

Het effect van stereotype dreiging op STEM-prestaties van leerlingen in het primair en secundair onderwijs

Student: M. C. Hoexum (s3418898)

Begeleider: Prof. dr. H. Korpershoek

Tweede beoordelaar: dr. D.D.N.M. Kostons

Rijksuniversiteit Groningen

Faculteit der Gedrags- en Maatschappijwetenschappen

Masterthesis Onderwijswetenschappen, track onderwijsinnovatie

26 mei 2023

Abstract

In STEM-professions (*Science, Technology, Engineering & Mathematics*), an unequal distribution can be found in gender and ethnicity of employees. Stereotyped career decisions follow from multiple contributing factors, for example performance and self-efficacy. Differences in performance between students occur from primary school onwards. Stereotype threat may explain the difference in learning performance between students in STEM-subjects. Stereotype threat is an explanatory concept in which groups perform according to the stereotype existing about the group.

Prior research shows that, in the mathematical domain, boys generally outperform girls in primary and secondary education. Besides, ethnic minorities generally perform lower academically than the ethnic majority. However, a systematic overview of the effect of stereotype threat on STEM-performance is missing. This review study focuses on summarizing recent literature on the effect of stereotype threat on STEM-performance of students in primary and secondary education. In this review, articles were collected in which the effect of stereotype threat on STEM-performance was examined. A meta-analysis is used to examine the average total effect of stereotype threat. Besides that, two separate analyses are executed for girls and ethnic minorities.

Fifteen articles were included in the analysis. The articles contain twenty-three samples with in total 4,522 students. The results show a small to medium size effect of stereotype threat on STEM-performance ($d = .260$). For girls a medium size effect was found ($d = .375$). On ethnic minorities only two studies were found. Both reported a large effect of stereotype threat on STEM-performance, but with a low statistical power. This seems to indicate that stereotype threat influences the learning performance of students in STEM subjects.

Inleiding

Van alle banen in Nederland die vallen onder de disciplines *Science, Technology, Engineering & Mathematics* (STEM), wordt slechts 14,4% ingevuld door vrouwelijke medewerkers (Techniekpact, 2021). Het proces van beroepskeuze begint al met de profielkeuze op de middelbare school. Meisjes (13%) kiezen minder vaak voor een technisch profiel dan jongens (31%). Het doorstroompercentage van meisjes naar STEM-opleidingen ligt vooral op het MBO laag (8%, jongens 49%). Deze cijfers zijn al tien jaar redelijk stabiel. Naast meisjes lijken ook etnische minderheden ondervertegenwoordigd in STEM-beroepen. In Amerika is onderzoek gedaan naar de mate waarin mensen van kleur vertegenwoordigd zijn in STEM-beroepen (Fry et al., 2021). Dit onderzoek bestudeerde verschillende databases die de verdeling van achtergrondkenmerken van medewerkers en hun beroep registreren. Hieruit kwam naar voren dat 67% van de medewerkers in STEM-beroepen als wit geregistreerd stond, terwijl dit percentage in de totale beroepsbevolking 63% bedraagt. Er zijn echter meerdere beroepen waarin witte personen sterk oververtegenwoordigd zijn, zoals civiele technici (73% wit) en boswachters (90% wit).

Er zijn diverse onderzoeken gedaan om factoren te identificeren die verschillen in de verdeling van geslacht en etniciteit per beroep mogelijk kunnen verklaren. Zo zouden een vroege interesse en self-efficacy in STEM op school de keuze voor een carrière in de vakgebieden stimuleren (Lent et al., 2018). In deze studie in de Amerikaanse context is voor meisjes en etnische minderheden een sterkere negatieve correlatie waargenomen tussen barrières (bv. armoede of gebrek aan steun van familie) en het uiteindelijk kiezen van een beroep in STEM dan bij jongens en de witte meerderheid. Ook de prestaties in STEM-vakken lijken voorspellend te zijn voor de toekomstige beroepskeuze van leerlingen (Wang et al., 2015). Leerlingen met een hogere score in bijvoorbeeld wiskunde kiezen vaker voor een STEM-beroep dan leerlingen met een lagere score. Prestaties in STEM en zowel interesse, self-efficacy als voorkennis in STEM zijn niet onafhankelijk van elkaar (Nugent et al., 2015). Een leerling met lagere prestaties heeft waarschijnlijk minder interesse in STEM dan een leerling met hoge prestaties. Ondanks deze inzichten is het vooralsnog niet gelukt om in de samenleving een evenredige verdeling over alle groepen (ongeacht hun achtergrondkenmerken) te krijgen wat betreft de beroepskeuze.

Een mogelijke verklaring voor de ongelijke leerprestaties en daarop volgend de beroepskeuze van leerlingen in STEM-vakgebieden is *stereotype threat* ofwel stereotype dreiging. Dit concept, beschreven in de sociaalpsychologische theorie, tracht verschillen in

leerprestaties tussen mannen en vrouwen of etnische groepen te verklaren (Spencer et al., 1999; Steele & Aronson, 1995; Steele, 1997). De stereotype dreiging komt als volgt naar voren: wanneer leden van een groep, over wie een negatief stereotype bestaat betreffende hun leerprestaties in een bepaald domein, een moeilijke taak uitvoeren in dat domein, dan wordt van deze groep verwacht dat zij door de druk van het stereotype slechter presteren op de taak. In onderzoek wordt doorgaans op twee manieren naar de invloed van stereotype dreiging gekeken. In een experiment kan sprake zijn van het expliciet aanspreken van het stereotype, bijvoorbeeld door te vertellen dat de andere groep beter is, of het impliciet aanspreken van het stereotype, bijvoorbeeld het geslacht laten invullen (Flore & Wicherts, 2015). In de controlegroep kan geen informatie worden gegeven of er kan gekozen worden voor het neutraliseren van het stereotype, bijvoorbeeld door uitleg te geven over stereotype dreiging aan de deelnemers.

In de eerste experimentele studies naar stereotype dreiging, zoals dat van Steele en Aronson (1995), lag de focus op verschillen in academische prestaties tussen etnische minderheden en de etnische meerderheid. Deelnemers kregen een mondeling examen. Voor de controlegroep werd dit gepresenteerd als een probleem-oplossingstaak en voor de experimentele groep als een test van intelligentie. De introductie was bedoeld om het stereotype, waarbij etnische minderheden lager presteren als het om academische prestaties gaat, te activeren. Deelnemers van etnische minderheden in de experimentele conditie presteerden lager dan deelnemers van de etnische meerderheid in diezelfde conditie. Later werd het construct ook gemeten bij andere groepen, zoals bij meisjes en vrouwen in kwantitatieve vakgebieden, zoals wiskunde of natuurkunde (Spencer et al., 1999). In dit onderzoek werd in de experimentele groep vooraf uitleg gegeven over het stereotype dat vrouwen lager zouden presteren in wiskunde dan mannen. Het vergroten van het bewustzijn over het stereotype had invloed op de verschillen in wiskundeprestaties tussen mannen en vrouwen. In de experimentele groep waren de verschillen kleiner dan in de controlegroep. Daarmee presteren vrouwen in de experimentele groep gemiddeld beter dan de vrouwen in de controlegroep en dichterbij de mannen in de experimentele groep. Het verkleinen van het verschil kan verklaard worden door het wegnemen van de stereotype dreiging, dit effect lijkt in de experimentele conditie irrelevant geworden.

