Knowledge mobilization and lessons for communicating meta-analytic results

Katie Fitzgerald
Azusa Pacific University

Elizabeth Tipton
Northwestern University

David Khella, Avery Charles
Azusa Pacific University

AERA SRMA SIG
8 December 2023

2022

The Meta-Analytic Rain Cloud Plot: A New Approach to Visualizing Clearinghouse Data

Kaitlyn G. Fitzgerald & Elizabeth Tipton

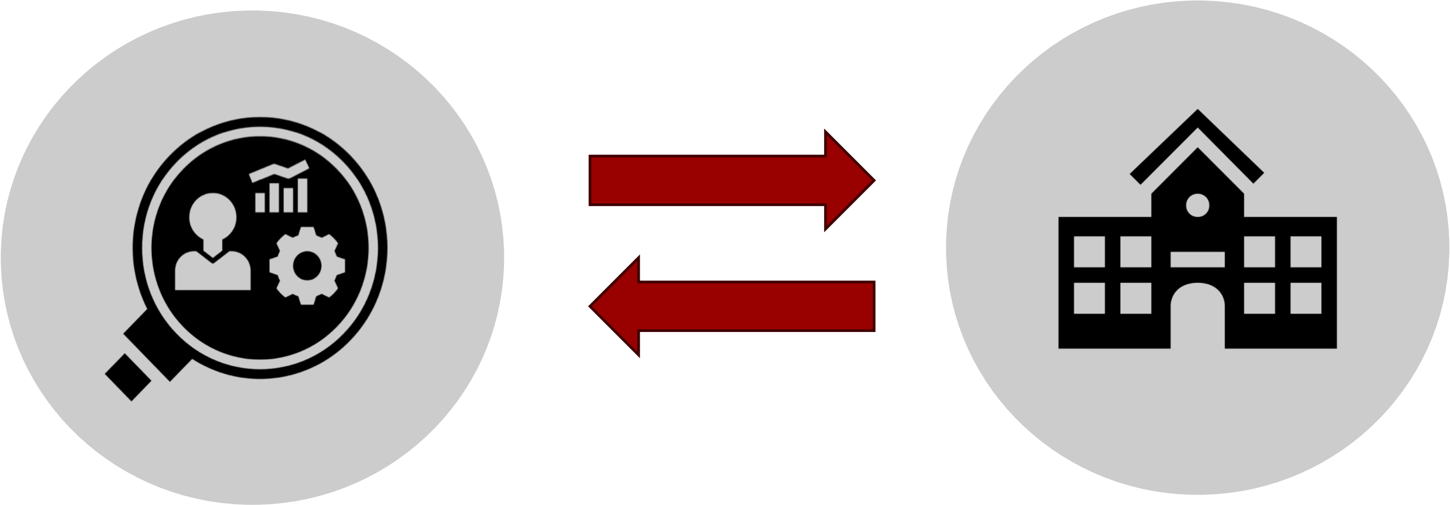
+ New MARC plot updates!

2023

A Knowledge Mobilization Framework: Toward Evidence-Based Statistical Communication Practices in Education Research

Kaitlyn G. Fitzgerald & Elizabeth Tipton

Evidence Use & Knowledge Mobilization



Research Evidence

Educational practice

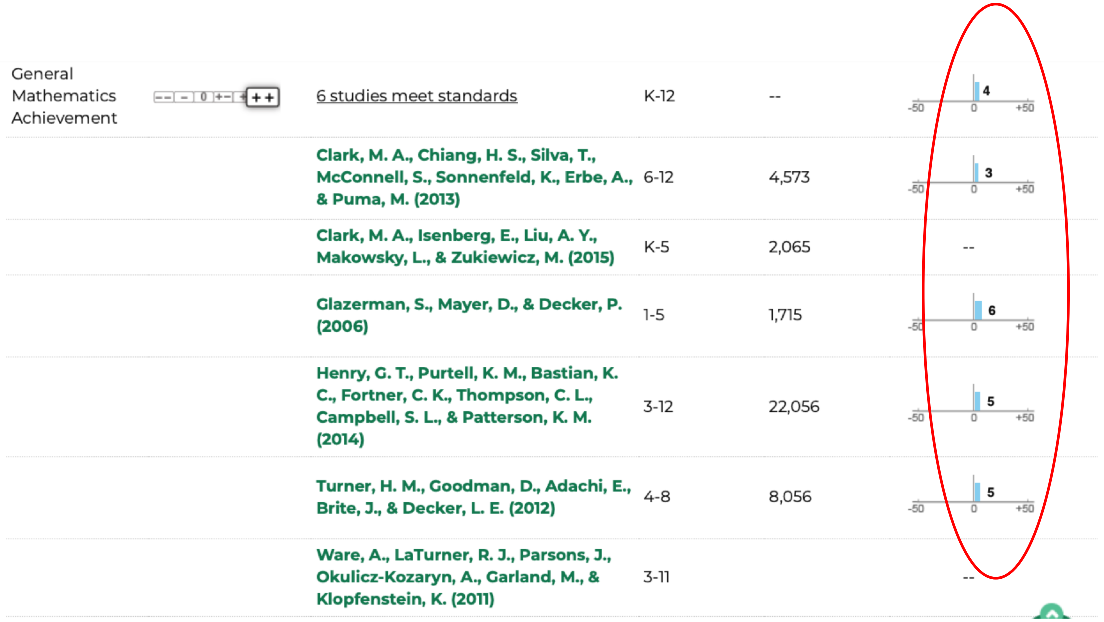
Shouldn’t assume evidence is used and usefu

Our corner of the problem

We're trying to communicate effect sizes, statistical uncertainty, meta-analytic results

People have poor statistical reasoning skills; minimal evidence on meta-analytic reasoning

