De complexiteit van evidence informed werken

Symen van der Zee

History _

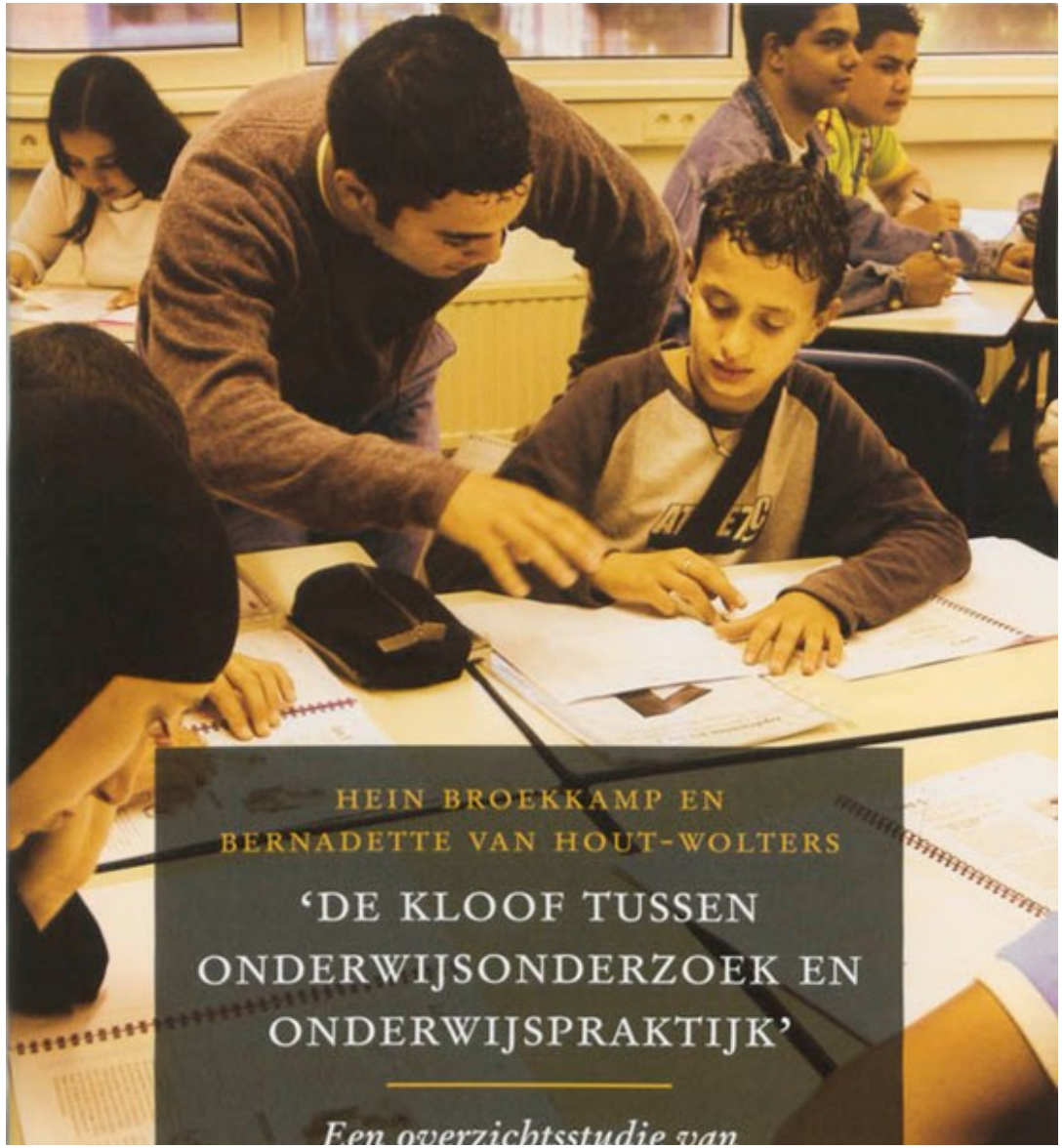
RESEARCH NEWS AND COMMENT

The Awful Reputation of Education Research

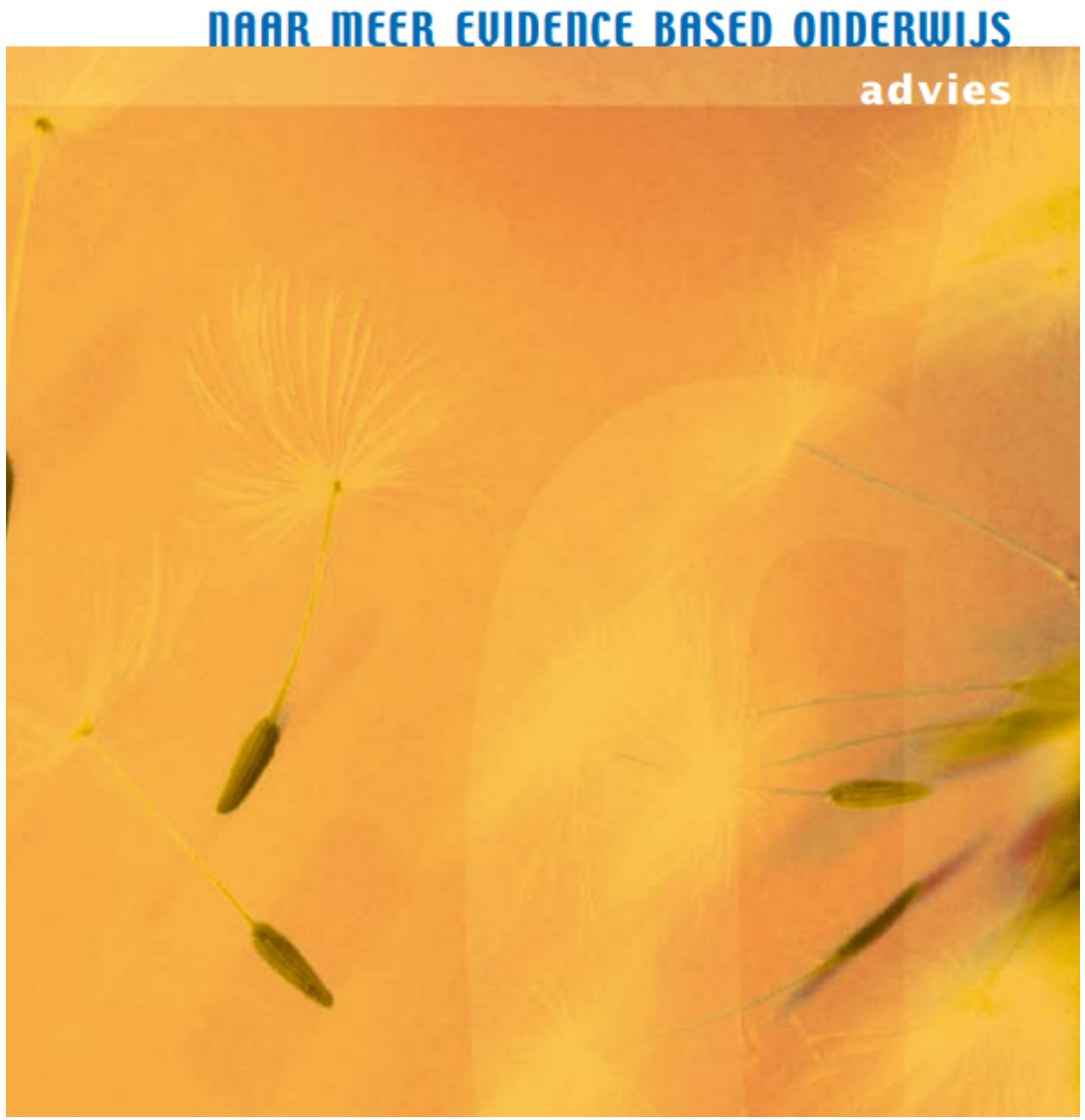
Original Articles

Educational Research and Teaching: a response to David Hargreaves' TTA lecture

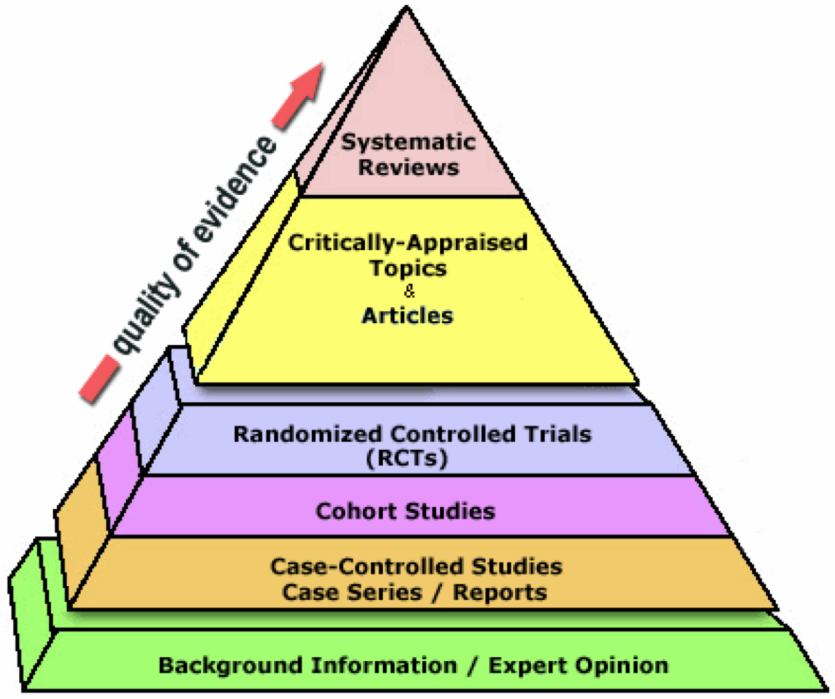
Martyn Hammersley

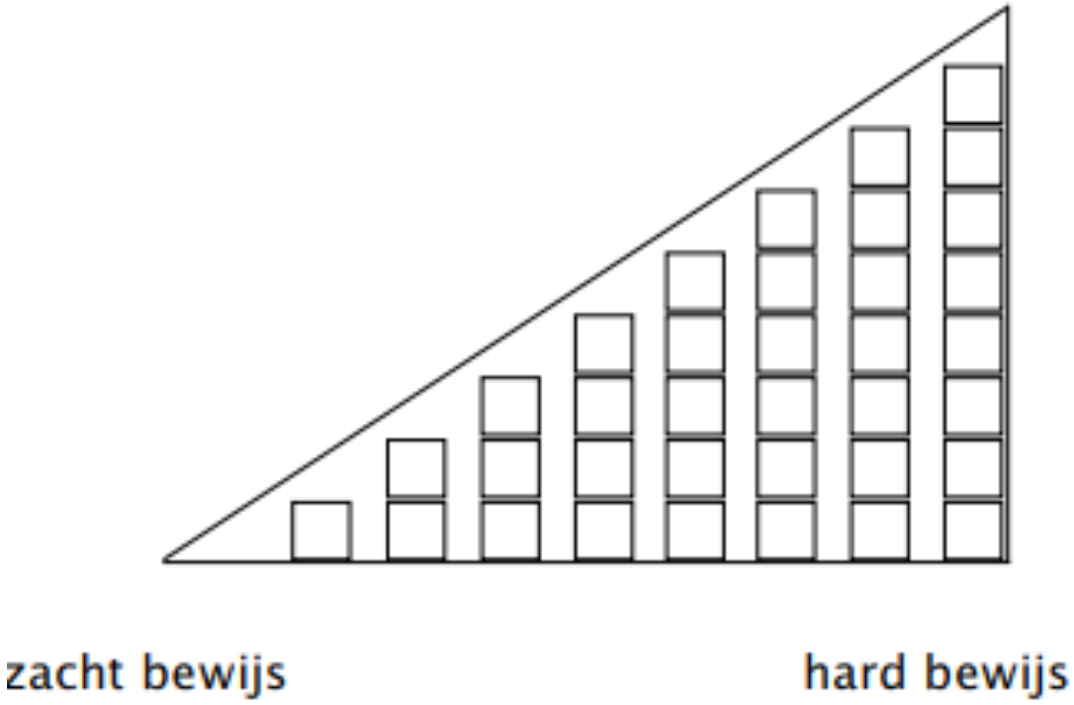

HEIN BROEKKAMP EN
BERNADETTE VAN HOUT-WOLTERS
'DE KLOOF TUSSEN ONDERWIJSONDERZOEK EN ONDERWIJSPRAKTIJK'
Een overzichtsstudie van


STRATION / DAVE CUTLER


NAAR MEER EVIDENCE BASED ONDERWIJS
advies


DE RELATIE TUSSEN ONDERWIJSONDERZOEK EN ONDERWIJSPRAKTIJK IN VLAANDEREN
Studie in opdracht van het viWTA – Samenleving en technologie
vlaams instituut voor wetenschappelijk en technologisch aspectenonderzoek


quality of evidence
Systematic Reviews
Critically-Appraised Topics & Articles
Randomized Controlled Trials (RCTs)
Cohort Studies
Case-Controlled Studies
Case Series / Reports
Background Information / Expert Opinion


zacht bewijs
hard bewijs

***Obs.*: 5957**	**Fixed effects**			**Random effects**		
***ICC*: 27.62%**	**γ**	***SE* γ**	**95% *CI* γ**	**Per**	***SD***	**95% *CI SD***
Intercept	2.12	0.07	[1.98; 2.27]	Id	0.50	[0.47; 0.53]
				Subject	0.23	[0.11; 0.35]
Group	−0.18	0.06	[−0.31; −0.06]	Subject	0.18	[0.07; 0.28]
Method	0.10	0.08	[−0.06; 0.26]	Subject	0.24	[0.11; 0.35]
Group × Method	0.30	0.07	[0.16; 0.44]	Subject	0.20	[0.07; 0.32]
Marginal R^2		0.02		Conditional R^2	0.36	

Obs., number of observations; ICC, intraclass correlation; γ, (unstandardized) regression coefficient; SE, standard error; SD, standard deviation; CI, confidence interval (on 1,000 bootstrapping samples); R^2, coefficient of determination.