In eerder gepubliceerde meta-analyses over stereotype dreiging (bijv. Flore & Wicherts, 2015; Nguyen & Ryan, 2008) lag de focus voornamelijk op het effect van stereotype dreiging op wiskunde-prestaties van meisjes en vrouwen. Zo rapporteerden Nguyen en Ryan (2008) een gemiddelde effectgrootte van $d = 1.26$. Er is hierbij sprake van een

negatief effect van stereotype dreiging op de wiskundeprestaties van meisjes uit etnische minderheidsgroepen. Daarnaast kwam in dit onderzoek naar voren dat er mogelijk meerdere factoren een modererende invloed hebben op de stereotype dreiging. In de meta-analyse van Flore en Wicherts (2015) is een gemiddelde effectgrootte van $\bar{g} = -0.22$ gerapporteerd. Stereotype dreiging had een negatief effect op de wiskunde prestaties van meisjes. De focus in deze analyse lag op het verschil tussen jongens en meisjes onder de 18 jaar. Zij vonden geen significante moderatoren, maar wijzen wel op mogelijke publicatiebias in de literatuur. Een derde meta-analyse naar stereotype dreiging concludeerde dat het mogelijk zou kunnen zijn dat huidige studies grotere effectgroottes rapporteren dan daadwerkelijk aanwezig zijn (Pichokiroga et al., 2021). Zij stellen dat er wellicht te veel onduidelijkheid bestaat over de specificaties van stereotype dreiging in de verschillende onderzoeken. Hierdoor ontstaat ruis en mogelijk een grotere variatie dan in de populatie daadwerkelijk te vinden is. Interpretatie van de resultaten zal, aldus de auteurs, daarom altijd met enige voorzichtigheid moeten gebeuren.

In dit onderzoek wordt stereotype dreiging bekeken in een bredere context dan alleen de wiskunde. Er wordt gekeken naar het hele STEM-gebied. Uit een preliminaire meta-analyse van Becker en Park (2011) blijkt dat integratief onderwijs in STEM-vakgebieden een positief effect heeft op de STEM-prestaties van leerlingen. Door de rijke leeromgeving die door integratie geboden wordt, ontwikkelen leerlingen ook een positievere attitude tegenover STEM. Ook latere meta-analyses vinden een gemiddeld tot groot effect van geïntegreerd STEM-onderwijs op academische prestaties in de STEM-gebieden (Belland et al., 2016; D'Angelo et al., 2014; Kazu & Kurtoglu Yalcin, 2021). Bij onderzoek naar STEM-onderwijs is het van belang rekening te houden met de definitie van STEM die in het onderzoek gehanteerd wordt. Over de invulling van het acroniem STEM (*Science, Technology, Engineering & Mathematics*) binnen het onderwijs bestaat nog geen eenduidige visie (McComas & Burgin, 2020). Dit maakt dat onderzoeken mogelijk niet altijd vergelijkbaar zijn. Het begrip STEM stamt uit 2001 en is sindsdien populair geworden in alle vier de betrokken velden. Er is veel debat over wat STEM-onderwijs precies inhoudt. Definities lopen uiteen van een losse aanpak waarin elk van de vier vakgebieden wordt onderwezen door een eigen leraar, tot een volledig geïntegreerd curriculum waarbij alle vakgebieden in verband worden aangeboden. Voor dit onderzoek wordt uitgegaan van de definitie waarbij STEM-onderwijs gericht is op de gezamenlijke constructie van kennis in meer dan één STEM-vakgebied (Corlu et al., 2014). Bij een integratieve visie worden minstens twee of meer van de vakgebieden tegelijk aangeboden. Voor het basisonderwijs wordt er bijvoorbeeld gekeken

naar het integreren van rekenen, natuur en techniek. Voor het middelbaar onderwijs gaat het om het samenvoegen van wiskunde, natuurkunde, scheikunde en techniekvakken.

Hoewel integratief STEM-onderwijs volgens onderzoek zorgt voor betere leerprestaties in de STEM-vakken, zijn de leerprestaties daarmee nog niet voor alle groepen gelijk. Dit blijkt vooral onder meisjes en etnische minderheden het geval. Meisjes presteren gemiddeld genomen lager in STEM-vakgebieden dan jongens (Ambady et al., 2004; Beasley & Fischer, 2012; Spencer et al., 1999). Daarnaast presteren etnische minderheden gemiddeld genomen lager in academische vaardigheden dan de (vaak witte) meerderheid (Armenta, 2010; Steele & Aronson, 1995). De academische leerprestaties van leerlingen en studenten hebben invloed op het kiezen van de vervolgopleiding en daarna de beroepskeuze (Parker et al., 2012). Zoals eerder besproken, zou het een probleem kunnen vormen voor de doorstroming naar gerelateerde opleidingen en beroepen, wanneer een bepaalde groep leerlingen lager presteert in STEM-vakgebieden.

Deze review is erop gericht het onderzoek dat reeds bestaat over stereotype dreiging op STEM-vakgebied in het primair en secundair onderwijs nader te bestuderen. Hierbij wordt gekeken naar academische prestaties, omdat die voorspellend zijn voor de motivatie en daarmee de beroepskeuze (Parker et al., 2012; Wang et al., 2015). Er is al veel bekend over het effect van geïntegreerd STEM-onderwijs en toch blijven er verschillen zichtbaar tussen jongens en meisjes en meer- en minderheden in leerprestaties (Armenta, 2010; Beasley & Fischer, 2012). In eerdere meta-analyses over stereotype dreiging lag de focus alleen op wiskundeprestaties (Flore & Wicherts, 2015; Nguyen & Ryan, 2008). In deze studie wordt het hele STEM-gebied bekeken, waardoor andere onderzoeken meegenomen kunnen worden in de analyse. Van onderzoeken naar het effect van stereotype dreiging op STEM-prestaties, is nog geen systematisch overzicht waarin het gemiddelde effect van stereotype dreiging over alle studies beschreven wordt, zoals dat in het wiskundig domein wel voorhanden is. Het overzicht van de huidige literatuur wordt gegeven aan de hand van een meta-analyse, waarin het gemiddelde effect van stereotype dreiging kan worden vastgesteld op meerdere eerder uitgevoerde experimentele studies. Daarnaast kan gekeken worden naar waar al veel over bekend (bv. welke STEM-domeinen) is en waar mogelijk nog witte vlekken in het onderzoeksveld bestaan. Daarmee kan er richting gegeven worden voor mogelijk vervolgonderzoek over de invloed van en eventueel het interveniëren op stereotype dreiging in de STEM-gebieden. Voor de praktijk kan dit onderzoek dienen als samenvatting van de bestaande onderzoeken naar stereotype dreiging binnen STEM-onderwijs. Professionals, zoals

leraren en schoolleiders kunnen geïnformeerd handelen naar de invloed van stereotype dreiging. De hoofdvraag is:

Heeft een stereotype dreiging invloed op de testcores op STEM-gebied van stereotype minderheden in het primair en secundair onderwijs?

Om richting te geven aan de analyses is deze vraag opgesplitst in twee subvragen:

1. *Heeft een stereotype dreiging invloed op de testcores op STEM-gebied van meisjes in het primair en secundair onderwijs?*

2. *Heeft een stereotype dreiging invloed op de testcores op STEM-gebied van etnische minderheden in het primair en secundair onderwijs?*

Er wordt verwacht dat de stereotype dreiging een negatief effect heeft op de scores op STEM-gebied van meisjes en etnische minderheden, zoals dit ook gevonden is in het wiskundig domein (Flore & Wicherts, 2015; Nguyen & Ryan, 2008). De verwachting is dat een klein tot middelgroot effect gevonden wordt, gezien dit in het wiskundige domein het geval is.