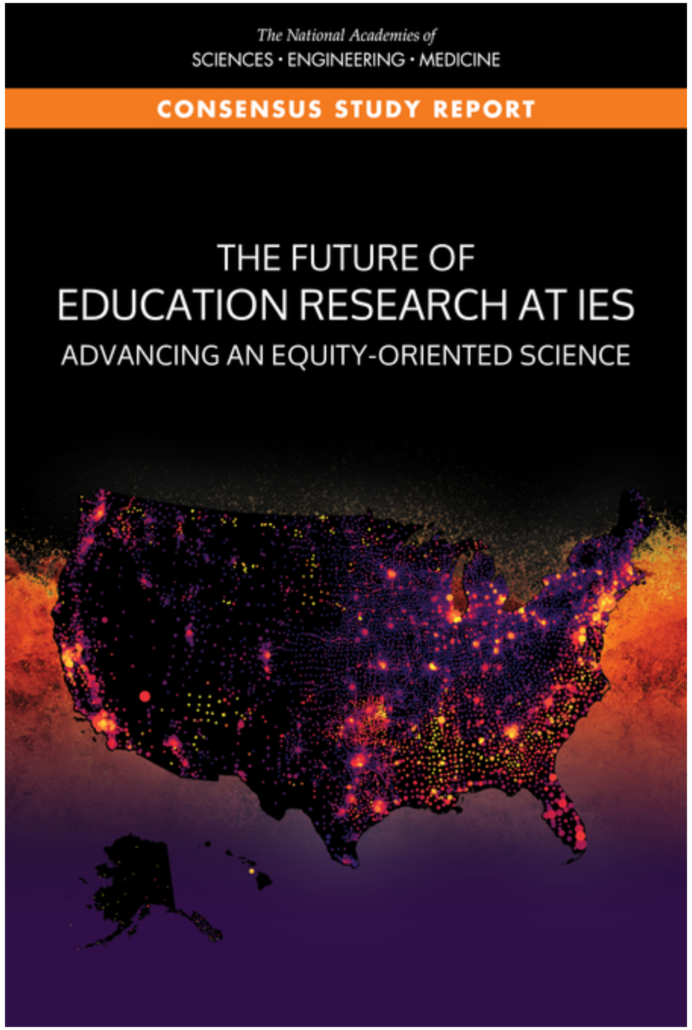
Common visualizations don't align with data viz best practices



General Mathematics Achievement	-- - 0 +- + ++	6 studies meet standards	K-12	--	4
		Clark, M. A., Chiang, H. S., Silva, T., McConnell, S., Sonnenfeld, K., Erbe, A., & Puma, M. (2013)	6-12	4,573	3
		Clark, M. A., Isenberg, E., Liu, A. Y., Makowsky, L., & Zukiewicz, M. (2015)	K-5	2,065	--
		Glazerman, S., Mayer, D., & Decker, P. (2006)	1-5	1,715	6
		Henry, G. T., Purtell, K. M., Bastian, K. C., Fortner, C. K., Thompson, C. L., Campbell, S. L., & Patterson, K. M. (2014)	3-12	22,056	5
		Turner, H. M., Goodman, D., Adachi, E., Brite, J., & Decker, L. E. (2012)	4-8	8,056	5
		Ware, A., LaTurner, R. J., Parsons, J., Okulicz-Kozaryn, A., Garland, M., & Klopfenstein, K. (2011)	3-11		--

How do people reason about the evidence we put in front of them? How can we improve?

National Academies Report (2022)



Knowledge Mobilization as one of five types of needed research

“how schools and decision-makers identify problems and develop solutions; which interventions, curricula, and programs are currently used in schools; **how to get promising evidence into their hands**; **how educational leaders harness that evidence to guide action; and what conditions support educational leaders to use research more centrally and substantively in their decision making**.”

([Bolding added]; National Academies of Sciences et al., 2022; Farley-Ripple et al., 2018; Jackson, 2022)

“Strategies to mobilize knowledge [should] be studied directly”

“IES should prioritize research on research use itself” (Conaway, 2021).

How might we structure these knowledge mobilization studies?

Goal: more effective, evidence-based statistical communication practices in education

Organize Knowledge Mobilization into three facets…

Prescriptive

How can we *help* facilitate better evaluations of the evidence?

Knowledge Mobilization

Normative

How *should* people evaluate statistical evidence?

Descriptive

How *do* education decision-makers evaluate statistical evidence?

We need to…

Examine *norms* embedded in evidence we communicate

Descriptively understand how decision-makers reason about this evidence as well as their varied decision-making needs

Prescriptively develop and evaluate communication strategies that facilitate better use of evidence by decision-makers

Bell et. al. (1988), Beyth-Marom et. al. (2008)

Case study: What Works Clearinghouse Evidence

Outcome domain	Effectiveness rating	Studies meeting standards	Grades examined	Students	Improvement index
Algebra	++	5 studies meet standards	8-PS	6,854	4
		Cabalo, J. V., Jaciw, A., & Vu, M.-T. (2007)	8-PS	344	--
		Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009)	8-9	270	--
		Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014)	8-12	5,738	3
		Ritter, S., Kulikowich, J., Lei, P., McGuire, C., & Morgan, P. (2007)	9	255	15
		Wolfson, M., Koedinger, K., Ritter, S., & McGuire, C. (2008)	9-12	247	19

Normative:

How should people reason about a collection of studies?

What's the appropriate way to make sense of the 6 lines of research presented here?

Prescriptive:

What are effective strategies and means of communication to bridge the gap?

What info should be included, how should it be displayed?

Descriptive:

How do decision-makers reason about and interpret this information?

Is this information relevant to their decision-making needs?

IMPORTANTLY: gaps between intended use and actual use of an information display are not always a result of decision-maker misunderstanding. Such gaps can also result when researchers misunderstand the information needed for decision-making.

Thesis: Knowledge Mobilization is an invitation to be more evidence-based in our own practices, and we think this framework can help

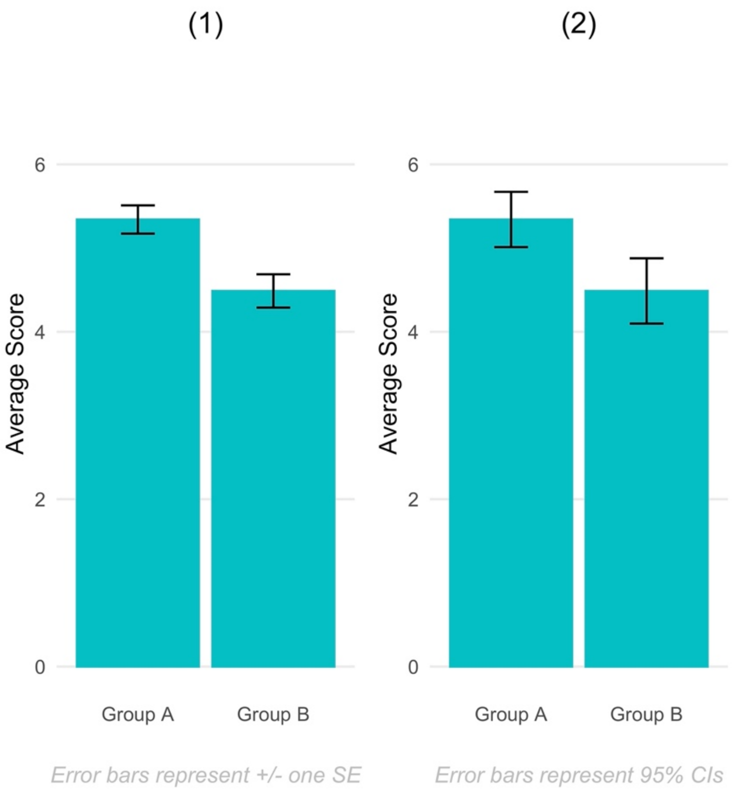
So what evidence should we turn to? And where do we need to generate new evidence?

