

Beschouwing

Evidence informed werken op de lerarenopleiding? Over ruis, onzekerheid en complexiteit

Symen van der Zee (Saxion Hogeschool)

Samenvatting

Evidence informed werken, de toepassing van effectieve onderwijsprincipes en aanpakken, klinkt logisch. Waarom zou je als leerkracht immers niet doen 'wat werkt'? Hoewel veel bekend is over 'wat werkt', menen pleitbezorgers van evidence informed onderwijs dat de onderwijspraktijk teveel ideologie en intuïtie volgt. In lerarenopleidingen zou eveneens te weinig oog zijn voor effectieve onderwijsprincipes. Hierdoor, zo wordt gesteld, blijft het onderwijs doormodderen. Of het zo simpel is, valt te betwisten. Is de kennis uit effectonderzoek werkelijk zo stevig? Maakt het echt duidelijk wat zou moeten in het onderwijs? In deze bijdrage betwist ik de zekerheid waarmee over evidence wordt gesproken op basis van de uitkomsten van meta-onderzoek. De bevindingen hieruit trekken de status van de onderwijs-wetenschappelijke kennisstand in twijfel en roepen fundamentele vragen op over de relatie die wordt gelegd tussen kennis over effectiviteit en de verbetering van de onderwijspraktijk. De consequenties hiervan voor het opleiden van leraren worden beschouwd.

Weten wat werkt en zou moeten

Evidence informed werken is de afgelopen decennia populair geworden in het onderwijs. Het idee ervan is simpel: de wetenschap ontdekt wat effectief is, 'wat werkt', en de onderwijspraktijk past toe wat effectief is (Onderwijsraad, 2006). Niet doen 'wat werkt' en wel doen 'wat niet werkt' zou hoofdoorzaak zijn van veel van de problemen in het onderwijs (zie bijvoorbeeld Scheerens & Kirschner, 2021). Volgens Kirschner (2023) leeft het onderwijs in een romantisch avondland, waar filosofie, ideologie, intuïtie en mooie verhalen boven "wat wij weten" staan. Hij stelt: "Het is niet een kwestie of wij wel of niet weten wat zou moeten of dat er meer onderzoek nodig is om daarachter te komen maar dat veel mensen dit niet willen accepteren en dus blijft het onderwijs doormodderen." In klare taal wordt daarom opgeroepen te stoppen met experimenteel onderwijs (NRC, 2023), onderwijsmythes de deur uit te doen (Muijs, 2022) en leerlingen vooral niet ontdekkend te laten leren (Visser, 2017). En in net zulke klare taal wordt gepleit voor expliciete directe instructie, want dat is 'wat werkt' (zie bijvoorbeeld Hollingsworth et al., 2020).

Evidence informed werken klinkt logisch. Waarom zou je als leerkracht niet doen 'wat werkt' en wel iets doen waarvan je niet weet of het werkt of waarvan je weet dat het niet werkt? Het logische karakter ervan wordt door pleitbezorgers van evidence informed werken ook beklemtoond. Het is immers duidelijk, zoals Kirschner (2023) aangeeft, wat zou moeten. Ondanks deze duidelijkheid toont onderzoek van Surma et al., (2018) aan dat de effectief bewezen strategieën 'retrieval practice' en 'distributed practice' vrijwel niet worden behandeld in de leerboeken van lerarenopleidingen in Nederland en Vlaanderen. Daar moet dan ook verandering inkomen. Immers, zoals Surma et al. (2018, p.235) concluderen: "Future teachers should be well prepared to teach with good, evidence-informed knowledge of how students learn."

10 redenen voor twijfel

Kennis uit effectonderzoek wordt in het publieke debat vaak gepresenteerd als onbetwistbare zekerheden die door de praktijk geaccepteerd en toegepast moeten worden. Het doormodderen, zoals Kirschner stelt, moet stoppen. Vraag is of het zo simpel is. Is de kennis uit effectonderzoek werkelijk zo stevig? Maakt het echt duidelijk wat zou moeten? In deze bijdrage betwist ik de zekerheid waarmee over evidence wordt gesproken, door 10 redenen voor twijfel uit te werken. Ik baseer mij hierbij op resultaten uit meta-onderzoek, onderzoek met als studieobject onderzoek, en beperk mijzelf tot technische kritieken. Wijsgerige mitsen en maren blijven dus buiten beschouwing (zie hiervoor bijvoorbeeld Biesta, 2022). De 10 redenen die achtereenvolgens worden besproken, zijn:

1. Meer dan de helft van het onderzoek bevat ontoelaatbare praktijken en 1 op 18 onderzoekers fabriceert of vervalst zelfs onderzoeksgegevens.
2. Als de ontwikkelaar het onderzoek doet, worden veel grotere effecten gevonden.
3. Grote positieve effecten kunnen net zo goed grote negatieve effecten zijn.
4. Uitkomsten van enkele experimenten zeggen maar heel weinig.
5. Cognitieve leerstrategieën die werken in het lab, werken niet zonder meer in de onderwijspraktijk.
6. De uitkomsten van veel gerandomiseerde veldexperimenten zijn niet informatief.
7. Gerandomiseerde veldexperimenten die veelbelovende resultaten tonen, overschatten de effecten.

8. Maar 1 op de 500 studies in de onderwijswetenschappen is een replicatieonderzoek.
9. Alleen positieve resultaten worden nog gepubliceerd.
10. Reviews overschatten de effecten van aanpakken waarschijnlijk enorm.

Doel van de uiteenzetting is een meer realistische kijk op de waarde van kennis over onderwijseffectiviteit te faciliteren en een vruchtbare dialoog hierover in lerarenopleidingen te stimuleren.

1. Een groot deel van het onderzoek bevat ontoelaatbare praktijken

Hoewel wordt gesteld dat we dankzij effectonderzoek weten wat werkt en wat in de praktijk moet, is niet al het onderzoek goed en wordt het niet altijd uitgevoerd zoals het moet. Onderzoekers kunnen op oneigenlijke manieren onderzoeksprocessen en uitkomsten beïnvloeden. Ze kunnen de analyse en presentatie van data bijvoorbeeld op tal van manieren plooien naar de eigen wensen, de kennisstand selectief weergeven, niet helemaal eerlijk zijn over de onderzoeksopzet, bewust bepaalde zaken achterhouden, enzovoorts. (Andrade, 2021). Blijkens de studie van Gopalakrishna et al., (2022) naar wetenschappelijke integriteit onder Nederlandse onderzoekers komen dit soort praktijken ook zorgwekkend vaak voor. Maar liefst de helft (50,2%) van de ondervraagde onderzoekers uit de sociale wetenschappen geeft aan in de afgelopen twee jaren iets te hebben gedaan dat niet door de beugel kan. 1 op de 18 onderzoekers geeft zelfs aan data te hebben gefabriceerd of vervalst. Omdat het hier om zelfrapportage gaat, is de kans bovendien groot dat het in de werkelijkheid om grotere aantallen en percentages gaat (Bekkers, 2022). De uitkomsten roepen een belangrijke vraag op. Als een aanzienlijk deel van het onderzoek ontoelaatbare praktijken bevat en 1 op de 18 wetenschappers zelfs data fabriceert of vervalst, hoeveel vertrouwen kunnen we dan hebben in de kennisstand over effectief onderwijs?

2. Als de ontwikkelaar het onderzoek doet, worden veel grotere effecten gevonden

Dat onderzoekers de analyse en presentatie van data kunnen plooien naar de eigen wensen en zo ook de conclusies van onderzoek kunnen sturen, is ook wat de studie van Wolf et al. (2019) suggereert. Hierin is gekeken naar de effectgroottes in het onderwijskundig onderzoek van de afgelopen dertig jaar en is nagegaan of het uitmaakt wie het onderzoek doet. Specifiek is gekeken of het ertoe doet of de ontwikkelaar van de onderwijsaanpak betrokken is bij het

onderzoek ernaar. In totaal zijn 170 studies op het gebied van taal en rekenen hiervoor onderzocht. Wat blijkt: als de ontwikkelaar betrokken is, zijn de gevonden effecten haast anderhalf keer zo groot. Anders uitgedrukt, als de ontwikkelaar betrokken is, zijn de leerresultaten van de onderwijsinterventie 70% hoger dan wanneer onafhankelijke onderzoekers het onderzoek doen. Als vooral de ontwikkelaars van onderwijsaanpakken het onderzoek doen of erbij betrokken zijn, hoe betrouwbaar is de kennisstand dan?