Methode

Procedure

Voor dit onderzoek is gekozen voor een meta-analyse. Hierbij is getracht overzicht te geven van de huidige stand van zaken in de literatuur. Voor de meta-analyse is gezocht naar artikelen in de databases van ERIC, Web Of Science en PsychInfo. Daarnaast is een informele zoekopdracht uitgevoerd via Google Scholar. De zoekterm die is gebruikt is de combinatie van 'stereotype threat' en 'STEM or Science and Technology and Engineering and Mathematics'. Ten slotte is gekeken naar de literatuurlijsten van de eerdere meta-analyses om zo via de sneeuwballmethode eventuele aanvullende artikelen te verzamelen. Alle gevonden artikelen zijn verzameld in de referentiesoftware van Endnote. Hier zijn eerst duplicaten verwijderd. Daarna zijn de titels en abstracts beoordeeld op relevantie/inclusie, waarbij gebruik is gemaakt van de criteria in Tabel 1. Voor de publicatiedatum is uitgegaan van de start van het benoemen van STEM als concept in de wetenschappelijke literatuur (McComas & Burgin, 2020). Dit was in 2001 en daarom zijn artikelen met een publicatiedatum in 2002 tot de datum van de zoekopdracht (januari 2023) meegenomen in de analyse. Alleen studies met een experimenteel of quasi-experimenteel onderzoeksdesign zijn meegenomen. In deze studies werd stereotype dreiging geactiveerd, of werd er gericht op preventie in de experimentele groep. Daarnaast moest er een controlegroep zijn waar deze manipulatie niet

plaatsvond of een controlegroep waarbij de interventie wél plaatsvond, maar waarbij de kenmerken van de controlegroep anders zijn (bijvoorbeeld meisjes in de experimentele groep vergeleken met jongens). Voor de uitkomstmaten is er gekeken naar academische prestaties in STEM-gebied. Door in de analyse uitkomstmaten van alle verschillende vakgebieden mee te nemen is getracht een samenvattende effectgrootte op STEM-prestaties te kunnen geven.

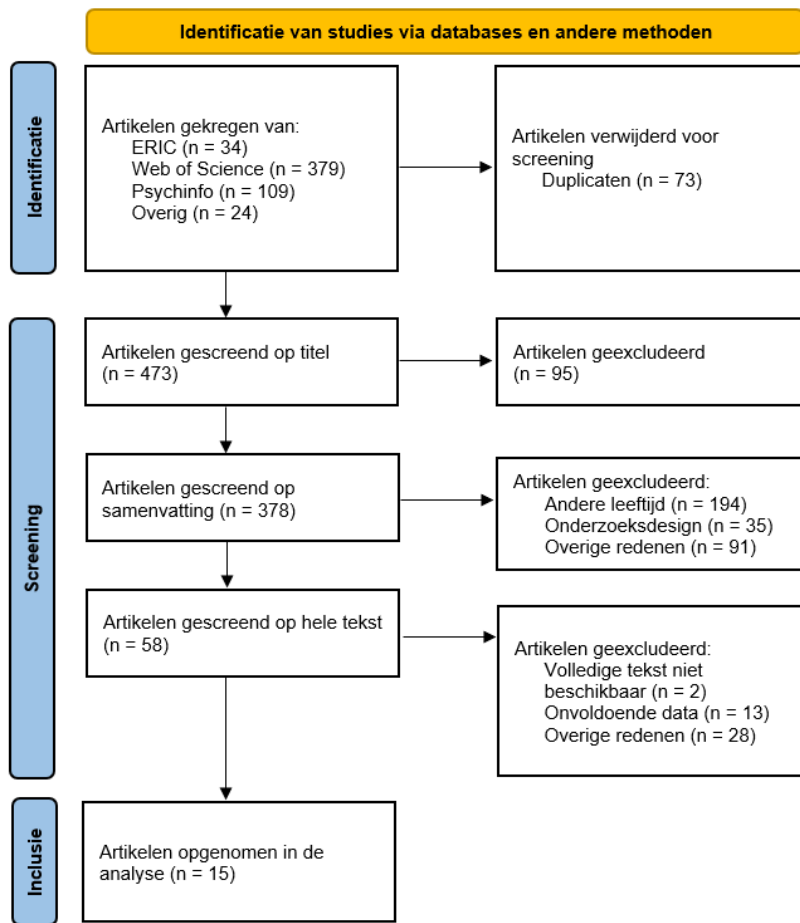
Tabel 1: Inclusiecriteria

 criterium	 Inclusie
Publicatieperiode	Van 2002 tot 2023
Publicatietype	Artikelen gepubliceerd in wetenschappelijke tijdschriften of dissertaties.
Taal	Nederlands, Vlaams of Engels
Deelnemers	Leerlingen in het primair of secundair onderwijs (tot 18 jaar)
Onderzoeksdesign	Een experimenteel of quasi-experimenteel design met stereotype activatie of preventie en een controlegroep.
Uitkomsten	Academische prestaties op STEM-gebied.
Implementatie	Meet het effect van stereotype dreiging in veld- of onderwijssetting. Daarbij wordt onderscheid gemaakt in etniciteit, in gender of in een combinatie van beide.
Toegankelijkheid	Volledige tekst beschikbaar
Beschikbare data	Steekproefgrootte, standaarddeviatie en gemiddelden of een effectgrootte voor het verschil tussen twee groepen (bijvoorbeeld, Cohen's <i>d</i> , Hedges <i>g</i>). Of gegevens waarmee een effectgrootte te berekenen is.

Uit de zoekopdrachten kwamen in totaal 546 artikelen naar voren. Het proces van artikelselectie is schematisch weergegeven in Figuur 1. Eerst werden duplicaten verwijderd, zodat enkel unieke artikelen overbleven. Daarna werd gekeken naar de titel van het artikel. Wanneer er geen focus lag op STEM in het basis- of middelbaar onderwijs werd een artikel niet verder meegenomen. Vervolgens werden de samenvattingen doorgenomen. Artikelen met een andere leeftijdsclassificatie of een afwijkend onderzoeksdesign werden zo gefilterd. Van de overige artikelen werd gekeken naar de hele tekst. Voor twee artikelen was de volledige tekst niet beschikbaar, waardoor deze artikelen zijn uitgesloten. Daarnaast bevatte een aantal artikelen onvoldoende data om een effectgrootte te kunnen berekenen. Uiteindelijk bleven er 15 artikelen over. De meeste artikelen zijn uitgesloten op de leeftijd van de deelnemers, het betrof hier voornamelijk studenten in het hoger onderwijs.

Figuur 1

Prisma diagram (Page et al., 2021)



Dataextractie

Voor het juist verkrijgen van de data is de volgende kerninformatie uit de artikelen gehaald en verzameld in Excel: bibliografische kenmerken (Auteur, publicatiejaar, land), bestudeerde populatie, inclusief steekproefgrootte, en uitkomstgegevens. Wanneer er onvoldoende statistische gegevens beschikbaar waren in de volledige tekst van een artikel, zoals gemiddelden, groeps groottes of standaarddeviaties, is alsnog gekozen voor exclusie van het artikel. Het ontbreken van deze gegevens maakte het onmogelijk de effectgrootte van de studie te berekenen, terwijl deze noodzakelijk is voor de analyse.

