffect van type tekst waren alle zeer klein en niet significant. Met betrekking tot het effect van type tekst lieten de resultaten zien dat het marginale gemiddelde voor de Yoni-taken in de populaire fictieconditie ( $M=.927$ , 95% Betrouwbaarheidsinterval .918; .936] vergelijkbaar was met dat in de literaire fictieconditie ( $M=.933$ , 95% Betrouwbaarheidsinterval .924; .942],  $F<1$ ,  $p=.329$ , partial  $\eta^2=.003$ .)

thought. His name was Howard Ritchie, and he was only a few years older than she was, but already equipped with a wife and a young family, as her father had immediately found out. Her expressions changed very quickly. She had bright-white teeth and short, curly, nearly black hair. High cheekbones that caught the light. Not a soft woman. Not much meat on the bone, which was the sort of thing her father might find to say next. Howard Ritchie thought of her as the type of girl who spent a lot of time playing golf and tennis. In spite of her quick tongue, he expected her to have a conventional mind.

*Too Many Have Lived* van Dashiell Hammett (fragment uit een populair fictieverhaal)

Spade rocked back in his chair. His face — given a not unpleasantly Satanic cast by the v's of his bony chin, mouth, nostrils, and thickish brows — was as politely interested as his voice. "Why?" The green-eyed man spoke quietly, with assurance: "I can talk to you, Spade. You've the sort of reputation I want in a private detective. That's why I'm here." Spade's nod committed him to nothing. The green-eyed man said, "And any fair price is all right with me."

Spade nodded as before. "And with me," he said, "but I've got to know what you want to buy. You want to find out what's happened to this — uh — Eli Haven, but you don't care what it is?" The green-eyed man lowered his voice, but there was no other change in his mien: "In a way I do." For instance, if you found him and fixed it so he stayed away for good, it might be worth more money to me." "You mean even if he didn't want to stay away?" The green-eyed man said, "Especially."