3. Het effect hangt af van wie er analyseert

De bevinding van Wolf et al. (2019) betekent niet automatisch dat ontwikkelaars van onderwijsaanpakken twijfelachtige onderzoekspraktijken toepassen. Want ook wanneer onderzoekers niet op oneigenlijke manieren onderzoeksprocessen sturen, kunnen onderzoeksresultaten beïnvloed worden. In de studie van Breznau et al., (2022) kregen 73 onderzoeksteams exact dezelfde dataset en werd ze gevraagd dezelfde hypothese te toetsen. Verwacht werd dat de teams tot min of meer dezelfde resultaten en conclusies zouden komen. Niets bleek minder waar te zijn. De resultaten liepen enorm uiteen. Van een groot negatief effect tot een groot positief effect. Eerdere onderzoeken kwamen tot dezelfde conclusie (Botvinik-Nezer et al., 2020; Landy et al, 2020; Menkveld et al., 2022; Silberzahn et al., 2018). Om te achterhalen hoe het komt dat de resultaten zo verschillen, hebben Breznau en collega's alle keuzes van de onderzoeksteams in kaart gebracht en vergeleken. In totaal werden 107 besluiten vergeleken, maar de verschillen in besluiten leverden geen verklaringen op voor de verschillen in resultaten. Breznau en collega's concluderen daarom: "...a hidden universe of uncertainty remains." Dit levert een fundamentele vraag op. Als een groot positief effect net zo goed een groot negatief effect kan zijn en het onduidelijk is waarom dit het geval is, wat betekent dit dan voor de wetenschappelijke kennisstand over effectief onderwijs? En bij iedere individuele studie is de vraag dus legitiem: als een ander onderzoeksteam deze data had geanalyseerd, hadden we dan niet iets heel anders van het onderzoek geleerd?

4. Eén of enkele experimenten zeggen weinig

Bij het onderzoek van Breznau et al., (2022) moet opgemerkt worden dat het gaat om de analyse van een grote dataset. Er zijn dan veel keuzes te maken, dus is de kans groter dat de resultaten van onderzoeksgroepen uiteenlopen. In experimentele studies zou dit probleem mogelijk minder spelen, aangezien hierin specifieke vragen worden onderzocht en alleen daarover data wordt verzameld (Breznau et al., 2022). Recent onderzoek van Huber et al. (2023) toont

echter een vergelijkbaar punt van zorg in het experimentele onderzoek. Doel van de studie was nagaan wat het effect van competitie is op moreel gedrag. Omdat eerder onderzoek geen consensus had opgeleverd, vroegen ze groepen onderzoekers een online-experiment hiervoor op te zetten. Van de 95 ingediende experimenten werden er 45 uitgevoerd, met in totaal 18.123 personen. Uiteindelijk vond men over al deze experimenten een klein negatief effect van competitie op moreel gedrag. Maar dat was niet de belangrijkste bevinding. Die had namelijk betrekking op de invloed van de inrichting van de experimenten op de gevonden effecten. De gevonden effecten varieerden namelijk sterk. In de studie werd daarom gekeken of kwaliteitsverschillen tussen de experimenten de gevonden verschillen in effecten zouden kunnen verklaren. Dit bleek niet het geval te zijn. Ook werd gekeken of verschillende ontwerpkeuzes de verschillen in gevonden effecten konden verklaren. Dit bleek eveneens niet het geval. Hoe het komt dat de resultaten zo uiteenlopen, is dus onduidelijk. Huber en collega's concluderen daarom dat de generaliseerbaarheid en informatiewaarde van één of enkele experimenten zeer beperkt is en roepen op tot veel grotere dataverzameling en de inzet van een groot aantal onderzoeksteams. Hoewel deze aanbeveling ter harte genomen moet worden, worden de conclusies van één of enkele studies momenteel nog met regelmaat als feit gepresenteerd richting de onderwijspraktijk en worden aanbevelingen op basis ervan gedaan. De vraag die dit oplevert luidt: hoe zeker zijn de uitspraken over onderwijseffectiviteit?