Lessons from Data Viz, Cog Sci, HCI

Beware of the curse of expertise!

Message sent $\neq$ message received

Descriptive – lessons from Data Viz, Cog Sci, HCI



Caution against:

Error bars for uncertainty

Bar plots for effect sizes

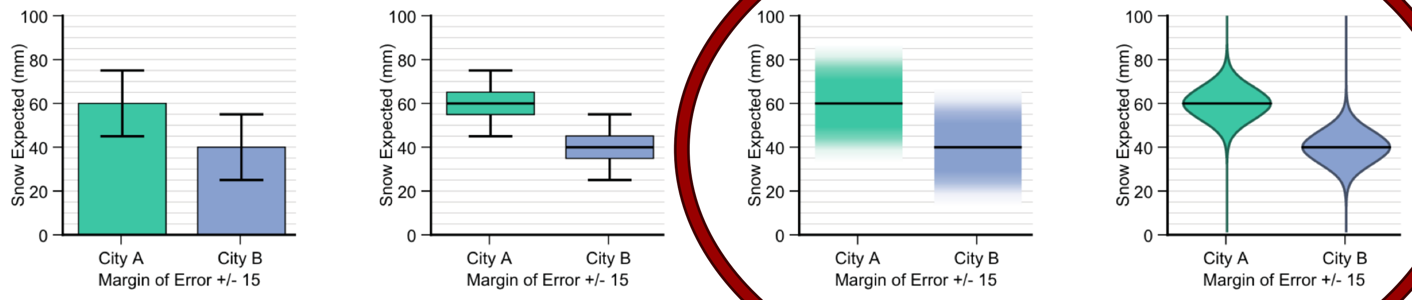
Belia et al. (2005), Correll and Gleicher (2014), Newman & Scholl (2012)

Prescriptive – lessons from Data Viz, Cog Sci, HCI

2142 IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 20, NO. 12, DECEMBER 2014

Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error

Michael Correll *Student Member, IEEE*, and Michael Gleicher *Member, IEEE*



(a) **Bar chart** with error bars: the height of the bars encodes the sample mean, and the whiskers encode a 95% t-confidence interval.

(b) **Modified box plot**: The whiskers are the 95% t-confidence interval, the box is a 50% t-confidence interval.

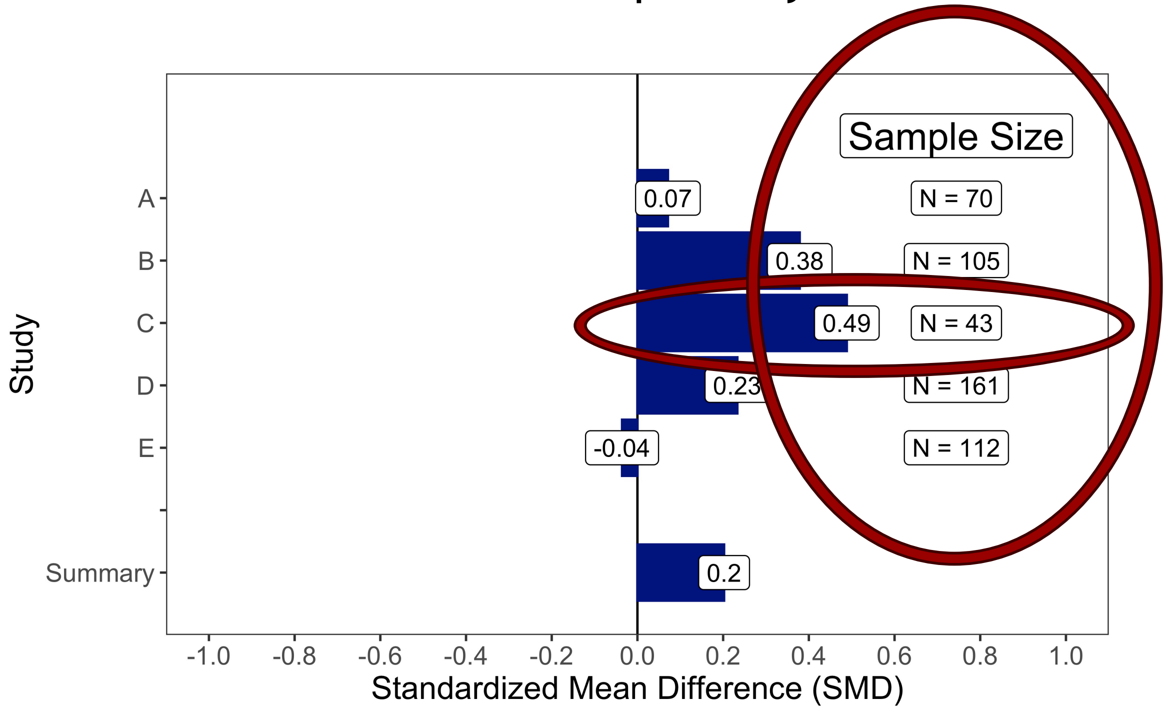
(c) **Gradient plot**: the transparency of the colored region corresponds to the cumulative density function of a t-distribution.

(d) **Violin plot**: the width of the colored region corresponds to the probability density function of a t-distribution.