Analyseplan

De effectgrootte data uit de artikelen zijn geanalyseerd met het random-effects model. Er werd namelijk verwacht dat de ware effectgrootte per studie zal verschillen in verband met verschillen tussen participanten in bijvoorbeeld leeftijd en hoeveel tijd zij hebben doorgebracht op school. Om dit te controleren is bij het model de heterogeniteitsassumptie

bekeken. Voor elke (sub)vraag is een afzonderlijk model geschat. De modellen zijn geschat in Jamovi versie 2.3 (The Jamovi Project, 2022), met behulp van het MAJOR pakket. Alle modellen zijn frequentistisch en er is gebruik gemaakt van een alfa van 5%. De aanname die daarbij geldt is dat er een normale verdeling is van ware effectgroottes. Voor elke studie zijn de effectgroottes gewogen. Wanneer deze niet in het artikel beschikbaar waren zijn de effectgrootte Cohen's d berekend met behulp van de gemiddelden en de standaarddeviatie. Hiervoor is gebruik gemaakt van de volgende formule (Borenstein et al. 2009):

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}$$

Waarbij \bar{X}_1 en \bar{X}_2 de gemiddelden zijn van de te vergelijken groepen en S_{within} de binnengroeps-standaarddeviatie berekend met de volgende formule:

$$S_{within} = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

Waarin n_1 en n_2 de steekproefgroottes van de groepen zijn en S_1 en S_2 de standaarddeviaties van de twee groepen.

Als een andere effectgrootte dan Cohen's d werd gerapporteerd, dan is deze omgezet naar Cohen's d (zie Borenstein et al., 2009). Met deze gegevens is een schatting gegeven voor de gemiddelde effectgrootte en de heterogeniteit tussen de studies. Ook voor studies met wiskunde en zonder wiskunde is de gemiddelde effectgrootte berekend om te kunnen inschatten of niet een te groot deel van het effect aan enkel het wiskundedomein toegedicht kan worden. Tot slot is *funnelplot* gemaakt per analyse om te kijken naar publicatiebias.

Resultaten

Databeschrijving

De kenmerken van de vijftien geïncludeerde artikelen zijn weergegeven in Tabel 2. Samen bevatten de artikelen 23 effectgroottes voor stereotype dreiging op STEM-prestaties. De uitkomstmaten bevatten doorgaans een toetsing van wiskunde, al dan niet gecombineerd met andere vakgebieden ($N = 15$). De steekproeven zijn grotendeels onafhankelijk. In de steekproeven 1.1 & 1.2, 12.1 & 12.3 en 12.2 & 12.4 zit overlap van leerlingen. In studie 1 zijn twee verschillende resultaten gemeten, namelijk wiskunde en scheikunde (Franklin & Rangel, 2022). In de steekproeven van artikel 12 wordt afwisselend onderscheid gemaakt in jongens en meisjes (12.1 & 12.2) en etnische minderheden en meerderheden (Wendt et al., 2018). In totaal is bij zeventien van de studies een onderscheid gemaakt in de steekproef tussen jongens en meisjes of zijn er alleen meisjes meegenomen in de steekproef. In al deze studies zijn de

meisjes ingedeeld in de experimentele groep. Bij slechts twee steekproeven (Wendt et al., 2018) is onderscheid gemaakt naar de etniciteit van de leerlingen. De leerlingen uit een minderheidsgroep zijn hierbij de experimentele groep. De activatie van het stereotype in de primaire studies kent veel verschijningsvormen. Er is een aantal artikelen gericht op het voorkomen van stereotype dreiging (studies 4, 14 en 15). Andere artikelen activeren het stereotype impliciet (bv. studies 3, 8 en 10). In sommige artikelen wordt ervan uitgegaan dat door een bepaalde conditie het stereotype in meer of mindere mate aanwezig is (bv. studies 1, 5 en 6). De meeste artikelen zijn gepubliceerd in een wetenschappelijk tijdschrift ($N = 13$) en zijn uitgevoerd in een Amerikaanse context ($N = 8$). De leeftijd van de leerlingen loopt uiteen van 4 tot 18 jaar. In totaal zijn gegevens beschikbaar van 4,522 leerlingen.

Tabel 2*Kenmerken van geïncludeerde artikelen.*

Artikel		Type	Leeftijd	Land	N	d	Manier van activeren stereotype dreiging	Uitkomstmaten	Onderscheid meisjes	Onderscheid etnische herkomst
Nr.	Auteurs (jaartal)									
1.1	Franklin & Rangel (2022)	Artikel	13-14	VS	468	0.38	Leerlingen in een gemixte of alleen meisjes klas	STAAR-scores wiskunde	Ja	Nee
1.2	Franklin & Rangel (2022)	Artikel	13-14	VS	662	0.97	Leerlingen in een gemixte of alleen meisjes klas	STAAR-scores wetenschappen	Ja	Nee
2	Gunderson et al. (2018)	Artikel	6-8	VS	634	-0.01	Eerst vragen over motivatie, daarna wiskunde gemeten	Woodcock - Johnson III Test of achievement	Ja	Nee
3.1	Huguet & Regner (2007)	Artikel	11-13	Frankrijk	40	0.94	Naam test is stereotyperend of niet	Het natekenen van een figuur	Ja	Nee
3.2	Huguet & Regner (2007)	Artikel	11-13	Frankrijk	454	0.27	Naam test is stereotyperend of niet	Het natekenen van een figuur	Ja	Nee
4.1	Kurtz-Costes et al. (2008)	Artikel	9-10	VS	90	0.02	Interventie bij meisjes op afname stereotype dreiging	Combinatiecijfer wiskunde en wetenschappen	Ja	Nee
4.2	Kurtz-Costes et al. (2008)	Artikel	11-12	VS	91	0.38	Interventie bij meisjes op afname stereotype dreiging	Combinatiecijfer wiskunde en wetenschappen	Ja	Nee
4.3	Kurtz-Costes et al. (2008)	Artikel	13-14	VS	114	0.30	Interventie bij meisjes op afname stereotype dreiging	Combinatiecijfer wiskunde en wetenschappen	Ja	Nee
5	Montuori et al. (2022)	Artikel	5-7	Italië	109	0.51	Leerlingen leren programmeren (stereotype verlagen)	Planning en accuraatheid van het ontwerpen van een spel	Ja	Nee
6.1	Pennington et al. (2018)	Artikel	11-12	Engeland	123	-0.09	Leerlingen in een gemixte of alleen meisjes klas	Gestandaardiseerde STEM score – wiskunde, wetenschappen en ICT	Ja	Nee
6.2	Pennington et al. (2018)	Artikel	11-12	Engeland	143	-0.05	Leerlingen in een gemixte of alleen meisjes klas	Gestandaardiseerde STEM score – wiskunde, wetenschappen en ICT	Nee	Nee

7	Plant et al. (2009)	Artikel	12-15	VS	106	0.49	Man, vrouw of computer als docent	Gecombineerde scheikunde, natuurkunde en wiskunde score	Nee	Nee
8	Randolph (2018)	Thesis	13-17	VS	20	0.03	Wel of niet benoemen van rekencomponent in instructie	Gestandaardiseerde wiskunde en inzicht test	Nee	Nee
9	Shenouda (2015)	Thesis	4-5	VS	22	1.61	Wel of geen plaatje om stereotype aan te spreken	Snelheid en accuraatheid in een blokkenbouwtaak	Ja	Nee
10	Shenouda & Danovitch (2014)	Artikel	4-5	VS	158	0.50	Wel of geen plaatje om stereotype aan te spreken	Snelheid en accuraatheid in een blokkenbouwtaak	Nee	Nee
11	Tellhed et al. (2022)	Artikel	11-12	Zweden	318	-0.18	Leerlingen leren programmeren (stereotype verlagen)	Leerkrachtbeoordeling op deelname in programmeercursus	Ja	Nee
12.1	Wendt et al. (2018)	Artikel	8-9	VS	188	-1.19	Wel of niet op een STEM-school zitten	STEM-scores – wetenschappen en wiskunde	Nee	Ja
12.2	Wendt et al. (2018)	Artikel	10-11	VS	174	-0.79	Wel of niet op een STEM-school zitten	STEM-scores – wetenschappen en wiskunde	Nee	Ja
12.3	Wendt et al. (2018)	Artikel	8-9	VS	176	0.02	Wel of niet op een STEM-school zitten	STEM-scores – wetenschappen en wiskunde	Ja	Nee
12.4	Wendt et al. (2018)	Artikel	10-11	VS	163	-0.04	Wel of niet op een STEM-school zitten	STEM-scores – wetenschappen en wiskunde	Ja	Nee
13	Wille et al. (2018)	Artikel	10-11	Duitsland	163	-0.09	Uitleg video met of zonder stereotypen	Aantal sommen binnen 2 minuten	Ja	Nee
14	Wulff et al. (2018)	Artikel	15-18	Duitsland	29	2.78	Interventie bij meisjes voor toename STEM-interesse	Inhoudelijke kennis over fysica	Ja	Nee
15	Zhao et al. (2018)	Artikel	12-14	China	77	0.43	Interventie bij meisjes op afname stereotype dreiging	Examenscores wiskunde en scheikunde	Ja	Nee