5. Cognitieve leerstrategieën die werken in het lab, werkt niet zondermeer in de onderwijspraktijk

De studie van Huber et al., (2023) heeft betrekking op experimenten in een laboratoriumachtige setting. Een bekende beperking van dergelijk onderzoek is dat de uitkomsten ervan niet zonder meer geldig zijn buiten die setting. Uit laboratoriumachtig experimenteel onderzoek kan bijvoorbeeld blijken dat bepaalde leertechnieken werken bij studenten psychologie die een vergoeding krijgen voor deelname, maar dit betekent niet automatisch dat ze ook werken bij jonge kinderen in een drukke kleuterklas. Hier is niet altijd voldoende oog en aandacht voor. Neem het cognitief psychologisch onderzoek. In dit onderzoeksgebied is veel evidentie te vinden voor de effectiviteit van de toepassing van cognitieve strategieën op het leren (*spaced learning, dual coding, interleaving, etc.*). Dit betekent echter niet zonder meer dat onderwijsinterventies die hierop gebaseerd zijn in de dagelijkse basisschoolpraktijk ook werken.

Sterker, een reviewstudie van Perry et al., (2022, p. 7) concludeert dat hier nog maar weinig bewijs voor is. Het onderzoek dat is gedaan, is over het algemeen klein van opzet, kort van duur en heeft maar binnen enkele vakgebieden plaatsgevonden. Daar komt bij dat de rol van de leerkracht vrijwel niet is onderzocht en de leerkracht in de studies juist vaak buiten spel is gezet. De vier beste studies die zijn verricht, dus de studies die wel grootschalig en langdurig waren en waar leerkrachtprofessionalisering wel onderdeel van was, vonden zelfs weinig tot geen effect. In één daarvan, die van Schunn et al., (2018), werden zelfs negatieve effecten gevonden. Ondanks het tekort aan evidence voor de effectiviteit wordt de inzet van strategieën uit de cognitieve leerpsychologie in evidence informed cursussen, trainingen en lezingen benoemd als noodzakelijk voor effectief onderwijs. De vraag die hieruit voortvloeit luidt: wanneer legitimeert welke evidence uitspraken over wat zou moeten in de praktijk?

6. Gerandomiseerde veldexperimenten zijn vaak niet informatief

Vanwege de voorgenoemde beperkingen zijn er naast laboratorium experimenten ook gerandomiseerde veldexperimenten. Dit zijn experimenten die in de daadwerkelijke onderwijspraktijk plaatsvinden. Instanties als de Engelse 'Education Endowment Foundation' (EEF) en de Amerikaanse Center for 'Educational Evaluation and Regional Assistance' (EERA) hebben fors geïnvesteerd in dit soort studies. Helaas blijken deze grootschalige gerandomiseerde experimenten vaak niet informatief te zijn. Lortie-Forgues en Inglis (2019) analyseerden de uitgevoerde gerandomiseerde experimenten en concluderen dat de resultaten ervan te onzeker zijn: gevonden effecten zijn klein en variëren sterk. Hierdoor kan er vaak geen betekenisvolle conclusie worden getrokken over de effectiviteit van de aanpak. Het kan namelijk zijn dat de aanpak werkt, maar het kan ook zijn dat de aanpak juist een negatief effect heeft. Op zijn scherpst gesteld zijn veel gerandomiseerde veldexperimenten onvoldoende informatief en kan de onderwijspraktijk er niks mee. De vraag is vervolgens ook of dit type onderzoek wel de moeite en de kosten waard is. Want dergelijke studies zijn duur. De Educational Endowment Foundation heeft er bijvoorbeeld al meer dan 115 miljoen pond in dergelijk onderzoek geïnvesteerd. Is investeren hierin wel kosteneffectief?

7. Veelbelovende studies overschatten de effecten

De reden dat veel gerandomiseerde veldexperimenten niet informatief zijn, heeft te maken met een gebrek aan zogeheten statistische power. Als een onderzoek te weinig power heeft, is de kans klein dat het werkelijke effect van

een aanpak wordt gevonden en is de kans groter dat de resultaten vertekend zijn door willekeurige en systematische fouten. Sims et al. (2022) tonen aan dat een deel van de gerandomiseerde experimenten veelbelovende effecten rapporteert, maar deze effecten vertekend zijn door het gebrek aan statistische power. Ze concluderen dat de effecten in deze studies met maar liefst 52% worden overschat. De vraag is wat de consequenties hiervan zijn. Hoe moeten we de effectgroottes van veelbelovende interventies in de evidence informed menukaarten, toolkits en 'warehouses' interpreteren? Sims et al. (2022) raden aan om er in ieder geval vanuit te gaan dat de effecten in werkelijkheid veel kleiner zullen zijn dan dat wordt aangegeven.