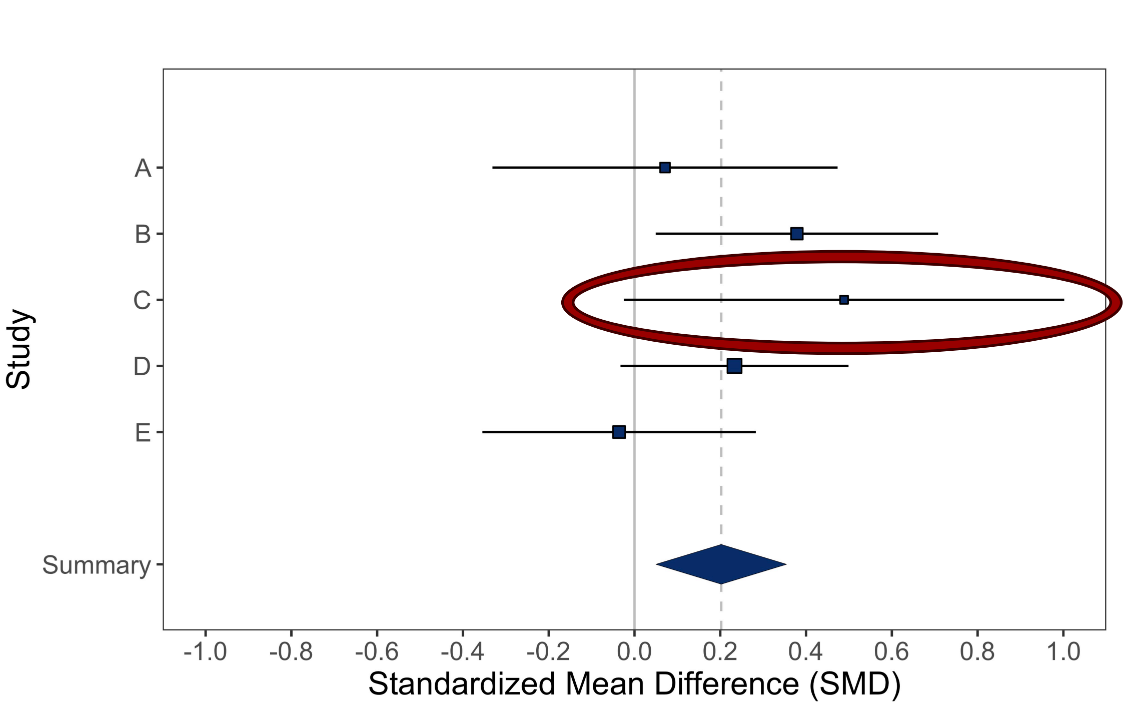
More effective

What about meta-analysis?

Effect sizes as bar plots – yikes!



CI bars for uncertainty – yikes!

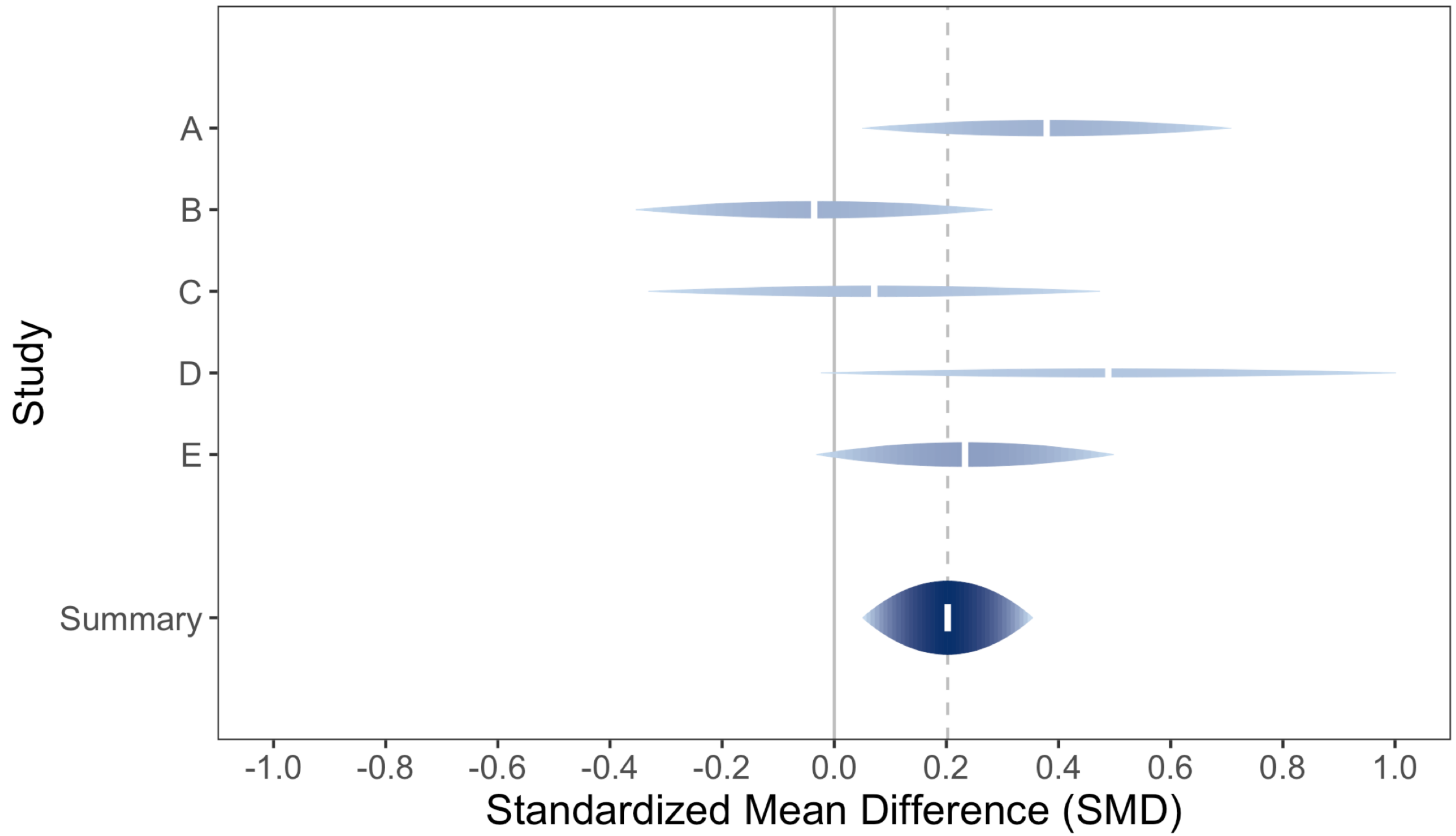


Caution against:

Bar plots & forest plots for meta-analytic data

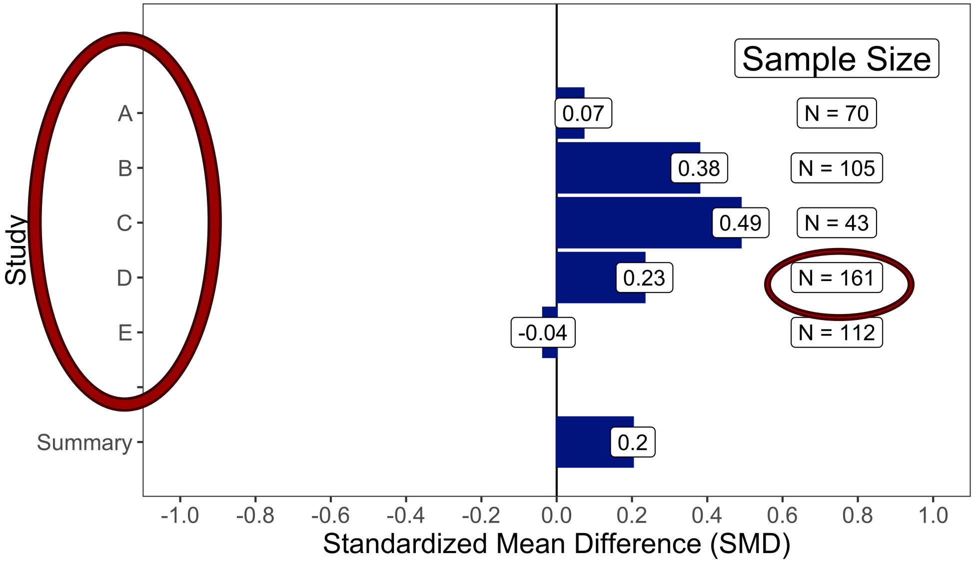
Fitzgerald & Tipton (2022)

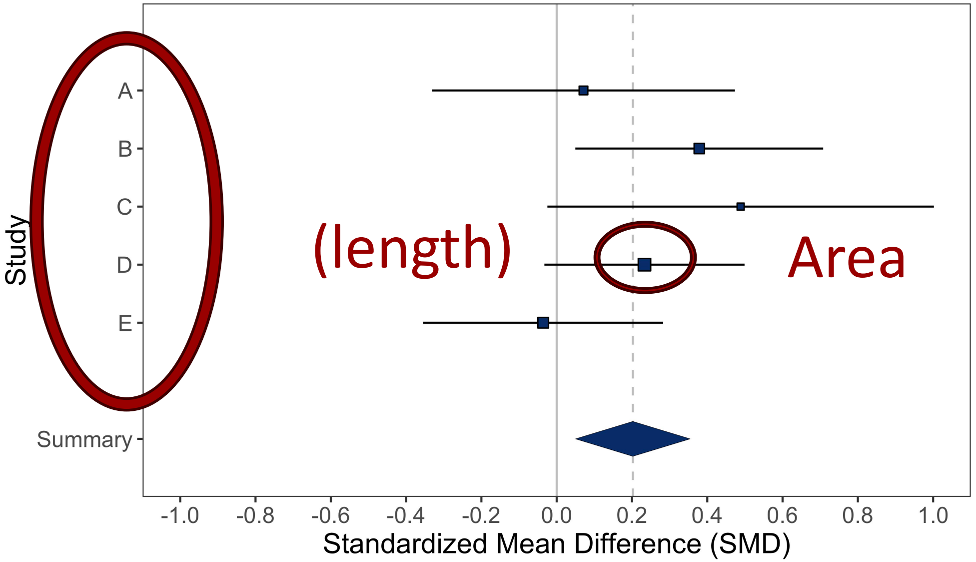
The rainforest plot seems promising?