Translate
Page Up
Page Down

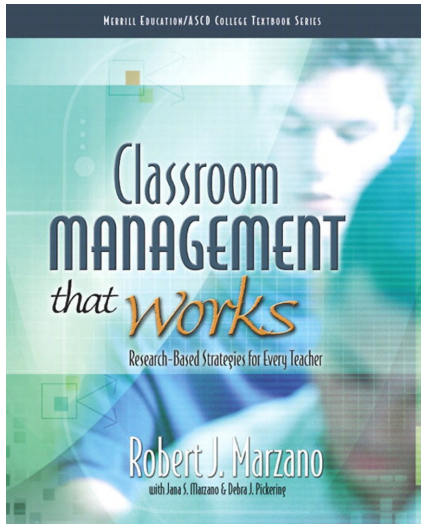
ies
WHAT WORKS
CLEARINGHOUSE

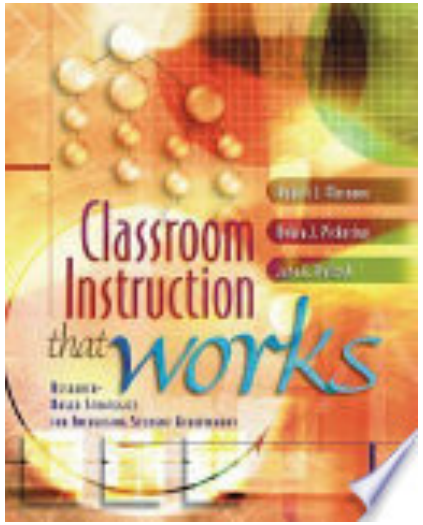
Education
Endowment
Foundation

BEE

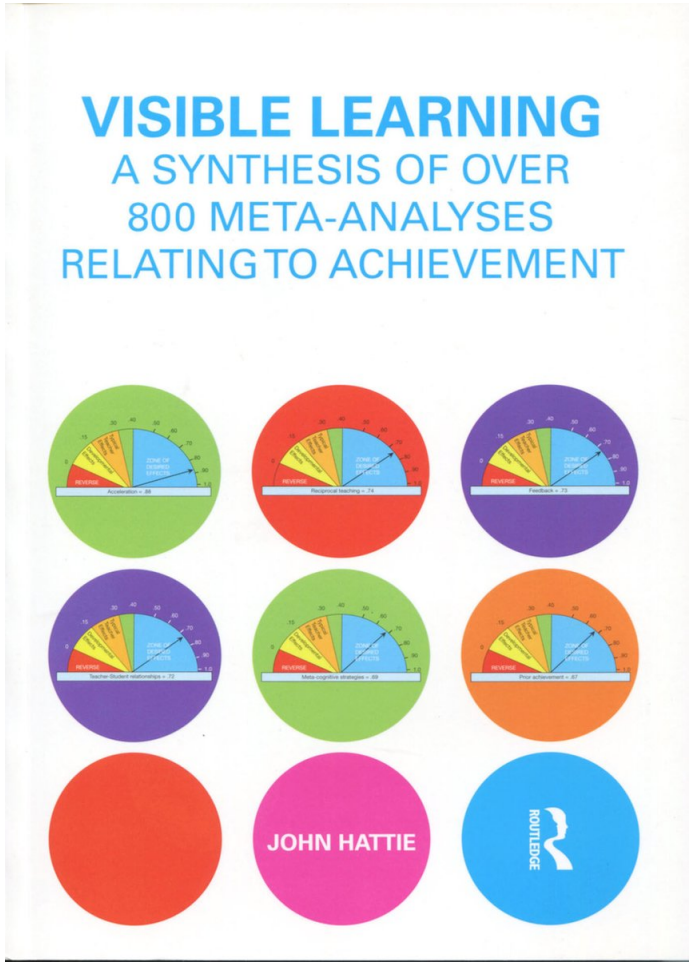
Best Evidence Encyclopedia

Empowering Educators with Evidence on proven programs


Merrill Education/ASCD College Textbook Series
Classroom
MANAGEMENT
that works
Research-Based Strategies for Every Teacher
Robert J. Marzano
with Jana S. Marzano & Debra J. Pickering


Classroom
Instruction
that works


WIJZE LESSEN
Twaalf bouwstenen voor effectieve didactiek
E=mc²
Tim Surma • Kristel Vanhoyweghen • Dominique Sluijsmans
Gino Camp • Daniel Muijs • Paul A. Kirschner


VISIBLE LEARNING
A SYNTHESIS OF OVER
800 META-ANALYSES
RELATING TO ACHIEVEMENT
JOHN HATTIE
ROUTLEDGE

Leergang Onderwijskundig Leiderschap

'Zonder onderwijskundig leiderschap geen goed onderwijs.'

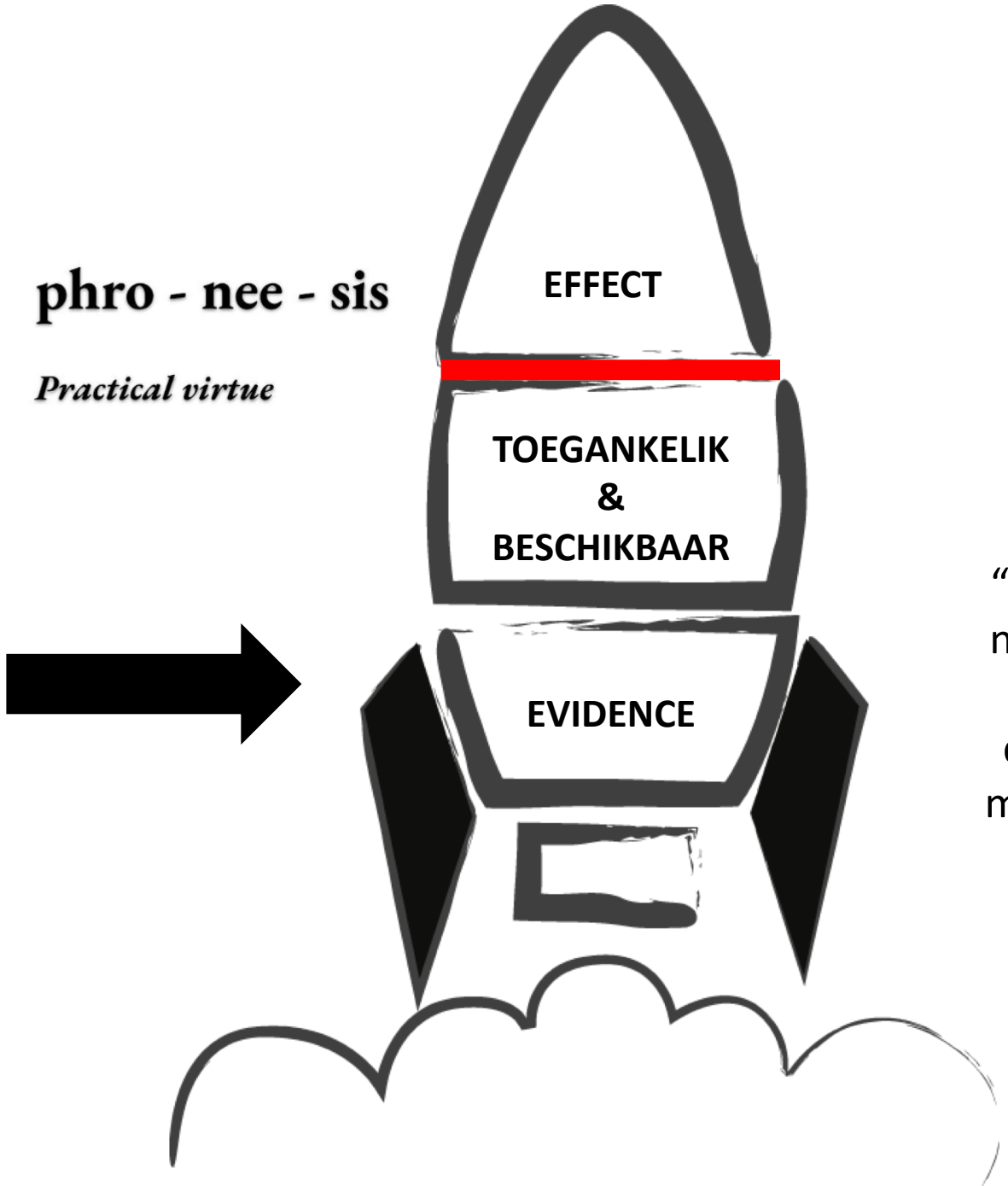
Weet u als leidinggevende in het onderwijs voldoende van de recente wetenschappelijke literatuur om te kunnen sturen op onderwijskwaliteit? Wilt u uw visie op onderwijs verder onderbouwen met wetenschappelijke inzichten of wilt u het onderwijs bij u op school meer evidence-informed vormgeven? Met de Leergang Onderwijskundig Leiderschap bouwt u aan krachtig én onderbouwd onderwijs binnen uw school of scholen.

De evidence-informed leraar

Onderdeel van Programma Excellent Teaching

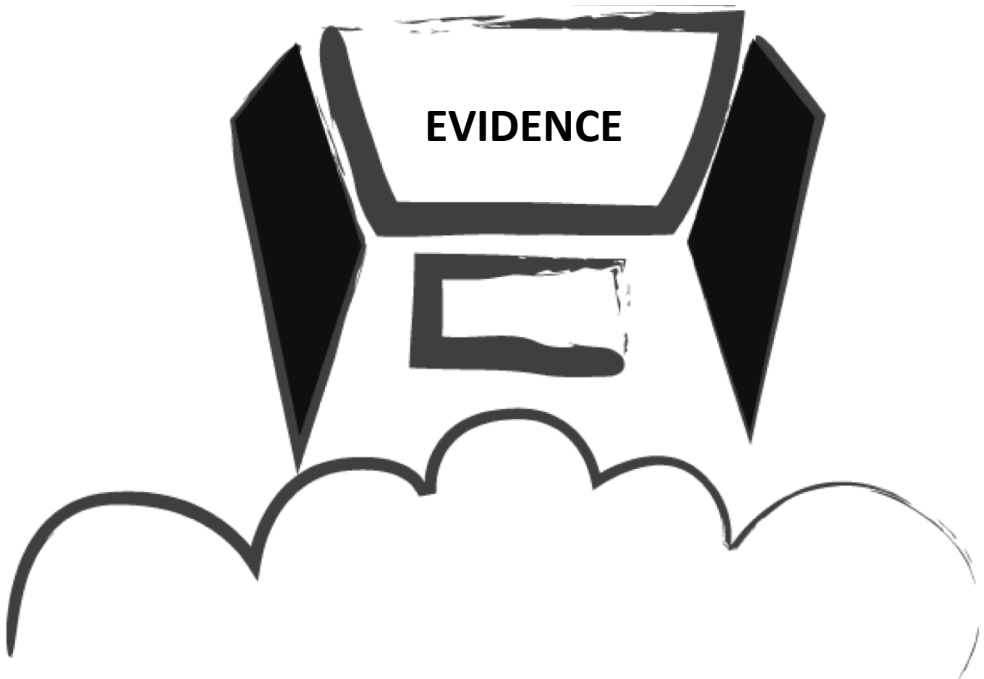
- Duik dieper in leertheorieën en cognitieve psychologie
- En leer deze toepassen in de praktijk
- Geheel wetenschappelijk onderbouwd