8. Maar 1 op de 500 is een replicatiestudie

Gezien de voorgaande bevindingen is het niet verwonderlijk dat de sociale wetenschappen te kampen hebben met een replicatiecrisis. In 2015 kwam deze crisis voor het voetlicht toen de Open Science Collaboration maar liefst 100 experimentele studies uit drie vooraanstaande wetenschappelijke tijdschriften op het gebied van psychologie gingen repliceren. Van de 100 studies had 97% een significant resultaat in het oorspronkelijke onderzoek, met een gemiddeld effect. In de onafhankelijk replicaties had nog maar 37% een significant effect en waren de gevonden effecten opeens klein. Of de onderwijswetenschappen net als de psychologie in een replicatiecrisis verkeert is onduidelijk. Dit komt omdat er bijna geen replicatieonderzoek wordt gedaan. Perry et al. (2022, p. 27) hebben een review hiernaar verricht en concluderen dat slechts één op de 500 studies in de onderwijswetenschappen een replicatiestudie is. De consequenties hiervan voor het idee en de praktijk van evidence informed werken zijn volgens Perry et al. (2022) verstrekkend. Als er geen echte pogingen tot replicatie worden gedaan, hoeveel vertrouwen mag je als leerkracht dan hebben in de kennisstand over effectief onderwijs?

9. Alleen positieve resultaten worden nog gepubliceerd

Een probleem dat samenhangt met de replicatiecrisis is dat positieve resultaten eerder en vaker worden gepubliceerd dan negatieve resultaten. Fanelli (2012) analyseerde 4600 gepubliceerde studies waarin een hypothese werd getoetst in de tijdspanne 1990 en 2007 en concludeert dat in de loop der tijd steeds vaker positieve resultaten worden gerapporteerd. Dit geldt met name voor de sociale wetenschappen, waartoe de onderwijswetenschappen behoren. De titel van het artikel beschrijft treffend de hoofdconclusie: *"Negative results are disappearing from most disciplines and countries"*. Gevolg van deze trend is dat de

replicatiegraad en de betrouwbaarheid van gepubliceerd onderzoek laag is. Als alleen het onderzoek wordt gepubliceerd dat laat zien dat een aanpak ‘werkt’ en het onderzoek dat aantoont dat deze aanpak niet werkt in de ladekasten blijft liggen, weten we dan echt wat werkt in het onderwijs?

10. Reviews overschatten de effecten

Reviewstudies lijken een oplossing voor het probleem van vrijwel geen replicatieonderzoek. In een review wordt uitvoerig gezocht naar al het beschikbare onderzoek over een aanpak en worden de uitkomsten ervan geïntegreerd met de techniek meta-analyse. Het gemiddelde van de effecten van de afzonderlijke studies bepaalt dan of de aanpak ‘werkt’ of niet. Naast vragen over wat het praktisch nut is van weten dat een aanpak gemiddeld werkt – zonder te weten voor wie wel en niet, onder welke omstandigheden, waardoor en waarvoor precies – is het de vraag of deze werkwijze een oplossing is voor het probleem van weinig replicatieonderzoek. Kvarven et al. (2020) hebben in het veld van de psychologie bijvoorbeeld de resultaten van herhaalde experimenten vergeleken met die van de uitkomsten van meta-analyses. De conclusie is dat meta-analyses veel grotere effecten rapporteren, tot wel drie keer groter dan replicatiestudies. Het kan dus goed zijn dat meta-analyses de effecten systematisch overschatten (zie ook Frankenbach, 2020). Dit levert nog een punt van zorg op. Als enkel positieve resultaten worden gerapporteerd en er nagenoeg geen replicatieonderzoek wordt gedaan, maar er worden wel veel reviews met meta-analyses geproduceerd, worden de effecten van aanpakken dan niet enorm overschat?

Onzekerheid en nederigheid

In de evidence informed retoriek wordt de kennis uit onderwijswetenschappelijk effectonderzoek vaak gepresenteerd als waar en zeker en wordt gesteld dat de onderwijspraktijk moet doen ‘wat werkt’ en lerarenopleidingen moeten aanleren ‘wat werkt’. In deze bijdrage heb ik deze gangbare manier van spreken en redeneren betwist. De gepresenteerde bevindingen geven tezamen genoeg aanleiding om de status van de onderwijswetenschappelijke kennisstand in twijfel te trekken en daarmee de relatie die wordt gelegd tussen kennis over effectiviteit en de verbetering van de onderwijspraktijk. De studies roepen fundamentele vragen. Hoe solide is de wetenschappelijke basis die leraren dient te informeren? Hoeveel evidence is er eigenlijk dat evidence informed werken werkelijk werkt? Zijn wetenschappers wel echt de ‘experts’ die weten wat werkt