Noot. N = totaal aantal leerlingen in steekproef, d = (omgerekende) totale effectgrootte, VS = Verenigde Staten

Gemiddelde effectgroottes en heterogeniteit

Op de hoofdvraag van dit onderzoek, heeft stereotype dreiging effect op STEM-prestaties van stereotype minderheden, komt uit de analyse een klein tot middelgroot effect naar voren van stereotype dreiging op STEM-prestaties (Tabel 3). Het betreft een significant effect onder een alfa van 5%. De waarde van 0 valt daarbij niet in het betrouwbaarheidsinterval (BHI). Het is onwaarschijnlijk dat er in de populatie geen effect van stereotype dreiging voorkomt.

De schatting voor heterogeniteit is significant ($Q = 188.368$, $df = 22$, $p < .001$). Een lage p -waarde geeft aan dat aan de assumptie van heterogeniteit van de effectgroottes is voldaan en dat er aangenomen mag worden dat de effectgroottes variëren in de populatie. Hiermee wordt gecontroleerd voor de legitimiteit van het random-effects model, welke alleen mag worden toegepast bij variërende effectgroottes (Borenstein et al. 2009).

Tabel 3

Gemiddeld effect stereotype dreiging op STEM-prestaties

	Geschatte effectgrootte	SE	p -waarde	BHI onderzijde	BHI bovenzijde
Intercept	.260	.128	.042	.009	.510

Noot. SE = Standaard error, BHI = Betrouwbaarheidsinterval

Voor de subvraag over het effect van stereotype dreiging op meisjes is gekeken naar de artikelen die onderscheid maken tussen jongens en meisjes in de steekproef (zie Tabel 2). Voor deze analyse zijn zeventien effectgroottes meegenomen. In Tabel 4 is te zien dat een middelgroot effect wordt gevonden van stereotype dreiging op de STEM-prestaties van meisjes. Dit effect is significant wanneer gekeken wordt met de vastgestelde alfa van 5%. De 0-waarde ligt buiten het BHI. Daarmee is het onwaarschijnlijk dat stereotype dreiging geen effect heeft op meisjes in de gehele populatie. Wanneer meisjes met een gelijke vaardigheid als jongens een STEM-toets afleggen scoren zij mogelijk lager dan jongens door de aanwezigheid van stereotype dreiging.

Bij het bekijken van de heterogeniteitsassumptie zijn geen problemen gevonden ($Q = 148.138$, $df = 16$, $p < .001$). De p -waarde is laag en daarom mag aangenomen worden dat de effectgroottes in de populatie verschillen. Dit maakt dat er geen beperkingen zijn aan het toepassen van het random-effects model als analyse.

Tabel 4*Gemiddeld effect stereotype dreiging bij meisjes op STEM-prestaties*

	Geschatte effectgrootte	SE	<i>p</i> -waarde	BHI onderzijde	BHI bovenzijde
Intercept	.375	.131	.004	.118	.633

Noot. SE = Standaard error, BHI = Betrouwbaarheidsinterval

Voor de subvraag over het effect van stereotype dreiging op etnische minderheden zijn er onvoldoende gegevens om een meta-analyse uit te voeren. Het betreft twee steekproeven (12.3 en 12.4) uit het artikel van Wendt en collega's (2018). In deze steekproef zijn leerlingen meegenomen op STEM-gecertificeerde scholen. Witte leerlingen vormden de controlegroep. Etnische minderheden werden als experimentele groep gezien. Het betrof twee steekproeven die op dezelfde scholen in overeenkomende periode zijn getrokken. De steekproef in het derde leerjaar bestond uit 12 witte leerlingen en 176 gekleurde leerlingen. De steekproef in vijfde leerjaar bestond uit 11 witte leerlingen en 163 gekleurde leerlingen. De berekende effectgroottes zijn $d = -1.194$ (derde leerjaar) en $d = -0.791$ (vijfde leerjaar). Dit is een groot effect. Interpretatie van dit effect in combinatie met de getrokken steekproef volgt in de conclusie.

Gezien studies met wiskunde in de meerderheid zijn is in tabel 5 de gemiddelde effectgrootte van studies met ($N = 15$) en zonder wiskunde ($N = 8$) weergegeven. Twee meta-analyses naar het effect van stereotype dreiging op wiskundeprestaties rapporteerden een klein ($\bar{g} = -0.22$) en een groot ($d = 1.26$) effect (Flore & Wicherts, 2015; Nguyen & Ryan, 2008). Het zou mogelijk kunnen zijn dat de gevonden effectgrootte in dit onderzoek enkel naar voren komt omdat stereotype dreiging invloed heeft op de wiskundeprestaties, zoals in de eerdere meta-analyses naar voren kwam. Wanneer er uit de studies zonder wiskundemeting een effect naar voren komt zou gesteld kunnen worden dat in de andere vakgebieden ook sprake lijkt te zijn van stereotype dreiging. Het totale gemiddelde effect van alles studies zou dan niet enkel te verklaren zijn door het wiskundige domein. Uit de analyse blijkt het gemiddelde effect van de studies met een wiskundige meting klein en niet significant. Een niet significant resultaat maakt dat het niet mogelijk is om een uitspraak te doen over een effect, danwel de grootte daarvan in de populatie aan de hand van dit onderzoek. Het gemiddelde effect van de studies zonder wiskunde is groot en significant. Dit maakt het waarschijnlijk dat er een effect van stereotype dreiging op STEM-prestaties in de populatie gevonden kan worden, ook wanneer wiskunde niet wordt meegenomen.