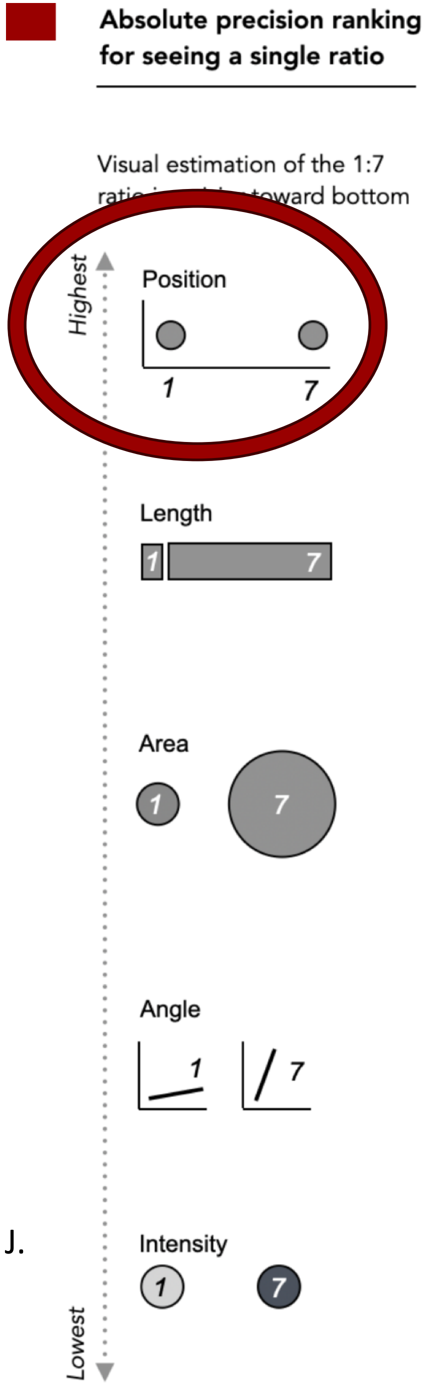


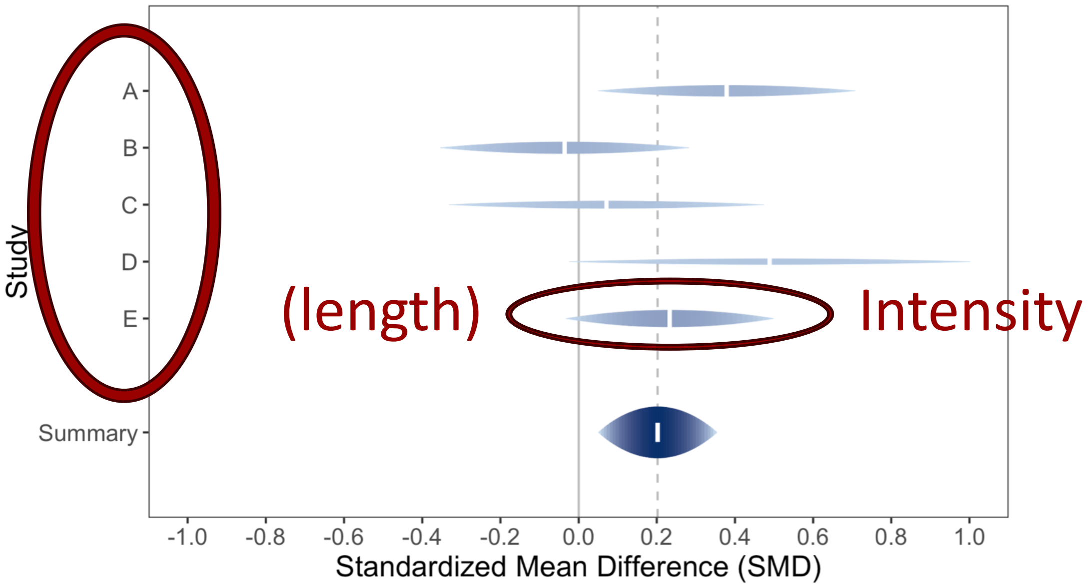
Curse of expertise!
Complex encodings

Schild & Voracek (2015)







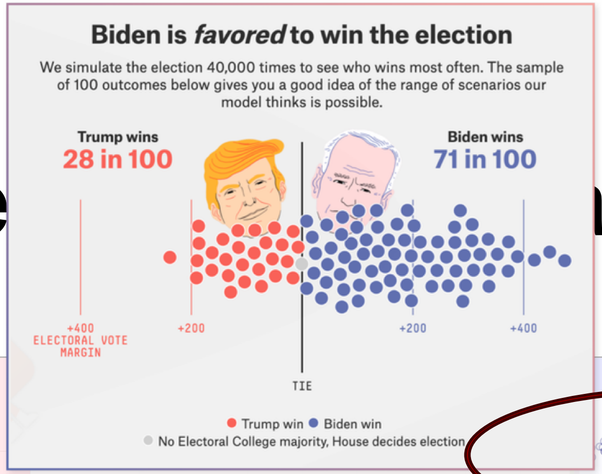


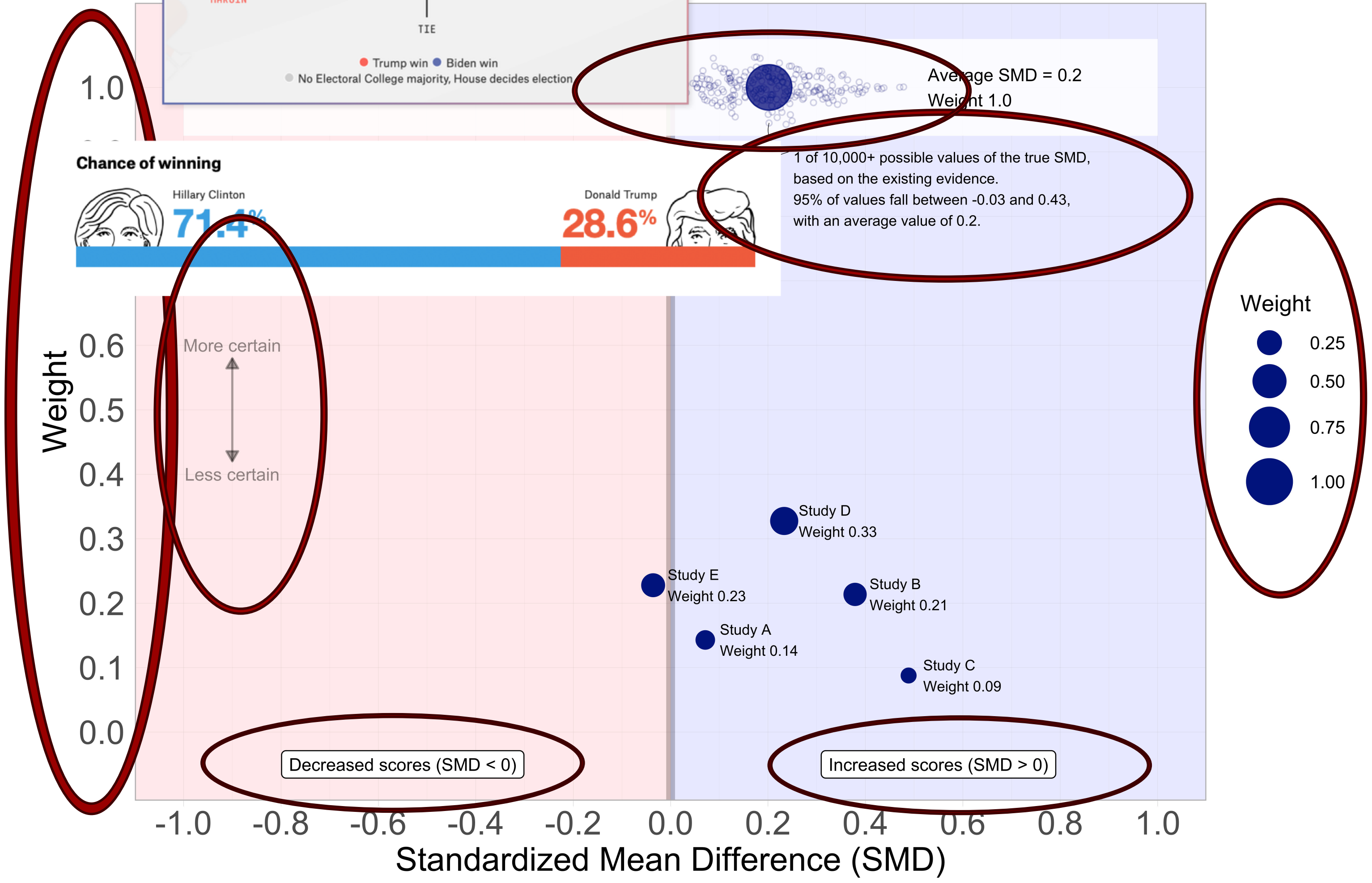
Key to meta-analytic reasoning:

More precise effects get more weight

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works.