in de klas? Het is tijd om de evidence informed retoriek aan te passen. In plaats van wetenschappelijke kennis te presenteren als waar, zeker en onbetwistbaar, zou het goed zijn om systematisch de mitsen, maren en manco's ervan te belichten en duidelijk te zijn over de veronderstellingen en keuzes die eraan te grondslag liggen. Dat betekent voor lerarenopleidingen dat we studenten inzichten moeten gaan bieden in de aard en beperkingen van onderzoek. De bevindingen in dit stuk kunnen wellicht helpen om het gesprek over de technische beperkingen van onderzoek op gang te helpen. Vanzelfsprekend dienen daarnaast ook de meer wijsgerige punten van zorg aan bod te komen, die in dit stuk buiten beschouwing zijn gelaten. Een nieuwe leereenheid op lerarenopleiding met de titel *'Kritisch denken met en over onderwijsonderzoek'* zou wellicht geen slecht idee zijn.

Voor de helderheid en zekerheid, dit is geen pleidooi tegen evidence informed werken op de lerarenopleiding en in de onderwijspraktijk. En het is ook geen oproep om effectonderzoek links te laten liggen. Het is juist een pleidooi voor meer wetenschappelijk geïnformeerd denken en spreken over de waarde en beperkingen van wetenschap voor de praktijk. Echt evidence informed werken en opleiden betekent systematisch oog hebben voor beperkingen, onzekerheden, ruis en complexiteit, zowel in het onderzoek als in de onderwijspraktijk. Oog daarvoor hebben draagt bij aan nederigheid. En nederigheid draagt bij aan nuance, zakelijkheid en daarmee, uiteindelijk en hopelijk, aan de onderwijs- en onderzoekskwaliteit.

Auteurs

Symen van der Zee is lector Vernieuwend Onderwijs op de Saxion Hogeschool. s.vanderzee@saxion.nl

Referenties

Andrade C. (2021). HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *The Journal of clinical psychiatry*, 82(1), 20f13804. <https://doi.org/10.4088/JCP.20f13804>

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454. <https://doi.org/10.1038/533452a>

Bekkers, R. (2022). Ten Meta Science Insights to Deal With the Credibility Crisis in the Social Sciences. SocArxiv Preprint, <https://osf.io/preprints/socarxiv/rm4p8>

Biesta, G. (2022). *Onderwijsonderzoek. Een onorthodoxe introductie*. Amsterdam: Boom.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F. *et al.* Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* 582, 84–88, <https://doi.org/10.1038/s41586-020-2314-9>

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., Alvarez-Benjumea, A., Andersen, H. K., Auer, D., Azevedo, F., Bahnsen, O., Balzer, D., Bauer, G., Bauer, P. C., Baumann, M., Baute, S., Benoit, V., Bernauer, J., Berning, C., Berthold, A., Żółtak, T. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*, 119(44), e2203150119. <https://doi.org/10.1073/pnas.2203150119>

Fanelli, D. (2012) Negative Results Are Disappearing from Most Disciplines and Countries. *Scientometrics*, 90, 891-904. DOI:10.1007/s11192-012-0925-0

Perry, T., Lea, R., Jørgensen, C. R., Cordingley, P., Shapiro, K., & Youdell, D. (2021). *Cognitive Science in the Classroom*. London: Education Endowment Foundation (EEF). Gedownload van <https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/cognitive-science-approaches-in-the-classroom>

Friese, M., & Frankenbach, J. (2020). *p*-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, 25(4), 456-471. <https://psycnet.apa.org/doiLanding?doi=10.1037%2Fmet0000246>

Gopalakrishna, G., Wicherts, J. M., Vink, G., Stoop, I., van den Akker, O. R., Ter Riet, G., & Bouter, L. M. (2022). Prevalence of responsible research practices among academics in The Netherlands. *F1000Research*, 11, 471. <https://doi.org/10.12688/f1000research.110664.2>

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., Abellán, M., Adayeva, X., Ay, F. C., Barron, K., Berry, Z., Bönte, W., Brütt, K., Bulutay, M., Campos-Mercade, P., Cardella, E., Claassen, M. A., Cornelissen, G., Dawson, I. G. J., Holzmeister, F. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences of the United States of America*, 120(23), 1-10. <https://www.pnas.org/doi/10.1073/pnas.2215572120>

Hollingsworth, J., Ybarra, S., & Schmeier, M. (2020). *Expliciete Directe Instructie 2.0*. Huizen: PICA.