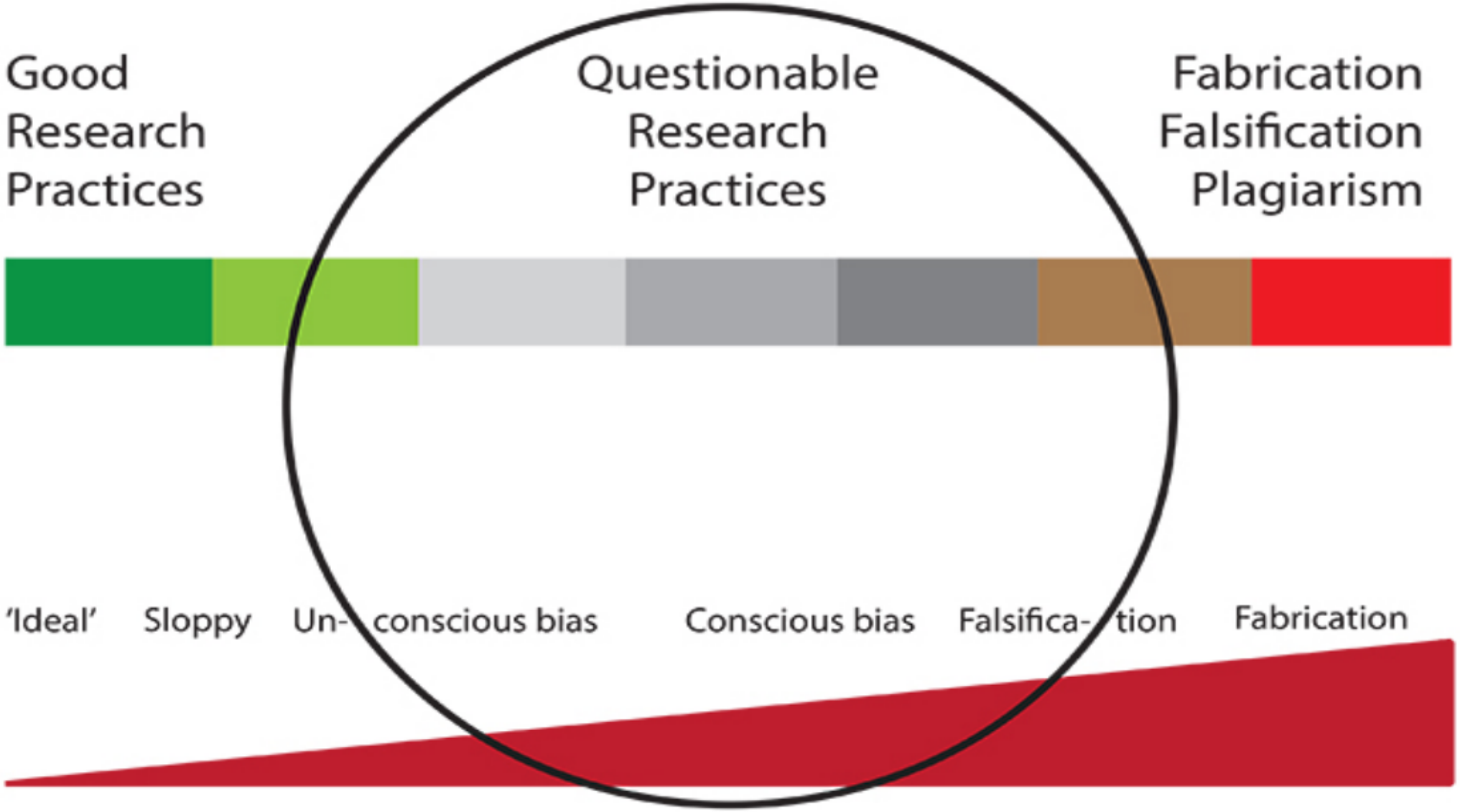
Persoonlijk advies? **Inschrijven**



“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.”

Questionable Research Practices




Good
Research
Practices
Questionable
Research
Practices
Fabrication
Falsification
Plagiarism
'Ideal'
Sloppy
Un-
conscious bias
Conscious bias
Falsifica-
tion
Fabrication

1. Stop collecting data once $p<.05$
2. Analyze many measures, but report only those with $p<.05$.
3. Collect and analyze many conditions, but only report those with $p<.05$.
4. Use covariates to get $p<.05$.
5. Exclude participants to get $p<.05$.
6. Transform the data to get $p<.05$.

"If you don't reveal some insights soon, I'm going to be forced to slice, dice, and drill!"

RESEARCH ARTICLE

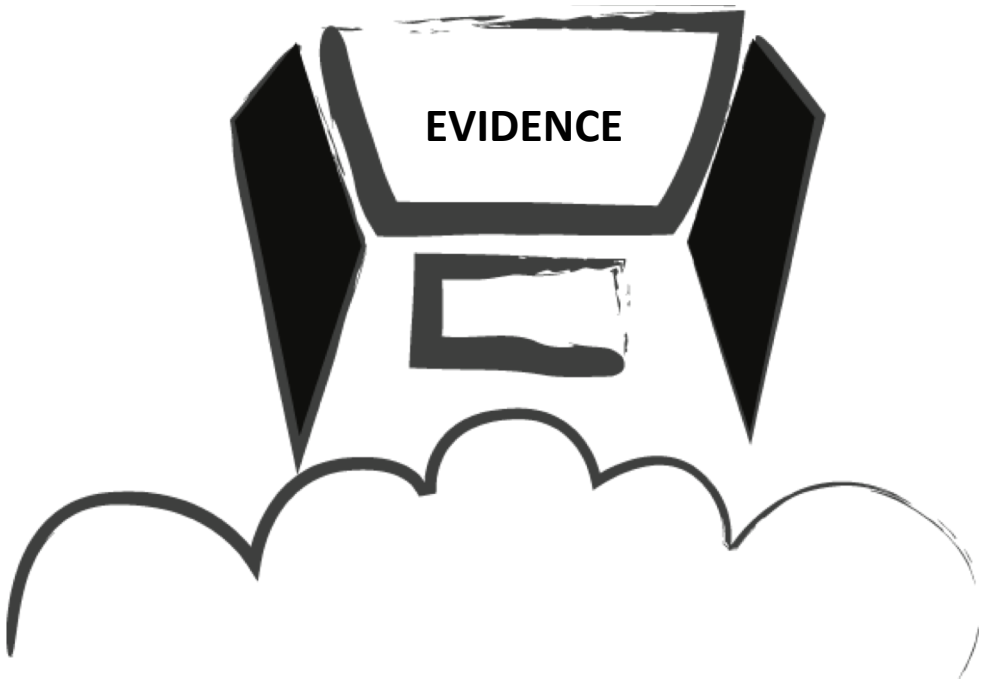
Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands

Gowri Gopalakrishna[1☯]*, Gerben ter Riet[2☯], Gerko Vink[3‡], Ineke Stoop[4‡], Jelte M. Wicherts[5☯], Lex M. Bouter[1,6☯]

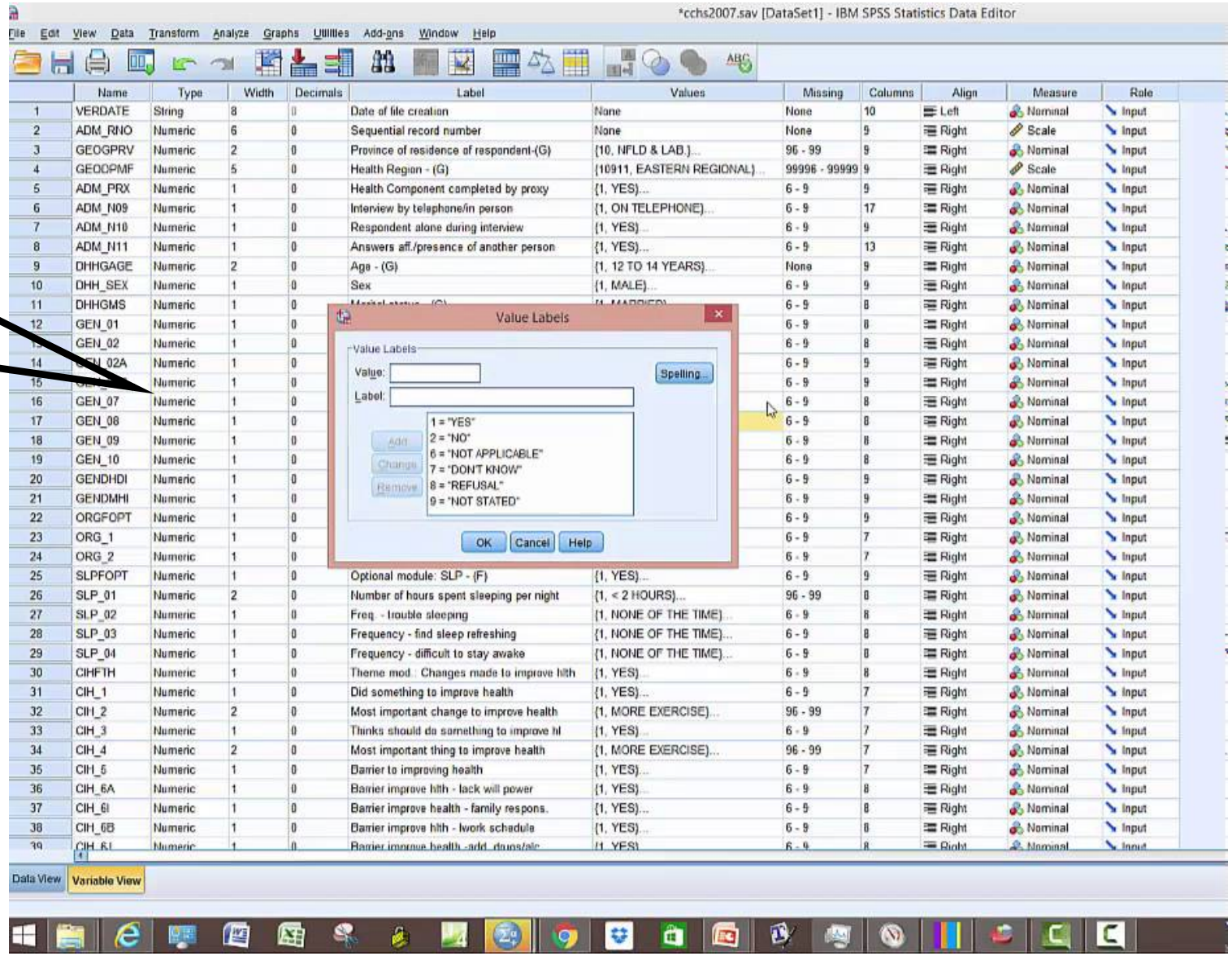
Table 2. Prevalence (95% confidence intervals) of the QRPs, any frequent QRP and fabrication or falsification stratified by disciplinary field and academic rank.