Tabel 5

Gemiddeld effect stereotype dreiging met en zonder wiskunde in studie

	Geschatte effectgrootte	SE	p-waarde	BHI onderzijde	BHI bovenzijde
Intercept met wiskunde	0.023	.096	.811	-0.165	0.211
Intercept zonder wiskunde	0.823	.285	.004	0.264	1.382

Noot. SE = Standaard error, BHI = Betrouwbaarheidsinterval

Publicatiebias

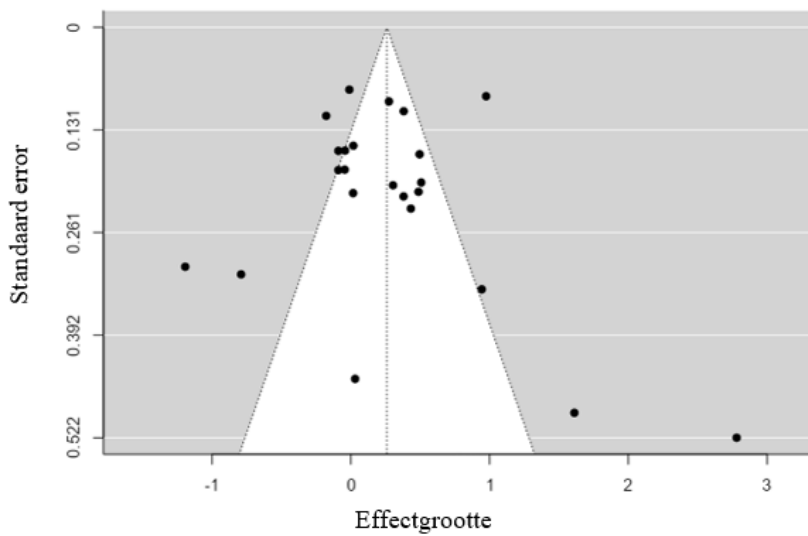
Artikelen met positieve of significante resultaten hebben een grotere kans op publicatie dan artikelen met een negatief of geen effect (Borenstein, et al., 2009; Song et al., 2010). Dit wordt het ‘file-drawer probleem’ of ‘publicatiebias’ genoemd. Om de dataset te toetsen op dit probleem wordt een *funnel plot* gemaakt voor beide analyses. Wanneer in de analyse geen sprake is van publicatiebias dan zullen de studies evenredig verdeeld zijn over de vlakken rechts en links van de verticale lijn. De verticale lijn is het door de meta-analyse verwachte gemiddelde. Bij een gerandomiseerd veld wordt verwacht dat de studies zich normaal verdelen. De studies zijn weergegeven als stippen.

In Figuur 2 is de *funnel plot* voor alle studies weergegeven. In het figuur is zijn zowel naar rechts als naar links zijn enkele uitschieters te zien. Dit duidt mogelijk op enige publicatie bias in de analyse of op heterogeniteit (Sterne et al., 2011). Dit laatste houdt in dat de metingen dermate anders zijn dat de resultaten hierdoor uiteen lopen. De *funnel plot* van alle studies geeft geen overtuigend beeld van publicatiebias, maar doordat in eerdere meta-analyses over stereotype dreiging wel sprake lijkt te zijn van publicatiebias (Flore & Wicherts, 2015; Picho-Kiroga et al., 2021) is het zeer goed mogelijk dat ook in dit onderzoek enkele studies missen.

In Figuur 3 volgt de *funnel plot* van de studies waarin een onderscheid in geslacht is meegenomen. In de *funnel plot* lijkt sprake te zijn van asymmetrie. Daarmee is het niet uit te sluiten dat er sprake is van publicatiebias. Mogelijk missen enkele studies waardoor een vertekend beeld ontstaat van de gemiddelde effectgrootte. Dit zou dan gaan om studies met een bovengemiddelde effectgrootte, omdat de meeste studies zich links van het gemiddelde bevinden.

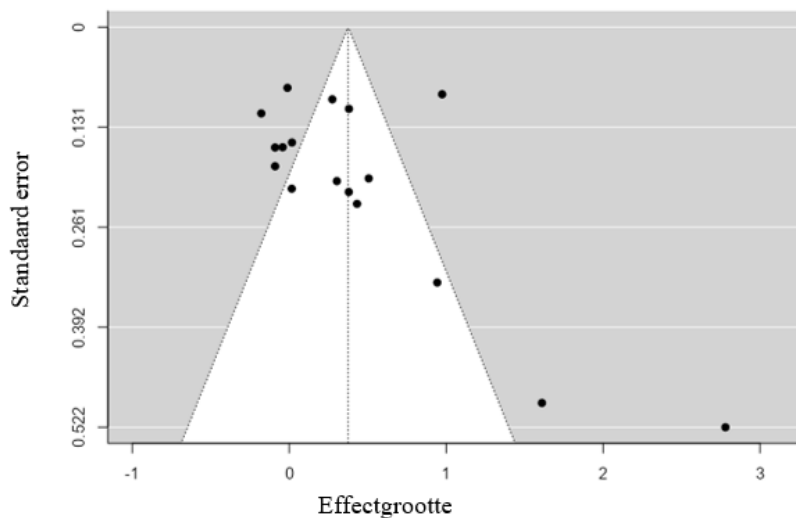
Figuur 2

Funnel plot alle geïncludeerde studies



Figuur 3

Funnel plot alle geïncludeerde studies met meisjes



Conclusie en discussie

In dit onderzoek is getracht de hoofdvraag ‘*Heeft een stereotype dreiging invloed op de test scores op STEM-gebied van stereotype minderheden in het primair en secundair onderwijs?*’ te beantwoorden. In de resultaten is een significant, klein tot middelgroot effect gevonden van stereotype dreiging op leerprestaties in het STEM-gebied. Dit sluit aan bij de verwachtingen in de gestelde hypothese. Vervolgens is ook gekeken naar of wiskunde in dit effect mogelijk de overhand zou kunnen hebben. Wanneer de studies gesplitst zijn in

metingen met en zonder wiskunde wordt voor de studies zonder wiskundemeting een groot, significant effect gevonden. Daarmee is het waarschijnlijk dat het gevonden effect van stereotype dreiging op leerprestaties in STEM-gebied in dit onderzoek niet enkel aan stereotype dreiging in de wiskunde toe te dichten valt.

Zeventien effectgroottes gaven de mogelijkheid de analyse voor de eerste subvraag, *‘Heeft een stereotype dreiging invloed op de testcores op STEM-gebied van meisjes in het primair en secundair onderwijs?’* uit te voeren. Hierbij is een significant, middelgroot effect gevonden. Daarmee lijkt het zo te zijn dat het effect van stereotype dreiging voor meisjes groter is dan voor de gehele populatie. Dit sluit aan bij de bevindingen over stereotype dreiging in wiskunde (Nguyen & Ryan, 2008; Spencer et al., 1999). Voor de tweede subvraag *‘Heeft een stereotype dreiging invloed op de testcores op STEM-gebied van etnische minderheden in het primair en secundair onderwijs?’* was het aantal studies ontoereikend om een gedegen meta-analyse uit te voeren. Uit de twee beschikbare studies lijkt een groot effect te komen van stereotype dreiging op de STEM-prestaties van etnische minderheden. Beide studies komen uit het onderzoek van Wendt en collega’s (2018). De controlegroep in dit onderzoek is echter dermate klein ten opzichte van de experimentele groep dat de power van deze vergelijking beperkt is. Wel laat onderzoek op algemene academische prestaties zien dat etnische minderheden meer stereotype dreiging ervaren dan de meerderheid (Armenta, 2010; Steele & Aronson, 1995). Er is echter op basis van deze review niet iets met zekerheid te zeggen over een mogelijk verschil in de mate van stereotype dreiging tussen etnische minderheden en meerderheden in het STEM-domein.

Ondanks het verwachte positieve effect van het steeds meer integratief aanbieden van STEM-vakken in het onderwijs op de prestaties van leerlingen (Belland et al., 2016; D’Angelo et al., 2014; Kazu & Kurtoglu Yalcin, 2021) lijkt stereotype dreiging een rol te blijven spelen in het onderwijs, voornamelijk bij meisjes en mogelijk ook bij etnische minderheden. Met het integratief aanbieden van STEM in het onderwijs lijkt de kloof in STEM-prestaties tussen jongens en meisjes en minder- en meerderheden nog niet gedicht. In dit onderzoek is getracht een overzicht te geven van de huidige literatuur over stereotype dreiging in STEM-gebied, maar het kent enkele beperkingen. Na de beschrijving van de beperkingen volgen enkele suggesties voor vervolgonderzoek.