Kirschner, Paul (2023). Onderwijs leeft in een romantisch avondland. Gedownload van: <https://redhetonderwijs.com/blog/onderwijs-leeft-in-een-romantisch-avondland/>

Kvarven, A., Strømmland, E. & Johannesson, M. (202). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behavior*, 4, 423–434. <https://doi.org/10.1038/s41562-019-0787-z>

Landy, J. F., Jia, M. L., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., Johannesson, M., Pfeiffer, T., Ebersole, C. R., Gronau, Q. F., Ly, A., van den Bergh, D., Marsman, M., Derks, K., Wagenmakers, E. J., Proctor, A., Bartels, D. M., Bauman, C. W., Brady, W. J., Cheung, F., ... Uhlmann, E. L. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological bulletin*, 146(5), 451–479. <https://doi.org/10.1037/bul0000220>

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166. <https://journals.sagepub.com/doi/10.3102/0013189X19832850>

Menkveld, Albert J., Non-Standard Errors (2021). CESifo Working Paper No. 9453, Gedownload van: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3979358

Muijs, Daniel (2021). 7 onderwijsmythes – en waarom het mythes zijn. Gedownload van: <https://www.academica-group.com/kennis/zeven-onderwijsmythes>

NRC (14 mei, 2023). Stop met experimenteel onderwijs, het werkt niet. Gedownload van: <https://www.nrc.nl/nieuws/2023/05/14/stop-met-experimenteel-onderwijs-het-werkt-niet-a4164650>

Onderwijsraad (2006). *Naar meer evidence based onderwijs*. Den Haag: Drukkerij Artoos.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349. <https://osf.io/preprints/psyarxiv/ksfvq>

Perry, T., Morris, R., & Lea, R. (2022). A decade of replication study in education? A mapping review (2011-2020). *Educational Research and Evaluation*, 27(1-2), 12-34. DOI: 10.1080/13803611.2021.2022315

- Scheerens, J. & Kirschner, P. *Progressief achteruit. Zwartboek over de last van slechte ideeën voor het funderend onderwijs*. Gedownload van: <https://www.kirschnered.nl/wp-content/uploads/2021/09/Zwartboek-over-de-Last-van-Slechte-Ideeen-voor-het-Funderende-Onderwijs-progressief.pdf>
- Schunn, C.D., Newcombe, N.S., Alfieri, L., Gromley, J.F., Massey, C., & Merlino, J.F. (2018). Using principles of cognitive science to improve science learning in middle school: What works when and for whom? *Applied Cognitive Psychology* 32(2), 225–240.
- Silberzahn, R., Uhlmann, E.L., Martin, D.P., et al. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*. 1(3), 337-356. doi:10.1177/2515245917747646
- Sims, S., Anders, J., Inglis, M., & Lortie-Forgues, H. (2022). Quantifying ‘promising trial bias’ in randomized controlled trials in education. *Journal of Research on Educational Effectiveness*, doi:10.1080/19345747.2022.2090470.
- Stevens, M. (2020). *The knowledge machine*. London: Allen Lane.
- Surma, T., Vanhoywegen, K., Camp, G., & Kirschner, P. (2018). The coverage of distributed practice and retrieval practice in Flemish and Dutch teacher education textbooks. *Teaching and Teacher Education*, 74, 229-237. doi.org:10.1016/j.tate.2018.05.007.
- Visser, J. (2017). Waarom leerlingen een leraar nodig hebben om te leren. Gedownload van: <https://decorrespondent.nl/6074/waarom-leerlingen-een-leraar-nodig-hebben-om-te-leren/fa5182a3-9b02-051f-1ed6-7d53665fef94>
- Vlaams Instituut Wetenschappelijk Technologisch Aspectenonderzoek (viWTA) (2007). *Onderzoek in onderwijs 8*. Brussel: Vlaams Parlement.
- Wolf, R., Morrison, J., Slavin, R., & Risan, K. (2019). *Do developer-commissioned evaluations inflate effect sizes?* Baltimore, MD: Johns Hopkins University, School of Education.