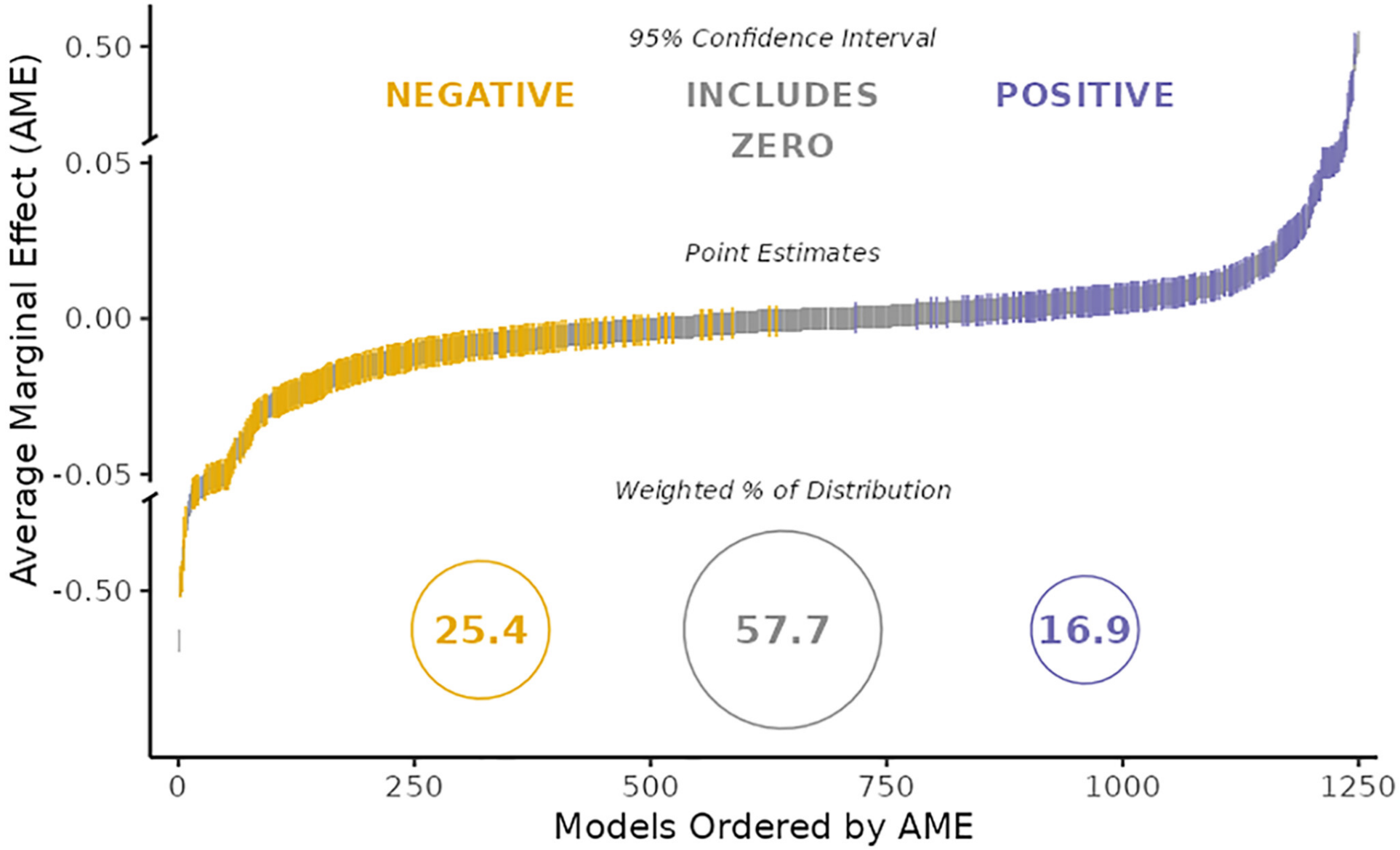
			Disciplinary field				Academic rank		
QRP	**Description (In the last three years.)**	**Life and medical sciences**	**Social and behavioural sciences**	**Natural and engineering sciences**	**Arts and humanities**	**PhD candidates and junior researchers**	**Postdocs and assistant professors**	**Associate and full professors**	**Overall**
Any frequent QRP	Score 5, 6 or 7 on at least 1 of the 11 QRPs	55.3 (53.4, 57.1)	50.3 (48.0, 52.5)	49.4 (46.8, 52.0)	42.1 (38.3, 46.1)	52.5 (50.3, 54.7)	52.3 (50.4, 54.2)	48.9 (46.7, 51.0)	51.3 (50.1, 52.5)
Fabrication	Making up of data or results	5.5 (3.2, 7.7)	4.[illegible] (2.[illegible], 7.5)	2.5 (0, 5.5)	0.7 (0, 5.1)	4.0 (1.4, 6.6)	4.9 (2.6, 7.1)	3.6 (1.1, 6.1)	4.[illegible] (2.[illegible], 5.7)
Falsification	Manipulating research materials, data or results	4.9 (2.7, 7.2)	2.0 (0, 4.6)	5.3 (2.2, 8.4)	6.1 (1.4, 10.9)	5.5 (2.8, 8.1)	2.6 (0.4, 4.8)	5.3 (2.7, 7.9)	4.[illegible] (2.[illegible], 5.6)
Any FF	Fabrication and/or Falsification	10.4 (7.1, 13.7)	5.7 (1.8, 9.5)	7.6 (3.1, 12.1)	8.4 (1.6, 15.3)	8.9 (5.0, 12.6)	7.3 (4.1, 10.6)	8.9 (5.1, 12.7)	8.3 (6.2, 10.3)

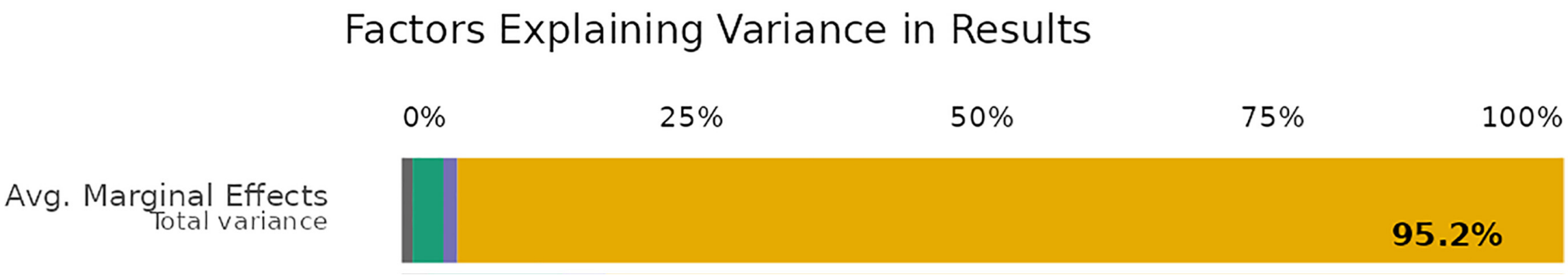
Multiple Analyst Problem



Breznau et al.
73 onderzoeksteams,
met 1 hypothese om
te toetsen




95% Confidence Interval
NEGATIVE
INCLUDES ZERO
POSITIVE
Point Estimates
Weighted % of Distribution
25.4
57.7
16.9
Average Marginal Effect (AME)
0.50
0.05
0.00
-0.05
-0.50
0
250
500
750
1000
1250
Models Ordered by AME



"…95,2% of the total variance is left unexplained, suggesting that massive variation in reported results originated from idiosyncratic decisions in the data analysis process."

Article | Published: 20 May 2020

Variability in the analysis of a single neuroimaging dataset by many teams

Rotem Botvinik-Nezer, Felix Holzmeister, Colin F. Camerer, Anna Dreber, Juergen Huber, Magnus Johannesson, Michael Kirchler, Roni Iwanir, Jeanette A. Mumford, R. Alison Adcock, Paolo Avesani, Blazej M. Baczkowski, Aahana Bajracharya, Leah Bakst, Sheryl Ball, Marco Barilari, Nadège Bault, Derek Beaton, Julia Beitner, Roland G. Benoit, Ruud M. W. J. Berkers, Jamil P. Bhanji, Bharat B. Biswal, Sebastian Bobadilla-Suarez, ... Tom Schonberg + Show authors

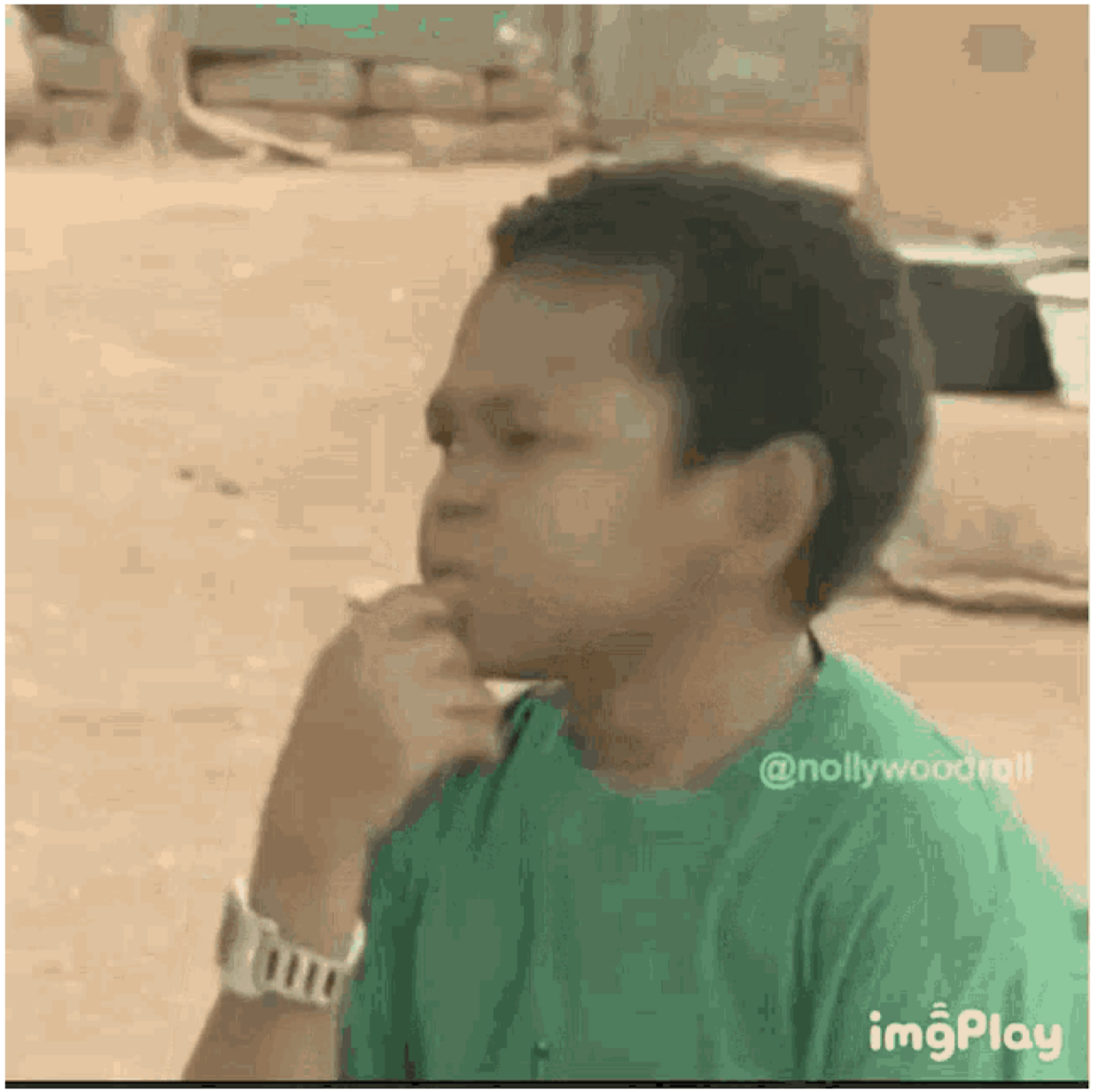
Nature **582**, 84–88 (2020) | Cite this article

Open access | Research article | First published online August 23, 2018

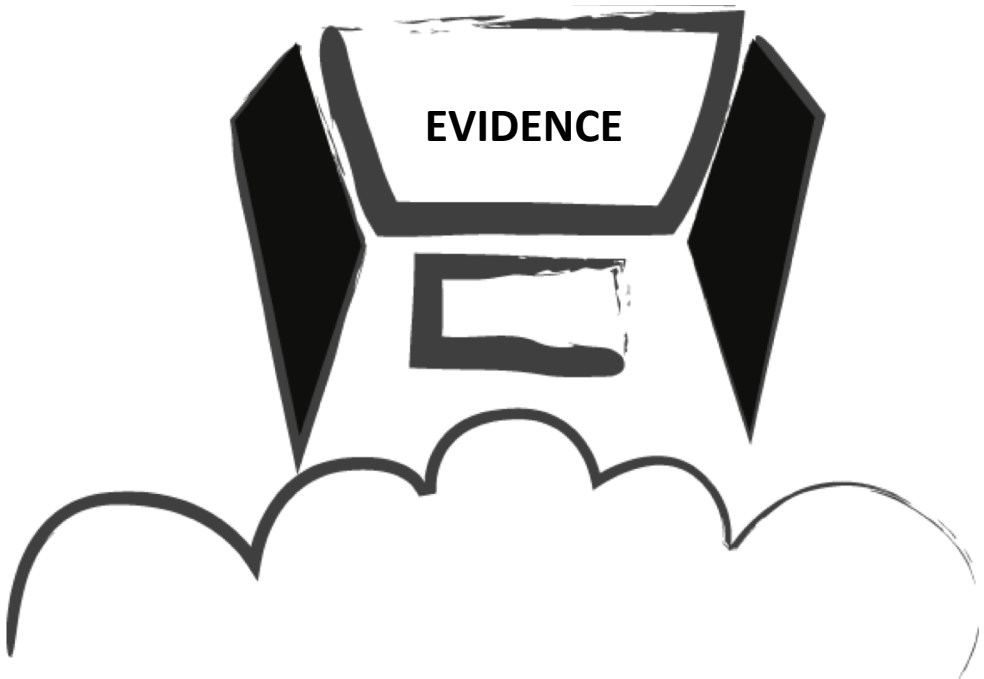
Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results