Er zijn in totaal vijftien artikelen meegenomen in de analyses. Acht van deze artikelen zijn uitgevoerd in de Amerikaanse context, zes in Europa en slechts één in Azië. Door enkel in het Engels, Vlaams en Nederlands te zoeken lijken westerse culturen oververtegenwoordigd in de steekproef van studies. Dit heeft effect op de externe validiteit

van het onderzoek. Mogelijk zijn in andere talen meer artikelen te vinden die een inzicht kunnen geven in hoeverre de huidige resultaten dekkend zijn voor de gehele wereldpopulatie op het gebied van stereotype dreiging in STEM-gebieden.

Een andere beperking is dat in deze meta-analyse drie artikelen mogelijk een grotere invloed hebben gehad op het resultaat van de analyse dan de andere artikelen en dat er sprake lijkt te zijn van publicatiebias. Het artikel van Wendt en collega's (2018) bevatte vier studies die geschikt waren voor gebruik in de analyse. Echter, vanuit geen enkel ander artikel zijn zoveel effectgroottes meegenomen. Daarnaast hebben twee artikelen (Shenouda, 2015; Wulff et al., 2018) zeer hoge effectgroottes gerapporteerd, welke in de analyses zijn meegenomen. Mogelijk hebben deze studies tezamen een groot aandeel in de gemiddelde effectgrootte die in de resultaten is gerapporteerd. Daarbij blijkt uit de *funnel plots* dat publicatiebias in dit onderzoek niet uit te sluiten is. Dit sluit aan bij eerder onderzoek dat publicatiebias constateerde in het veld van stereotype dreiging (Flore & Wicherts, 2015) en eerder onderzoek dat rapporteerde over mogelijk grotere variatie in effectgroottes in onderzoek dan er is in de populatie (Picho-Kiroga et al., 2021). Er lijken echter vooral studies te missen met een bovengemiddelde effectgrootte, waardoor ook heterogeniteit verklarend zou kunnen zijn voor de gevonden resultaten. Uit de heterogeniteitstest blijkt ook dat variatie tussen de studies te vinden is.

Deze studie heeft als voordeel dat veel nieuwe studies zijn meegenomen ten opzichte van eerdere meta-analyses (Flore & Wicherts, 2015; Nguyen & Ryan, 2008). Dit zou kunnen wijzen op een toename in literatuur en daarbij een grote behoefte voor een samenvattend artikel dat richting kan geven aan vervolgonderzoek. Daarnaast kent de review als sterk punt dat de leeftijd van de leerlingen in de studies de range dekt die verwacht mag worden in het primair en secundair onderwijs.

Bij het analyseren van de recente literatuur is gebleken dat relatief weinig (quasi-) experimenteel onderzoek is verricht naar verschillen in stereotype dreiging op STEM-vakgebieden tussen etnische minderheden en de meerderheid. Het is aan te bevelen om in vervolgonderzoek verder te kijken naar mogelijke verschillen tussen etnische minderheden en de meerderheid, voornamelijk buiten Amerika. Dit biedt de mogelijkheid te kijken naar de mogelijke impact van stereotype dreiging op deze specifieke groep. Voor geslacht is in de analyse een klein tot middelgroot effect van stereotype dreiging naar voren gekomen, dat maakt dat op dit gebied kan worden ingezet op onderzoek naar interventies die het effect van stereotype dreiging. Waar uit onderzoek is gebleken dat het inzetten van integratief STEM-onderwijs helpt met het verbeteren van de STEM-prestaties van alle leerlingen (Belland et al.,

2016; D'Angelo et al., 2014; Kazu & Kurtoglu Yalcin, 2021), is het nu van belang om verder te kijken naar interventies die specifiek het effect van stereotype dreiging tegengaan. Het doel daarbij zou kunnen zijn om de gemiddelde verschillen in leerprestaties tussen jongens en meisjes in STEM-vakgebieden te verkleinen.

Dit onderzoek geeft aanleiding te verwachten dat stereotype dreiging mogelijk niet beperkt blijft tot de wiskunde. In voorgaand onderzoek naar stereotype dreiging bij meisjes had wiskunde echter vaak de focus (Flore & Wicherts, 2015; Nguyen & Ryan, 2008; Spencer et al., 1999). Gezien nu blijkt dat stereotype dreiging in het hele STEM-gebied een klein tot middelgroot effect lijkt te hebben, zou gesteld kunnen worden dat deze focus op wiskunde in onderzoek naar (het tegengaan van) stereotype dreiging te beperkt is. Daarmee is het aan te raden om in onderzoek naar stereotype dreiging een bredere focus aan te nemen. De verschillende gebruikte uitkomstmaten in de geïnccludeerde artikelen, naast wiskunde bijvoorbeeld ook scheikunde, natuurkunde en vaardigheid in programmeren, maken duidelijk dat er op meer gebieden sprake kan zijn van stereotype dreiging. Het is zowel van theoretisch als van maatschappelijk belang te ontdekken waar stereotype dreiging de grootste invloed heeft op de leerprestaties, in zowel het primair als het secundair onderwijs, zodat gericht gekeken kan worden naar maatregelen om de negatieve effecten van stereotype dreiging te voorkomen en terug te dringen. Als stereotype dreiging voorkomen kan worden heeft dat mogelijk een positief effect op de verdeling van meisjes en jongens en van etnische minderheden en de etnische meerderheid in STEM-beroepen.

Literatuurlijst

- Ambady, N., Paik, S. K., Steele, J., Owen-Smith, A. & Mitchell, J. P. (2004). Deflecting negative self-relevant stereotype activation: The effects of individuation. *Journal of Experimental Social Psychology*, *40*(3), 401–408.
<https://doi.org/10.1016/j.jesp.2003.08.003>
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology*, *16*(1), 94–98. <https://doi.org/10.1037/a0017564>
- Beasley, M. A. & Fischer, M. J. (2012). Why they leave: the impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, *15*(4), 427–448. <https://doi.org/10.1007/s11218-012-9185-3>
- Becker, K. & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education*, *12*(5 & 6), 23–37.
<https://www.jstem.org/jstem/index.php/JSTEM/article/download/1509/1394>
- Belland, B. R., Walker, A. E., Kim, N. J. & Lefler, M. (2016). Synthesizing Results From Empirical Research on Computer-Based Scaffolding in STEM Education. *Review of Educational Research*, *87*(2), 309–344. <https://doi.org/10.3102/0034654316670999>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (1ste editie). Wiley.
- Corlu, M. S., Caparo, R. M. & Caparo, M. M. (2014). Introducing STEM education: Implications for educating our teachers for the age of innovation. *Educational and Science*, *39*(171), 74–85. <http://repository.bilkent.edu.tr/handle/11693/13203>