Me in Cloud (MARC) Plot





Recommend:

Make meta-analytic weight salient

Utilize y-axis

Simple encodings

Continuous (and individual outcomes) display of uncertainty

Use annotations to guide interpretation

Fitzgerald & Tipton (2022)

Experimental design $(4 * 2^2)$

Factor A: Visualization type

Factor B: Statistical significance of summary effect

Factor C: Magnitude of summary effect

A = Viz type

- $a_1 = FP$
- $a_2 = BP$
- $a_3 = RFP$
- $a_4 = MARC$

×

	C -	C +
B -	NS, 0.10	NS, 0.20
B +	S, 0.10	S, 0.20

Participants

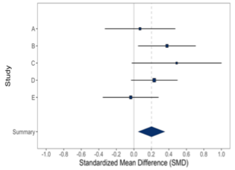
EDUCATION
PRACTITIONERS

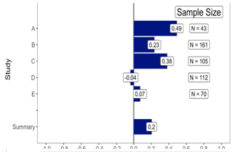
N = 83

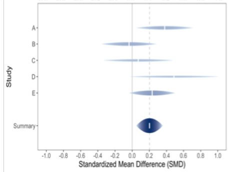
EDUCATION
RESEARCHERS

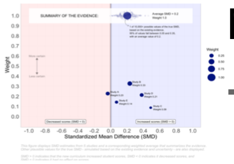
N = 94

Can practitioners accurately interpret the meta-analytic data?









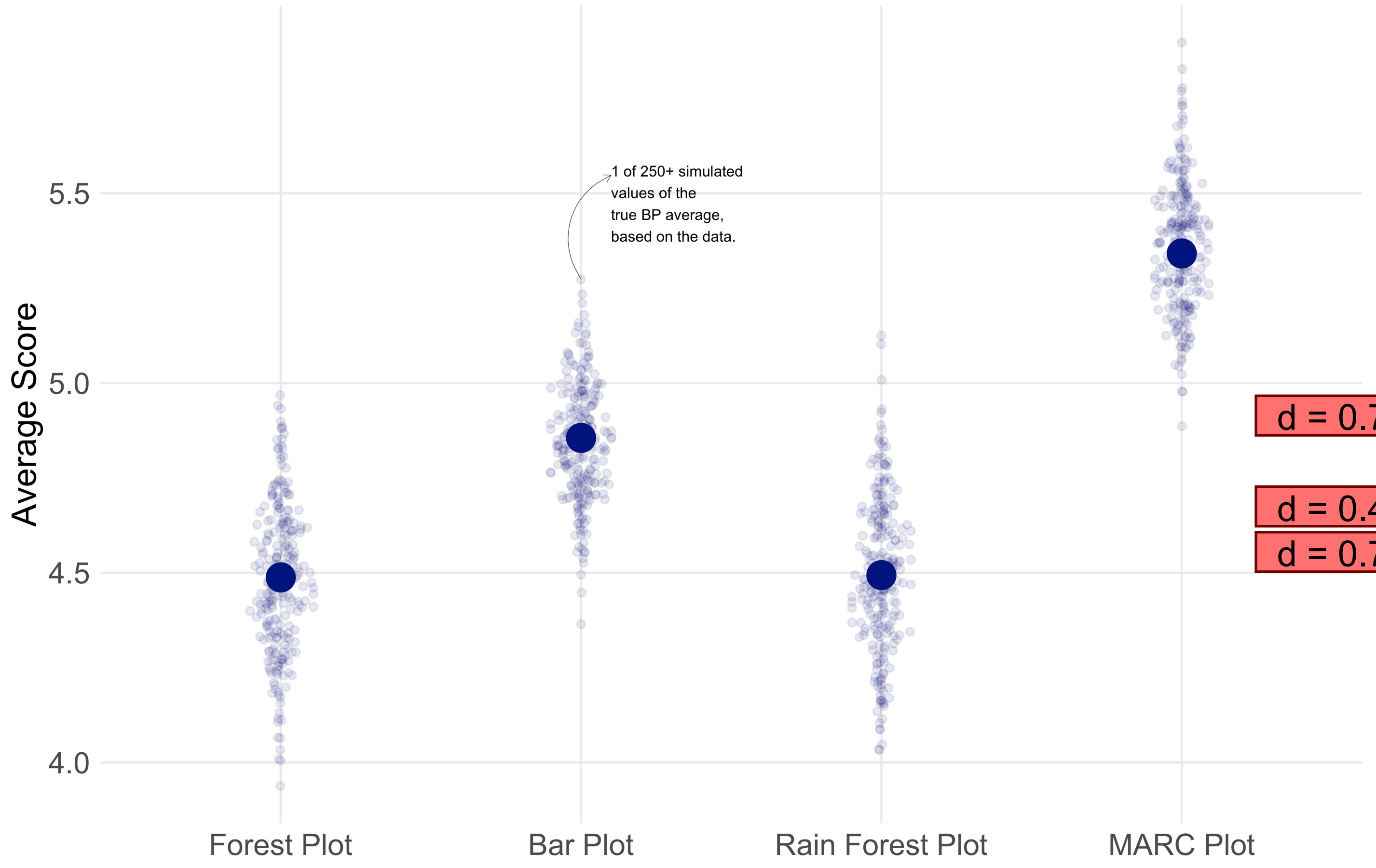
Visualization	n	Q1 Trust Most	Q2 Most Weight	Q3 Least Certain	Q4 Largest SMD	Q5 Avg SMD	Q6 Best estimate	Q7 Sufficient evidence
FP	82	0.512	0.573	0.683	0.866	0.780	0.598	0.476
BP	83	0.759	0.554	0.687	0.904	0.831	0.663	0.458
RFP	81	0.580	0.580	0.617	0.827	0.802	0.667	0.420
MARC	82	0.720	0.951	0.890	0.866	0.805	0.610	0.500