R. Silberzahn, E. L. Uhlmann, [...], and B. A. Nosek (+62) View all authors and affiliations

All Articles | https://doi.org/10.1177/2515245917747646 | View correction

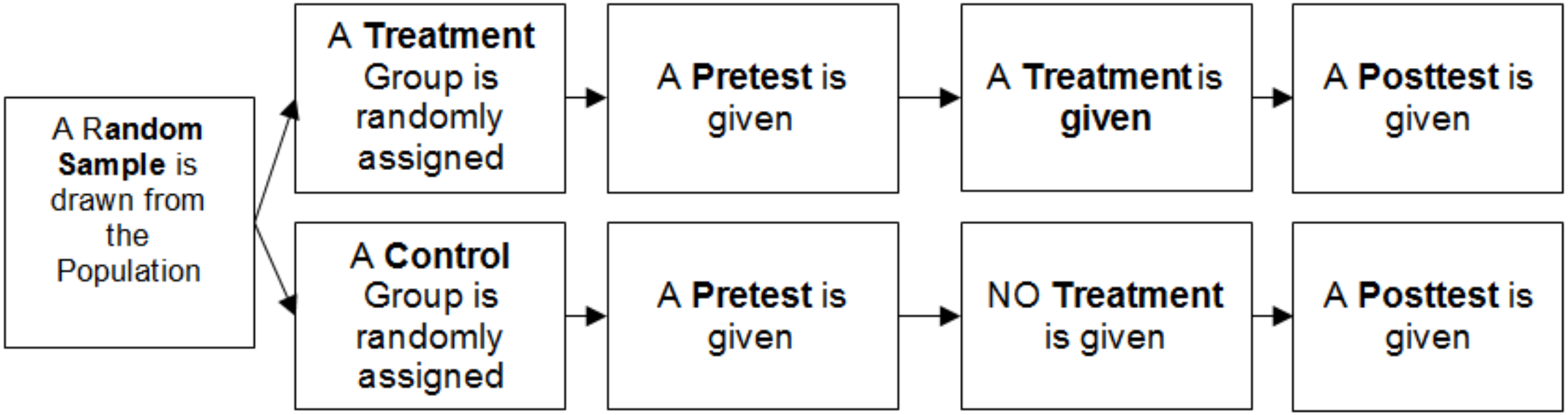


Experimenteren




NETFLIX

Breznau et al: “In experimental research, the data-generating model is often clearer or involves fewer decisions.”



RESEARCH ARTICLE | ECONOMIC SCIENCES |

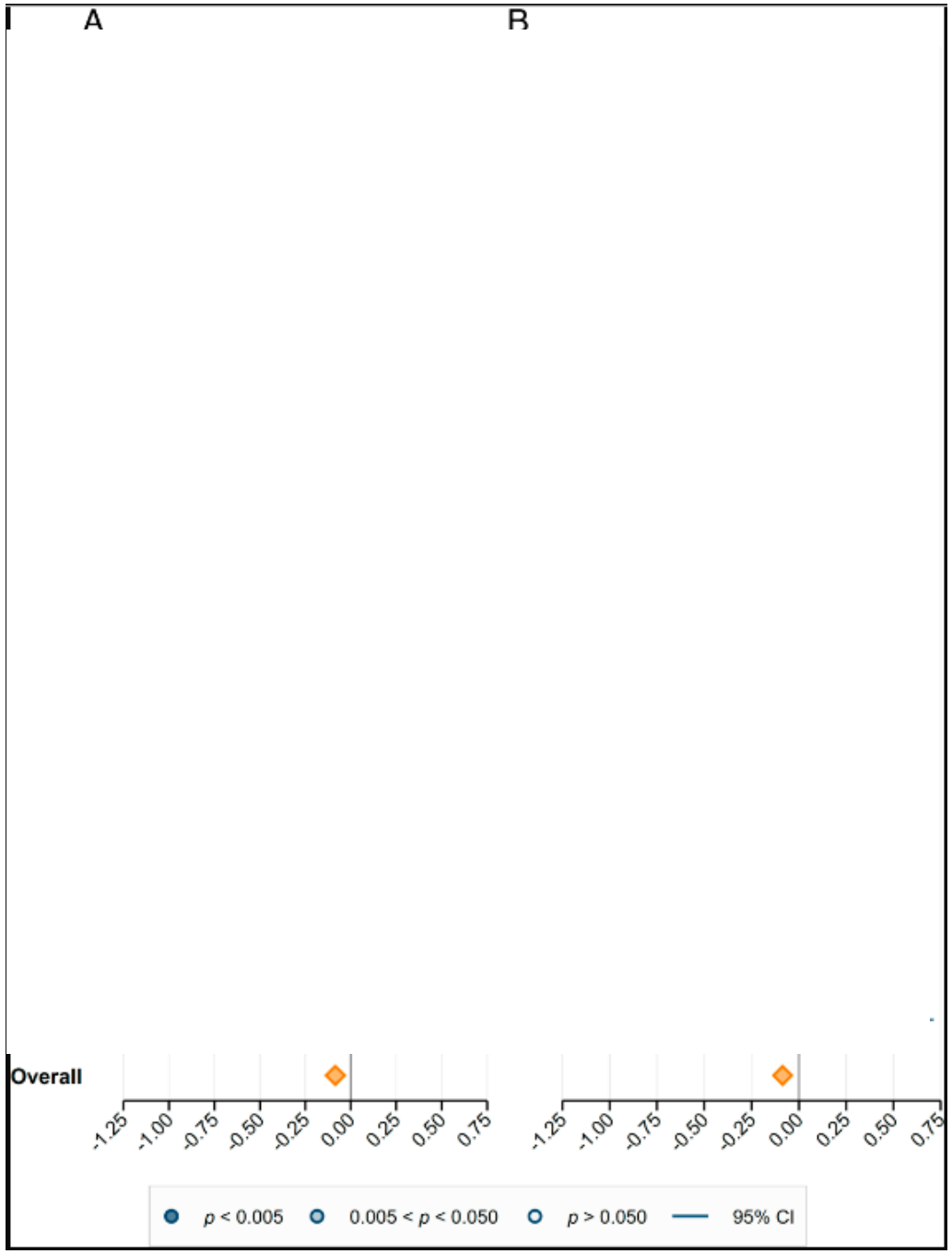
Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs

Christoph Huber, Anna Dreber, Jürgen Huber, +91, and Felix Holzmeister Authors Info & Affiliations

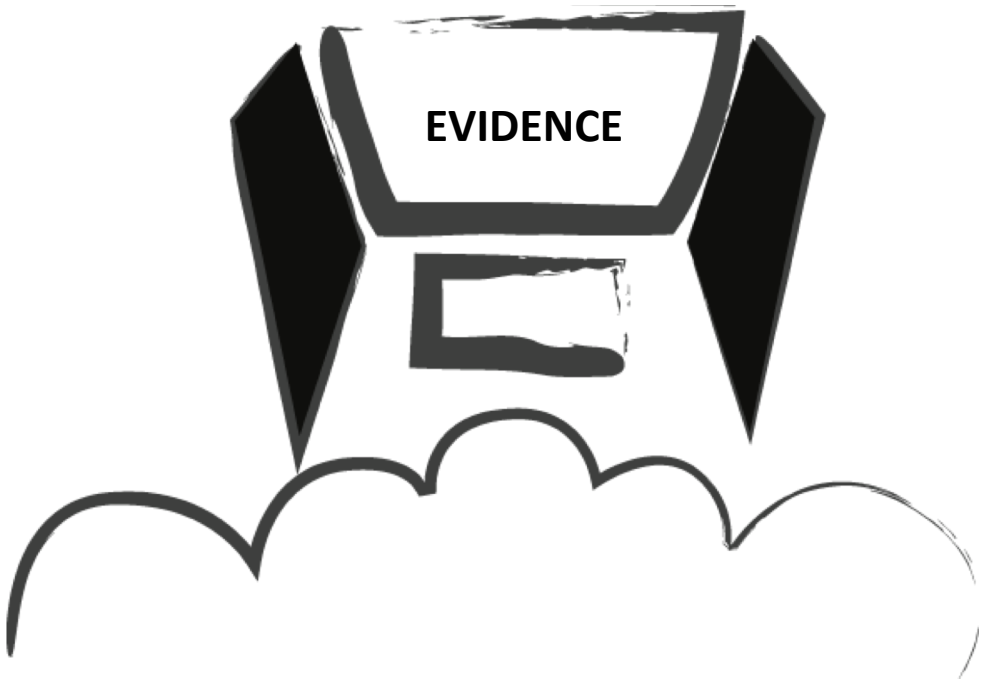
Edited by Timothy Wilson, University of Virginia, Charlottesville, VA; received September 13, 2022; accepted April 5, 2023

May 30, 2023 | 120 (23) e2215572120 | https://doi.org/10.1073/pnas.2215572120

“This highlights that the generalizability and informativeness of individual experimental designs are limited.”



Veldexperimenten



Internal Validity

External Validity


Julian

Zijn de resultaten van experimenten met Amerikaanse studenten in het psychologie lab aan de universiteit (die 20 dollar voor deelname ontvangen) geldig voor kinderen in groep 3 bij juf Janneke in Oosternijkerk Friesland?

EXAMPLE

Education
Endowment
Foundation

Cognitive Science in the Classroom:

Evidence and Practice Review

The evidence for the application of cognitive science principles in everyday classroom conditions (applied cognitive science) is limited, with uncertainties and gaps about the applicability of specific principles across subjects and age ranges.

Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?

Hugues Lortie-Forgues[1] and Matthew Inglis[2]

There are a growing number of large-scale educational randomized controlled trials (RCTs). Considering their expense, it is important to reflect on the effectiveness of this approach. We assessed the magnitude and precision of effects found in those large-scale RCTs commissioned by the UK-based Education Endowment Foundation and the U.S.-based National Center for Educational Evaluation and Regional Assistance, which evaluated interventions aimed at improving academic achievement in K–12 (141 RCTs; 1,222,024 students). The mean effect size was 0.06 standard deviations. These sat within relatively large confidence intervals (mean width = 0.30 *SD*s), which meant that the results were often uninformative (the median Bayes factor was 0.56). We argue that our field needs, as a priority, to understand why educational RCTs often find small and uninformative effects.

Keywords: educational policy; evaluation; meta-analysis; program evaluation

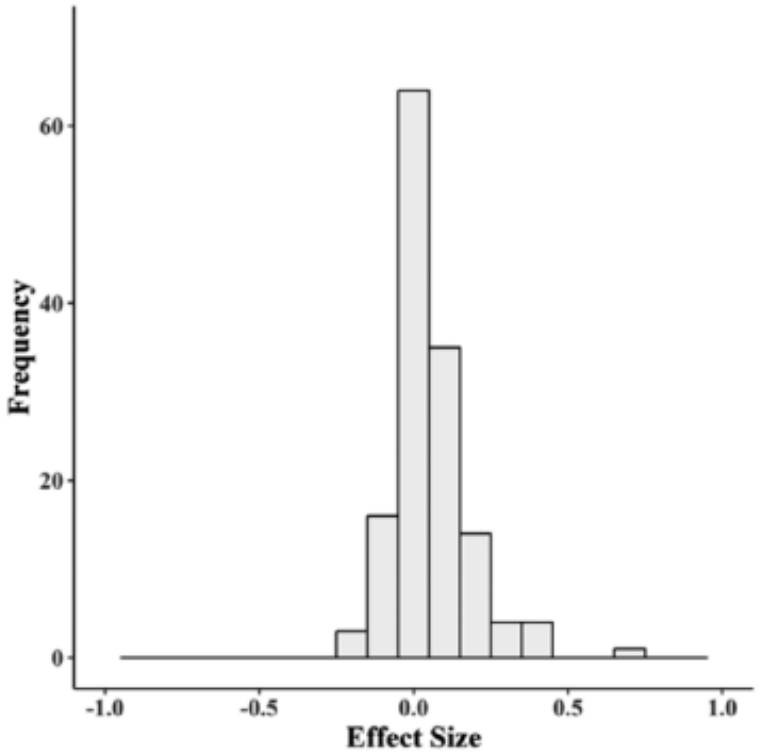


FIGURE 1. *The distribution of effect sizes from the 141 trials commissioned by the Education Endowment Foundation and National Center for Educational Evaluation and Regional Assistance.*