- D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., Haertel, G. (2014). *Simulations for STEM Learning: Systematic Review and Meta-Analysis*. Menlo Park, CA: SRI International.
- Flore, P. C. & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*(1), 25–44. <https://doi.org/10.1016/j.jsp.2014.10.002>
- *Franklin, D., & Rangel, V. S. (2022) Estimating the Effect of Single-Sex Education on Girls' Mathematics and Science Achievement. *Leadership and Policy in Schools*, Article 2108461. <https://doi.org/10.1080/15700763.2022.2108461>
- Fry, R., Kennedy, B. & Funk, C. (2021). *STEM Jobs See Uneven Progress in Increasing Gender, Racial and Ethnic Diversity*. Pew Research Center. https://www.pewresearch.org/science/wp-content/uploads/sites/16/2021/03/PS_2021.04.01_diversity-in-STEM_REPORT.pdf
- *Gunderson, E. A., Park, D., Maloney, E. A., Beilock, S. L., & Levine, S. C. (2018). Reciprocal relations among motivational frameworks, math anxiety, and math achievement in early elementary school. *Journal of Cognition and Development, 19*(1), 21-46. <https://doi.org/10.1080/15248372.2017.1421538>
- *Huguet, P., & Regner, I. (2007). Stereotype threat among schoolgirls in quasi-ordinary classroom circumstances. *Journal of Educational Psychology, 99*(3), 545-560. <https://doi.org/10.1037/0022-0663.99.3.545>
- Kazu, I. Y. & Kurtoglu Yalcin, C. (2021). The Effect of STEM Education on Academic Performance: A Meta-Analysis Study. *Turkish Online Journal of Educational Technology, 20*(4), 101–116. <https://files-eric-ed-gov.proxy-ub.rug.nl/fulltext/EJ1313488.pdf>

- *Kurtz-Costes, B., Rowley, S. J., Harris-Britt, A., & Woods, T. A. (2008). Gender Stereotypes about Mathematics and Science and Self-Perceptions of Ability in Late Childhood and Early Adolescence. *Merrill-Palmer Quarterly: Journal of Developmental Psychology*, 54(3), 386-409. <https://www.jstor.org/stable/23096251>
- Lenhard, W. & Lenhard, A. (2016). Computation of effect sizes. *Psychometrica*. https://www.psychometrica.de/effect_size.html. DOI: 10.13140/RG.2.2.17823.92329
- Lent, R. W., Sheu, H., Miller, M. J., Cusick, M., Penn, L. T., & Truong, N. N. (2018). Predictors of science, technology, engineering, and mathematics choice options: A meta-analytic path analysis of the social-cognitive choice model by gender and race/ethnicity. *Journal of Counseling Psychology*, 65(1), 17–35. <https://doi.org/10.1037/cou0000243>
- McComas, W. F. & Burgin, S. R. (2020). A Critique of “STEM” Education. *Science & Education*, 29(4), 805–829. <https://doi.org/10.1007/s11191-020-00138-2>
- *Montuori, C., Ronconi, L., Vardanega, T., & Arfe, B. (2022). Exploring Gender Differences in Coding at the Beginning of Primary School. *Frontiers in Psychology*, 13, Article 887280. <https://doi.org/10.3389/fpsyg.2022.887280>
- Nguyen, H. H. D. & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology*, 93(6), 1314–1334. <https://doi.org/10.1037/a0012702>
- Nugent, G., Barker, B., Welch, G., Grandgenett, N., Li, V. C., & Nelson, C. A. (2015). A Model of Factors Contributing to STEM Learning and Career Orientation. *International Journal of Science Education*, 37(7), 1067–1088. <https://doi.org/10.1080/09500693.2015.1017863>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J.,

- Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, n71. <https://doi.org/10.1136/bmj.n71>
- Parker, P. D., Schoon, I., Tsai, Y. M., Nagy, G., Trautwein, U. & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental Psychology*, 48(6), 1629–1642. <https://doi.org/10.1037/a0029167>
- *Pennington, C. R., Kaye, L. K., Qureshi, A. W., & Heim, D. (2018). Controlling for Prior Attainment Reduces the Positive Influence that Single-Gender Classroom Initiatives Exert on High School Students' Scholastic Achievements. *Sex Roles*, 78(5-6), 385-393. <https://doi.org/10.1007/s11199-017-0799-y>
- Picho-Kiroga, K., Turnbull, A. & Rodriguez-Leahy, A. (2021). Stereotype Threat and Its Problems: Theory Misspecification in Research, Consequences, and Remedies. *Journal of Advanced Academics*, 32(2), 231–264. <https://doi.org/10.1177/1932202x20986161>
- *Plant, E. A., Baylor, A. L., Doerr, C. E., & Rosenberg-Kima, R. B. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education*, 53(2), 209-215. <https://doi.org/10.1016/j.compedu.2009.01.013>
- *Randolph, K. Y. (2018). *The care factor: The importance of using domain identification measures to identify stereotype threat vulnerability in high school females* [Thesis]. Virginia Commonwealth University. <https://doi.org/10.25772/G8T9-C590>
- *Shenouda, C. K. (2015). *Effects of gender stereotypes on children's beliefs, interests, and performance in stem fields* [Thesis]. Michigan State University. <https://doi.org/doi:10.25335/M5NQ87>

- *Shenouda, C. K., & Danovitch, J. H. (2014). Effects of gender stereotypes and stereotype threat on children's performance on a spatial task. *Revue Internationale De Psychologie Sociale-International Review of Social Psychology*, 27(3-4), 53-77.
<https://www.cairn-int.info/journal--2014-3-page-53.htm>.
- Song, F., Parekh, S., Hooper, L., Loke, Y. K., Ryder, J., Sutton, A. J., ... & Harvey, I. (2010). Dissemination and publication of research findings: an updated review of related biases. *Health technology assessment*, 14(8), 1-220. <https://doi.org/10.3310/hta14080>
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 35(1), 4–28.
<https://doi.org/10.1006/jesp.1998.1373>
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797–811.
<https://doi.org/10.1037/0022-3514.69.5.797>
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613–629. <https://doi.org/10.1037/0003-066x.52.6.613>
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J. R., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, 343(jul22 1), d4002.
<https://doi.org/10.1136/bmj.d4002>
- Techniekpact. (2021). Vrouwen en de keuze voor bètatechnische opleidingen en loopbanen. In *Techniekpact*. <https://admin.techniekpact.nl/uploads/techniekpact/originals/33768687-57ac-4457-8e3a-cac8db30af3a.pdf>

- *Tellhed, U., Bjorklund, F., & Strand, K. K. (2022). Sure I can code (but do I want to?). Why boys' and girls' programming beliefs differ and the effects of mandatory programming education. *Computers in Human Behavior*, *135*, Article 107370.
<https://doi.org/10.1016/j.chb.2022.107370>
- Wang, M.-T., Degol, J., & Ye, F. (2015). Math achievement is important, but task values are critical, too: examining the intellectual and motivational factors leading to gender disparities in STEM careers [Original Research]. *Frontiers in Psychology*, *6*.
<https://doi.org/10.3389/fpsyg.2015.00036>
- *Wendt, J. L., Rockinson-Szapkiw, A., & Cordes, M. (2018). Examining the influence of a STEM certification model on female, minority science outcomes. *Journal of Educational Research*, *111*(6), 666-677. <https://doi.org/10.1080/00220671.2017.1396437>
- *Wille, E., Gaspard, H., Trautwein, U., Oschatz, K., Scheiter, K., & Nagengast, B. (2018). Gender Stereotypes in a Children's Television Program: Effects on Girls' and Boys' Stereotype Endorsement, Math Performance, Motivational Dispositions, and Attitudes. *Frontiers in Psychology*, *9*, Article 2435. <https://doi.org/10.3389/fpsyg.2018.02435>
- *Wulff, P., Hazari, Z., Petersen, S., & Neumann, K. (2018). Engaging young women in physics: An intervention to support young women's physics identity development. *Physical Review Physics Education Research*, *14*(2), Article 020113.
<https://doi.org/10.1103/PhysRevPhysEducRes.14.020113>
- *Zhao, F., Zhang, Y., Alterman, V., Zhang, B., & Yu, G. (2018). Can math-gender stereotypes be reduced? A theory-based intervention program with adolescent girls. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues*, *37*(3), 612-624. <https://doi.org/10.1007/s12144-016-9543-y>