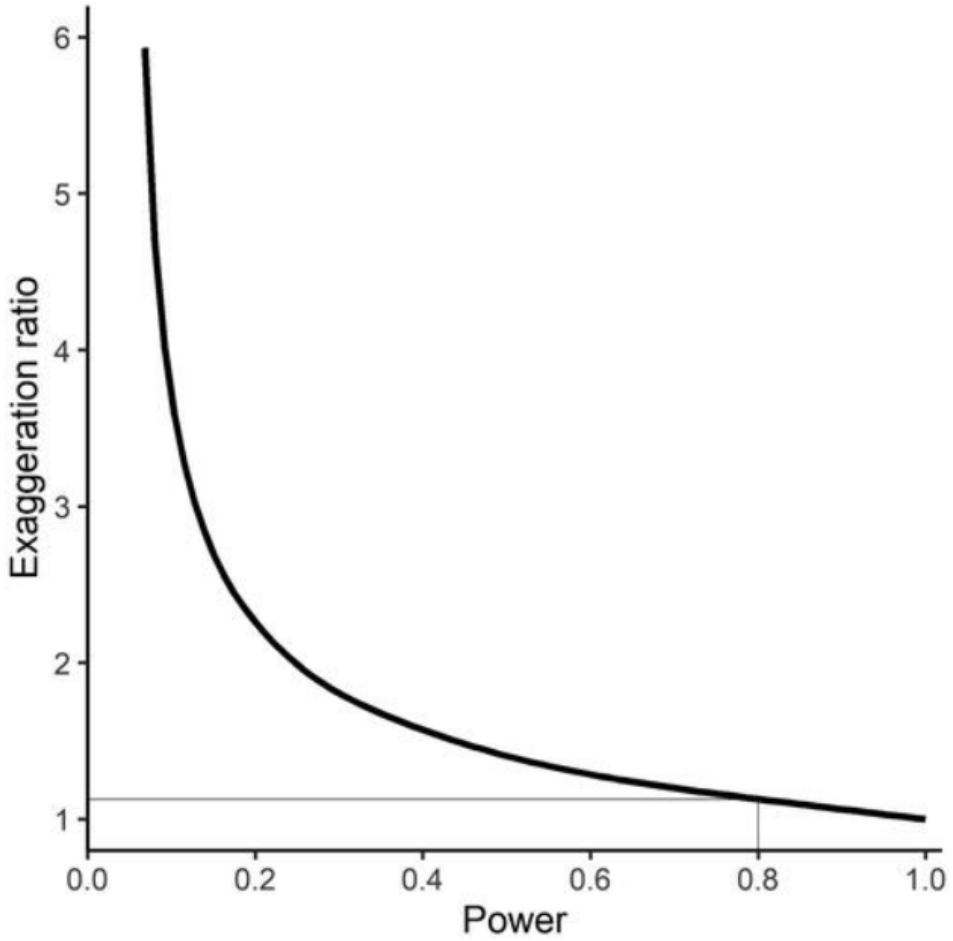
Quantifying "Promising Trials Bias" in Randomized Controlled Trials in Education

Sam Sims, Jake Anders, Matthew Inglis & Hugues Lortie-Forgues

Received 18 May 2020, Accepted 06 May 2022, Published online: 11 Jul 2022

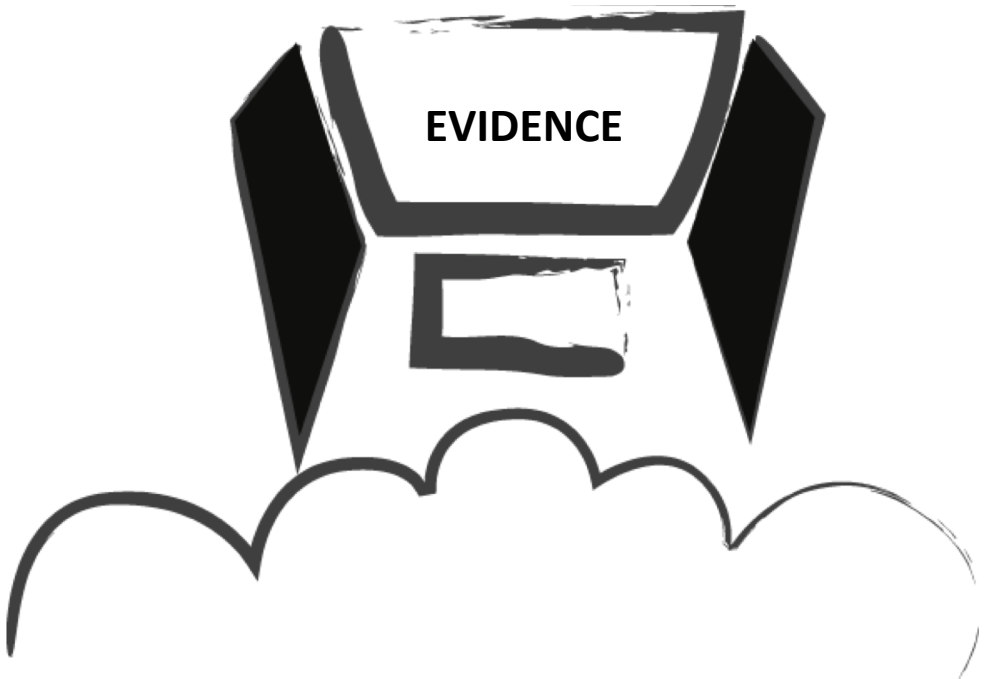
The estimated effect sizes are exaggerated by an average of 52% or more.

In particular, it would be prudent for them (educators) to expect smaller effects than would be suggested by single trial results listed in "warehouses" or "toolkits" of promising interventions. These are likely substantially upwardly biased.



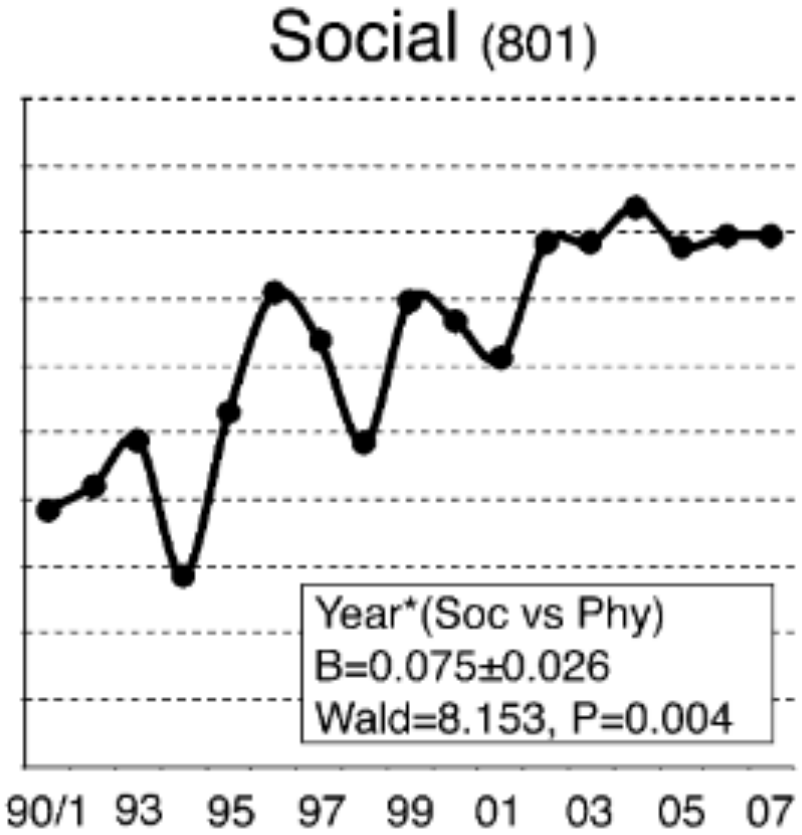

THISISELLIZ

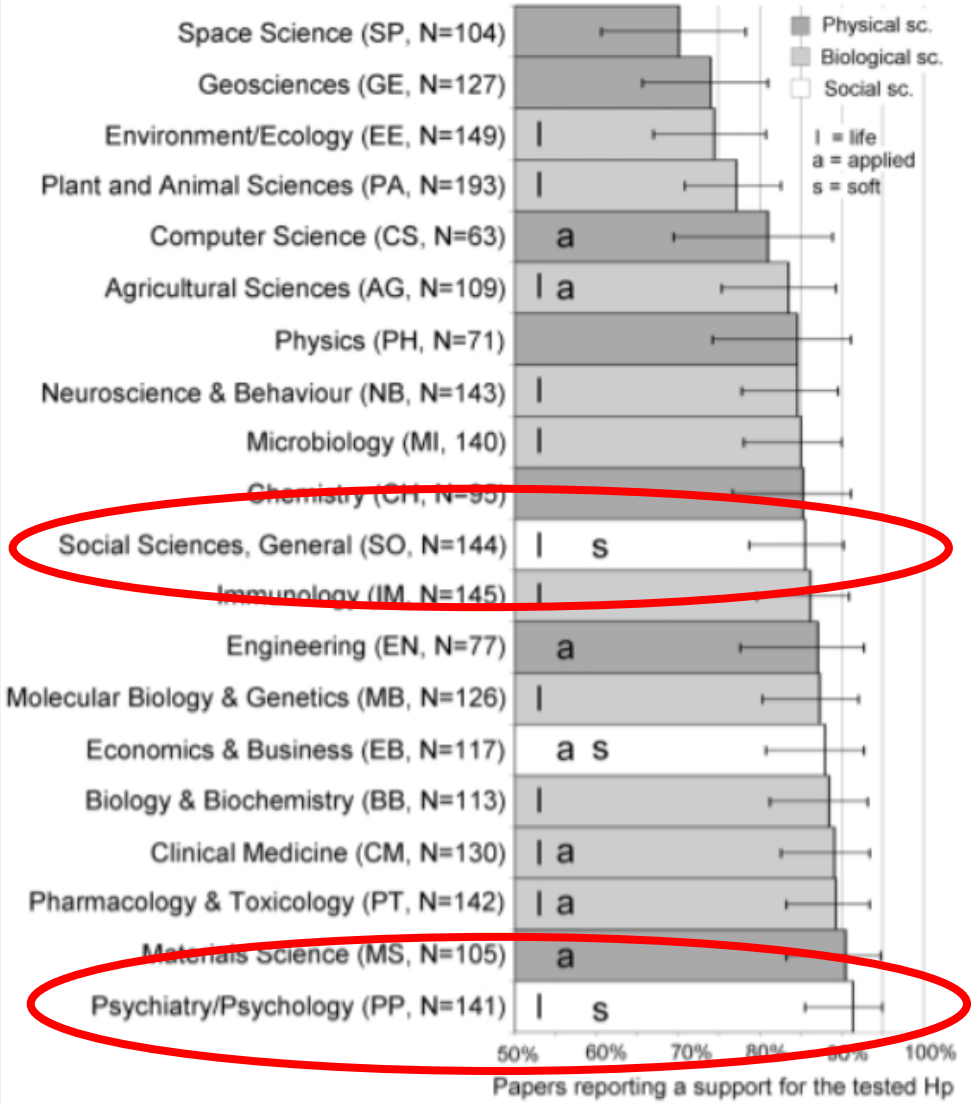
Werkt alles?



Negative results are disappearing from most disciplines and countries

Daniele Fanelli

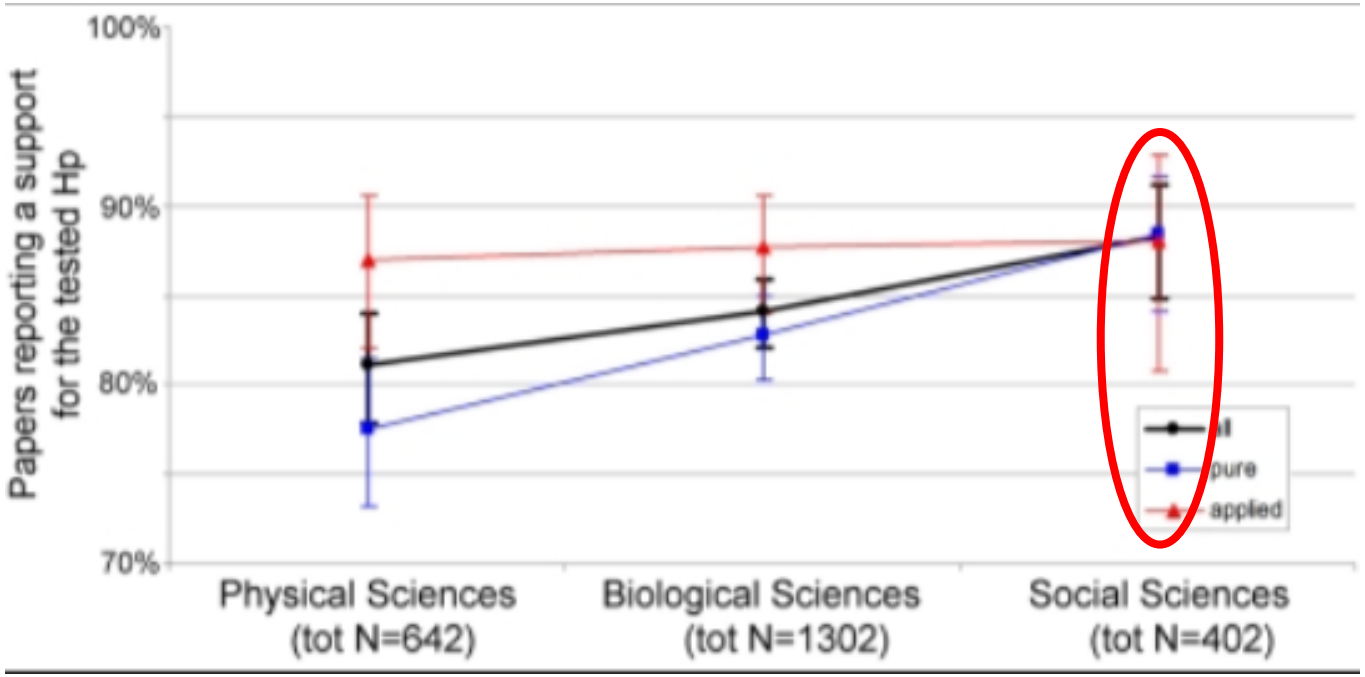




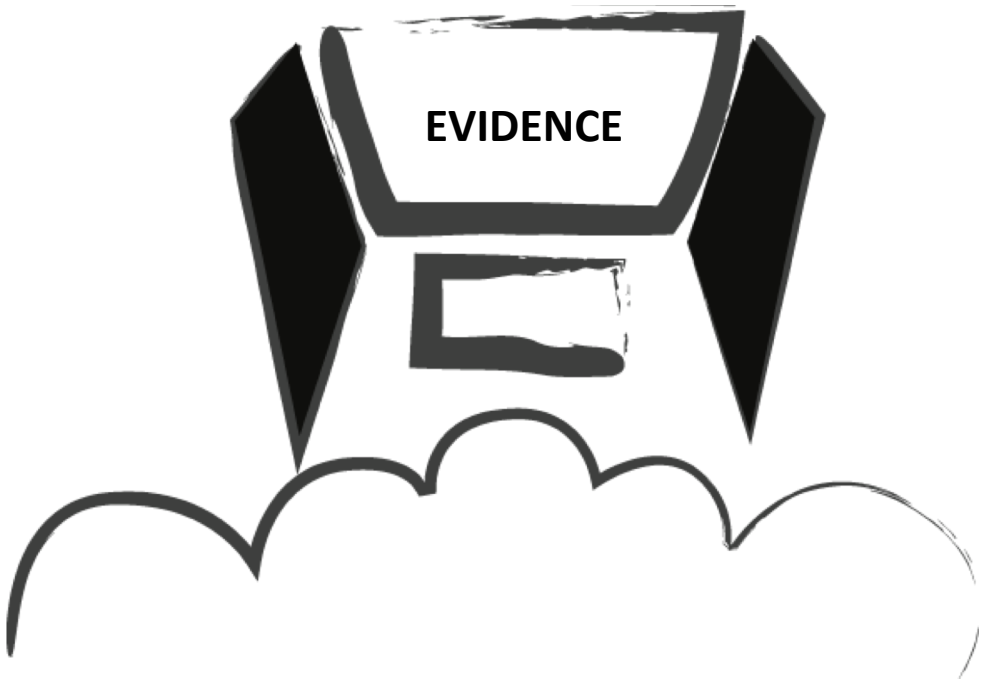
PLoS One. 2010 Apr 7;5(4):e10068. doi: 10.1371/journal.pone.0010068.

"Positive" results increase down the Hierarchy of the Sciences

Daniele Fanelli [1]



Replicatieonderzoek




We must replicate!!

Articles

A decade of replication study in education? A mapping review (2011–2020)

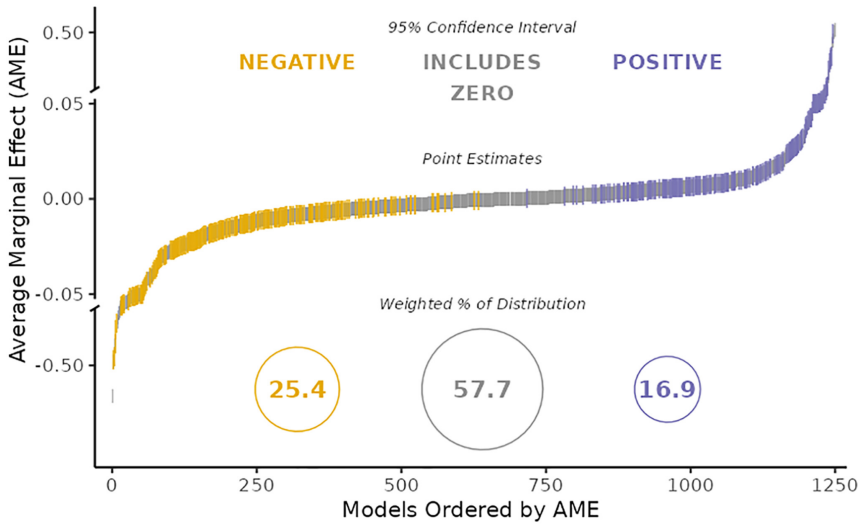
Thomas Perry, Rebecca Morris & Rosanna Lea

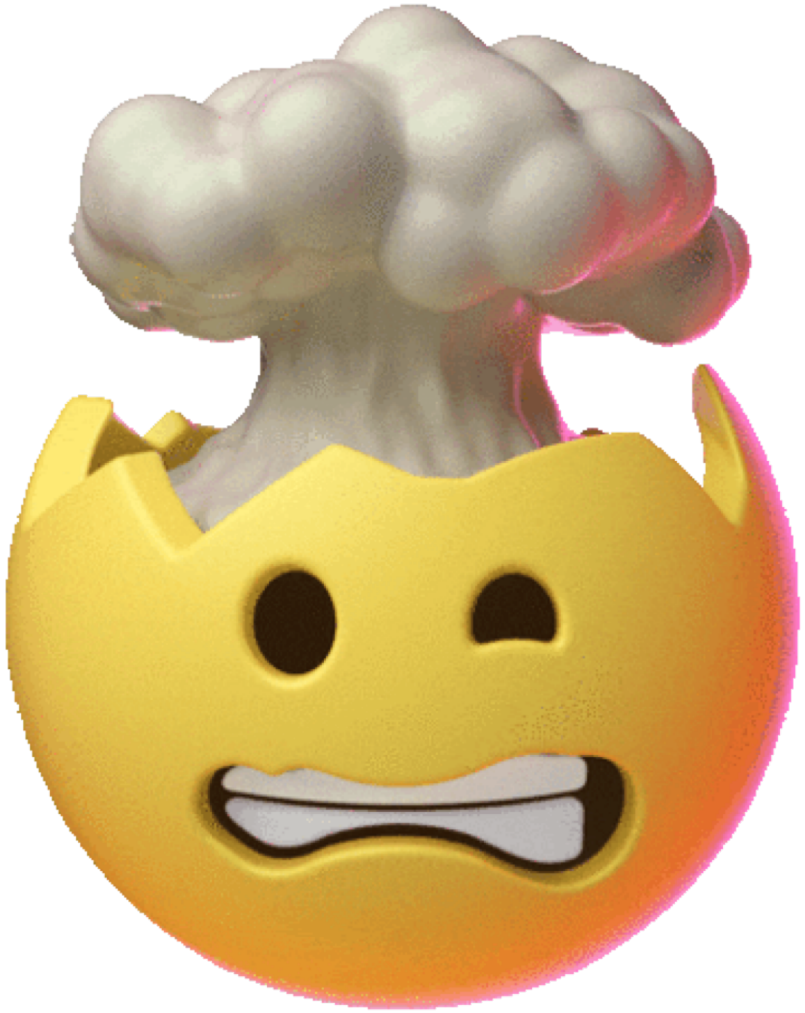
Pages 12-34 | Published online: 31 Jan 2022

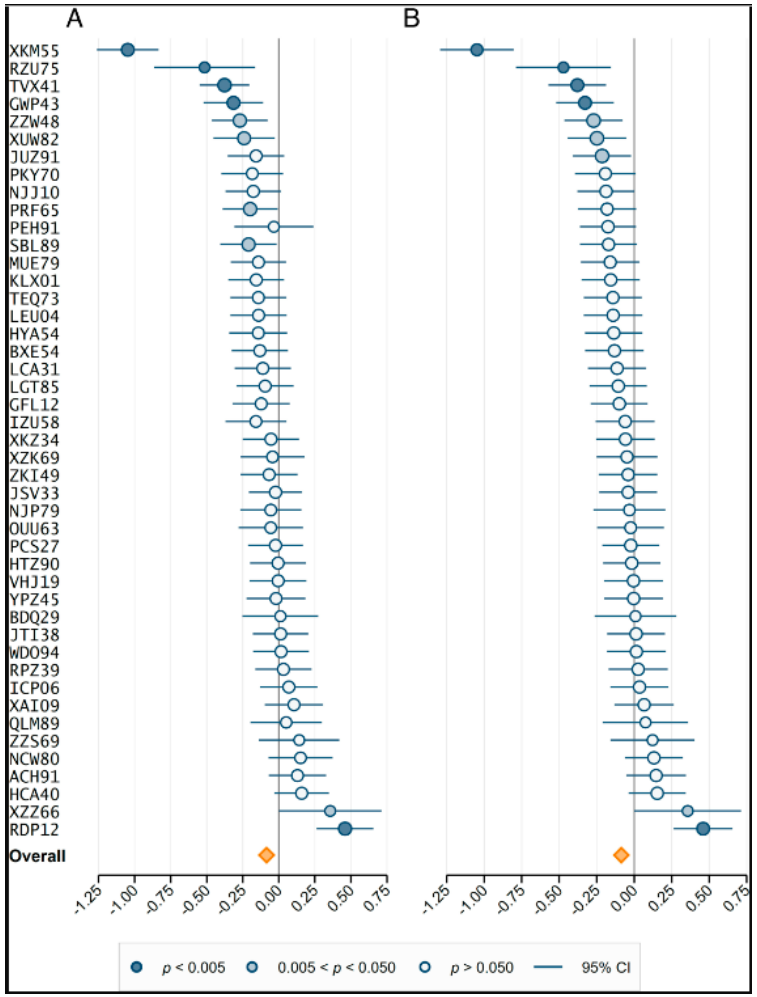
Cite this article https://doi.org/10.1080/13803611.2021.2022315 Check for updates

Like Makel and Plucker, we find very low rates of replication overall (at approximately 0.20% on average across the 2011–2020 period)….

1 op de 500 studies dus!!


95% Confidence Interval
NEGATIVE
INCLUDES ZERO
POSITIVE
Point Estimates
Weighted % of Distribution
25.4
57.7
16.9
Average Marginal Effect (AME)
0.50
0.05
0.00
-0.05
-0.50
0
250
500
750
1000
1250
Models Ordered by AME

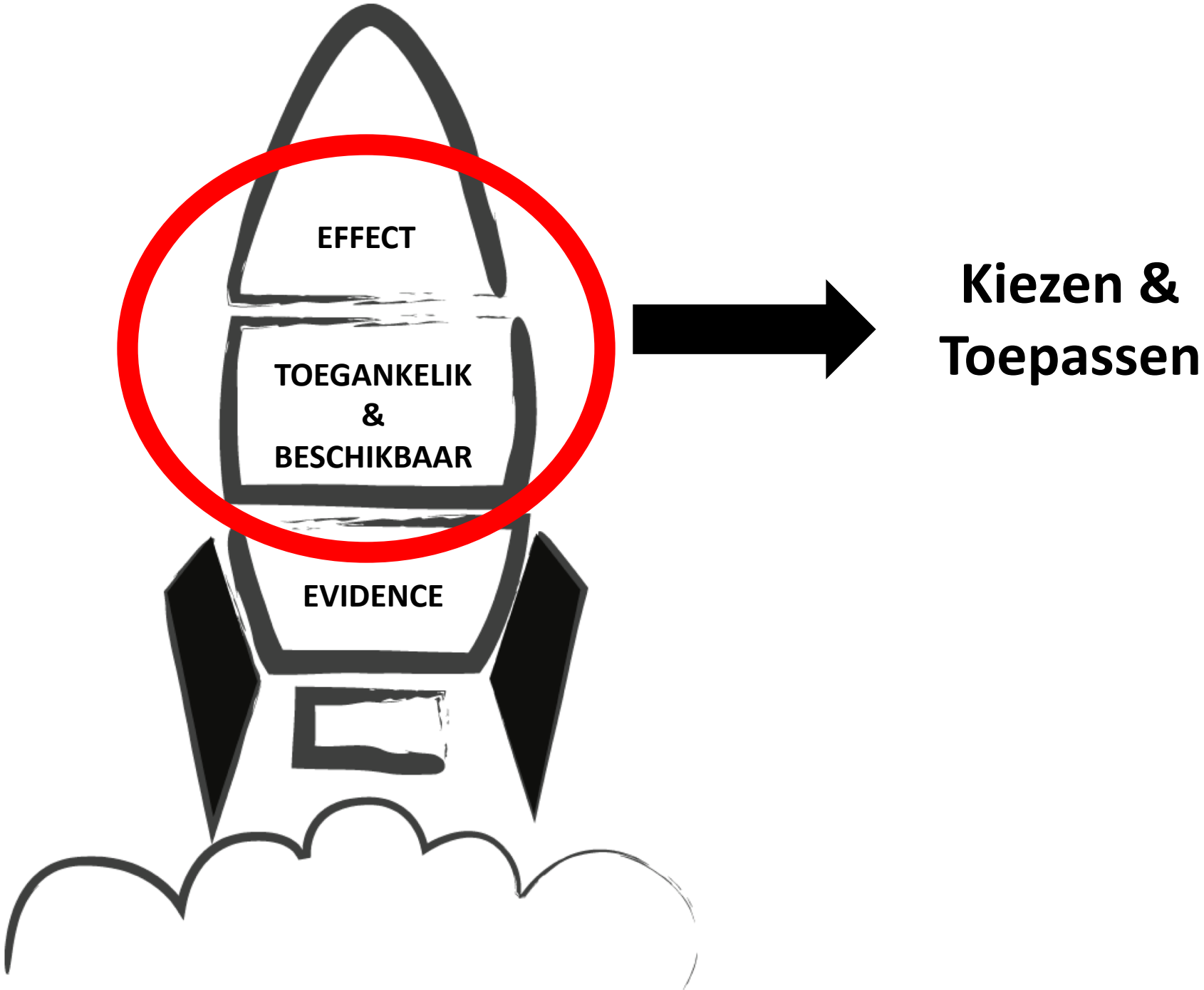



A
B
XKM55
RZU75
TVX41
GWP43
ZZW48
XUW82
JUZ91
PKY70
NJJ10
PRF65
PEH91
SBL89
MUE79
KLX01
TEQ73
LEU04
HYA54
BXE54
LCA31
LGT85
GFL12
IZU58
XKZ34
XZK69
ZKI49
JSV33
NJP79
OUU63
PCS27
HTZ90
VHJ19
YPZ45
BDQ29
JTI38
WDO94
RPZ39
ICP06
XAI09
QLM89
ZZS69
NCW80
ACH91
HCA40
XZZ66
RDP12
Overall
-1.25
-1.00
-0.75
-0.50
-0.25
0.00
0.25
0.50
0.75
p < 0.005
0.005 < p < 0.050
p > 0.050
95% CI

Comment

by David C. Berliner

Breznau et al (2022): “…scientists, especially those working with the complexities of human societies and behavior, should exercise humility and strive to better account for the uncertainty in their work.”


EFFECT
TOEGANKELIK
&
BESCHIKBAAR
EVIDENCE
Kiezen &
Toepassen

“

Bij realistisch rekenen ligt de nadruk op **begrip.**
Kinderen moeten **begrijpen wat**
ze doen als ze een som uitrekenen.

EFFECTIEFREKENONDERWIJS.NL

NO EVIDENCE


EVIDENCE

Research Article

Let me make mathematics and music together: A meta-analysis of the causal role of music interventions on mathematics achievement

Ayça Akın

Received 01 Nov 2022, Accepted 25 Feb 2023, Published online: 28 Jun 2023

Marcel Schmeier

Effectief
rekenonderwijs
op de basisschool

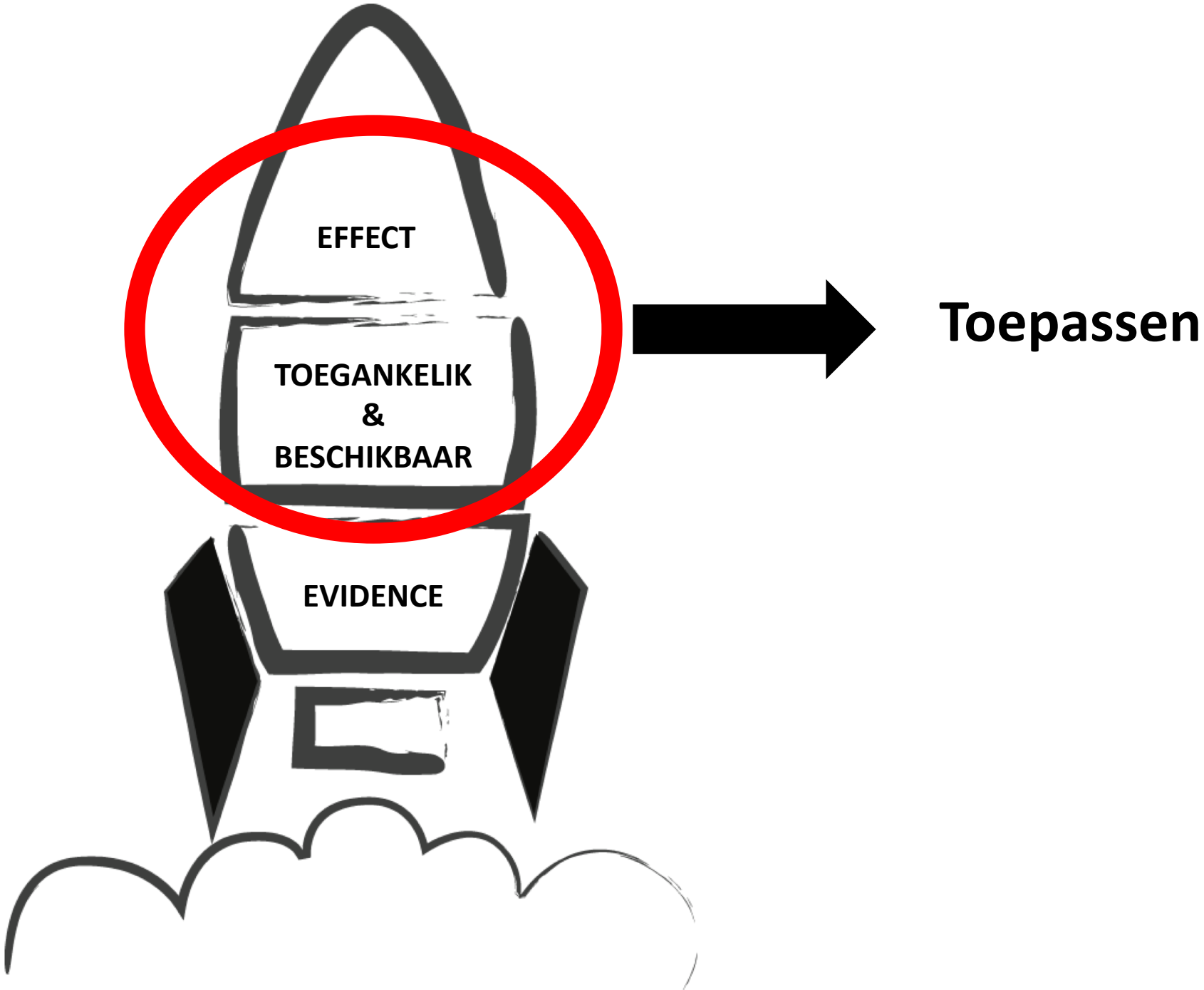
pica

CHOOSE

Table 3. Comparison of the overall effects estimated by three different methods.

	Multilevel meta-analysis	Robust variance estimation	Averaging effect sizes
Cognitive abilities	0.17	0.21*	0.20
Social skills	0.22*	0.26*	0.24**
Creativity	0.25	0.24	0.24
Motor skills	0.27	0.19	0.19
Academic achievement	1.10*	1.15*	1.15*

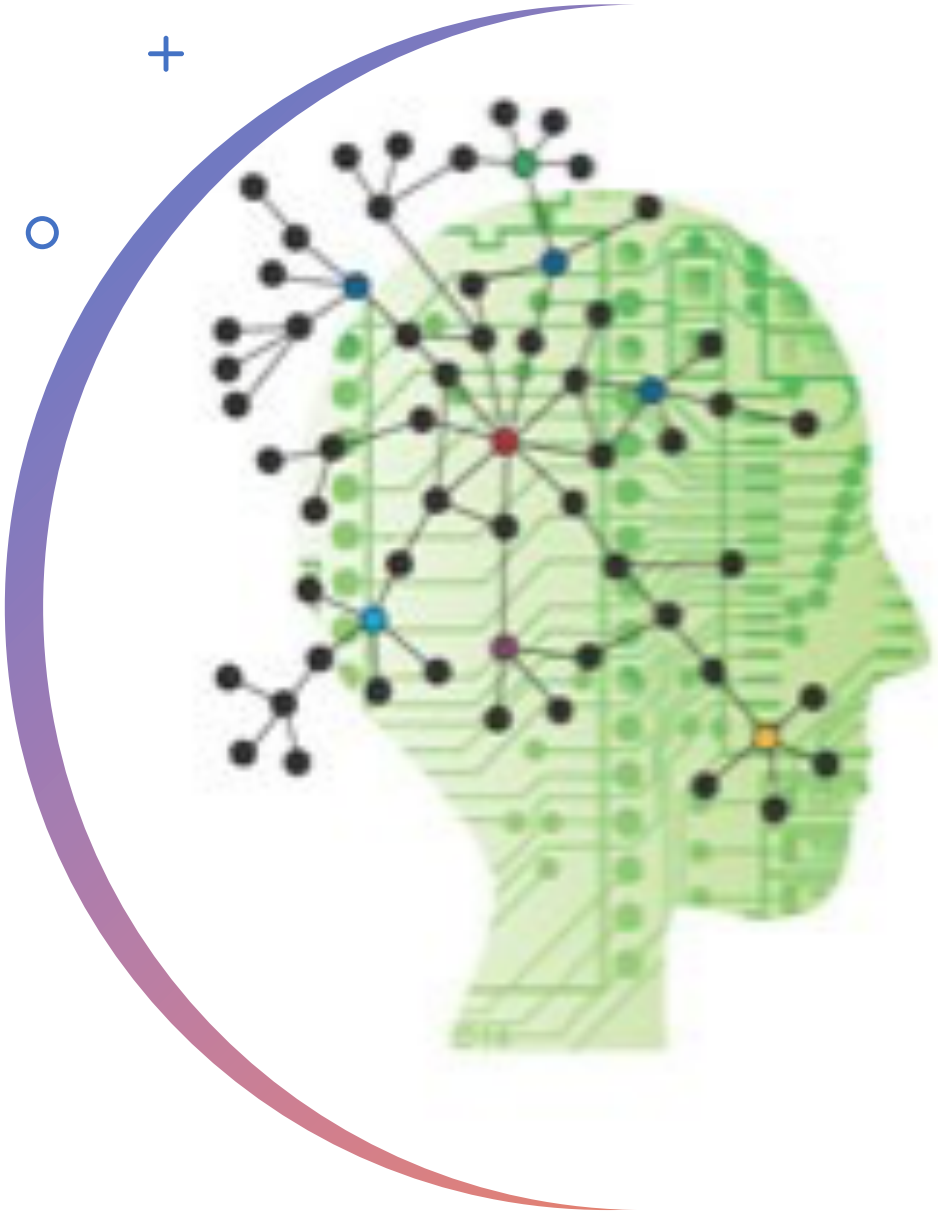
CHOOSE


EFFECT
TOEGANKELIK
&
BESCHIKBAAR
EVIDENCE
Toepassen

“Wij hebben geen gemiddelde leerlingen!”

“Gemiddelde leerlingen bestaan niet!”


kulfyapp.com
Hey, I'm average.



Wat werkt voor wie, waardoor, waarvoor en onder welke omstandigheden

Complexiteit van de toepassing

- Je werkt aan meerdere doelen op hetzelfde moment
- Doelen kunnen zich bevinden in verschillende doeldimensies
- Deze doelen kunnen conflicteren (wat effectief is voor het één kan minder effectief zijn voor het ander)
- Je werkt aan doelen in het hier en nu, maar met het oog op later
- Deze doelen kunnen conflicteren (wat voor nu effectief, kan voor het lange termijn doel minder effectief zijn)
- Wat nu bij deze groep voor dit doel werkt, hoeft niet te werken bij een andere groep op een ander moment
- Wat voor mij als leerkracht werkt, hoeft niet voor jou te werken

Comment

Teaching: The Absolute Best Job of All

by Symen van der Zee

1. Hoe stevig is de kennis uit effectonderzoek?
2. Hoe verhoudt effectonderzoek zich tot een onderwijsvisie?
3. Hoe zeker is het dat de toepassing leidt tot effectiever onderwijs?


QUESTION
AUTHORITY


QUESTION
ALL
AUTHORITY
WHY SHOULD I?