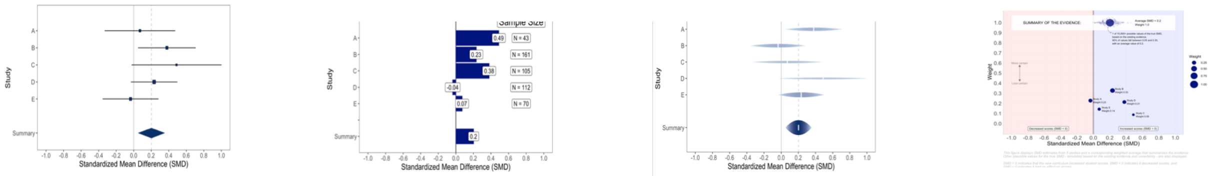
Which study w

in determining

Positional encodings work

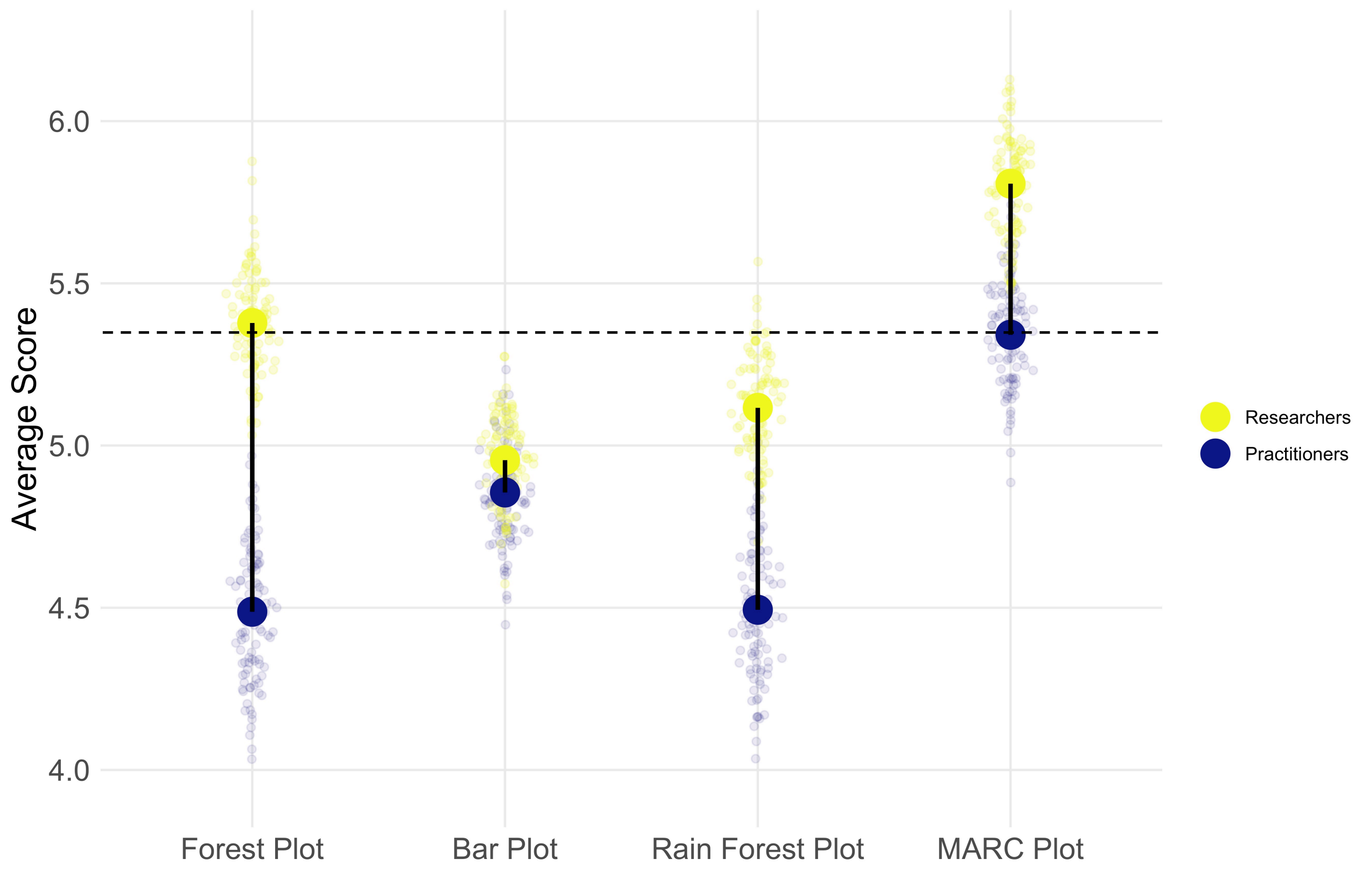
MARC Plots perform better



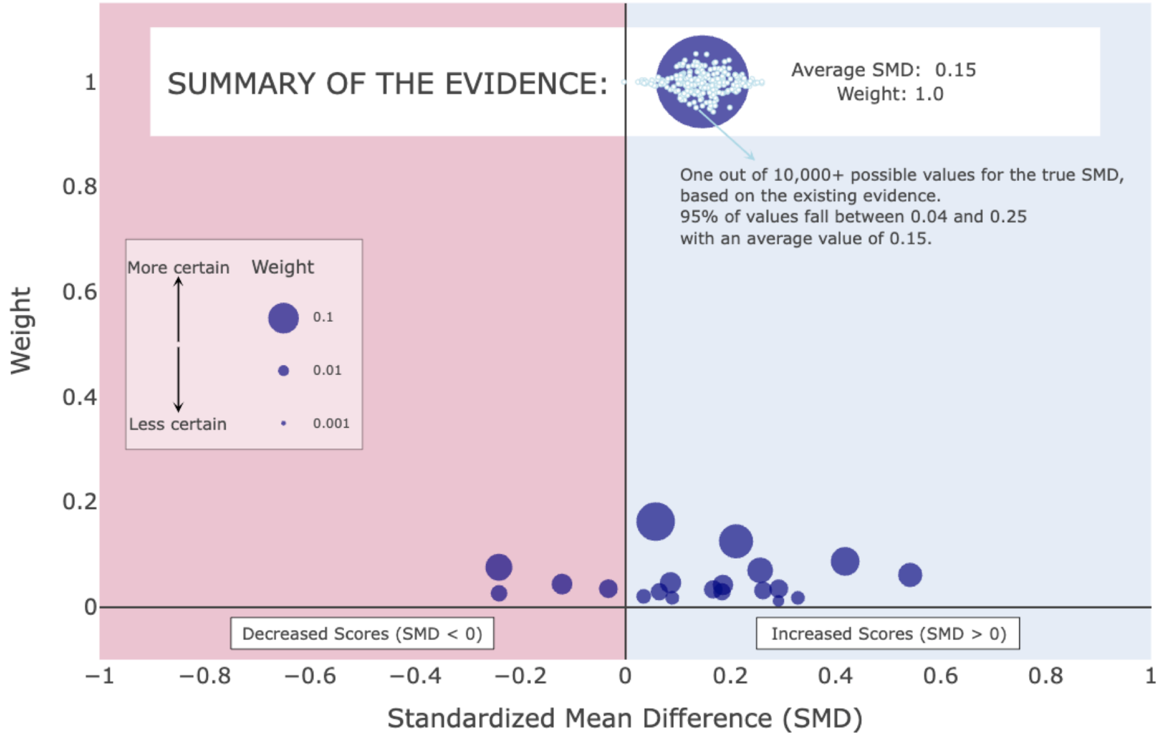


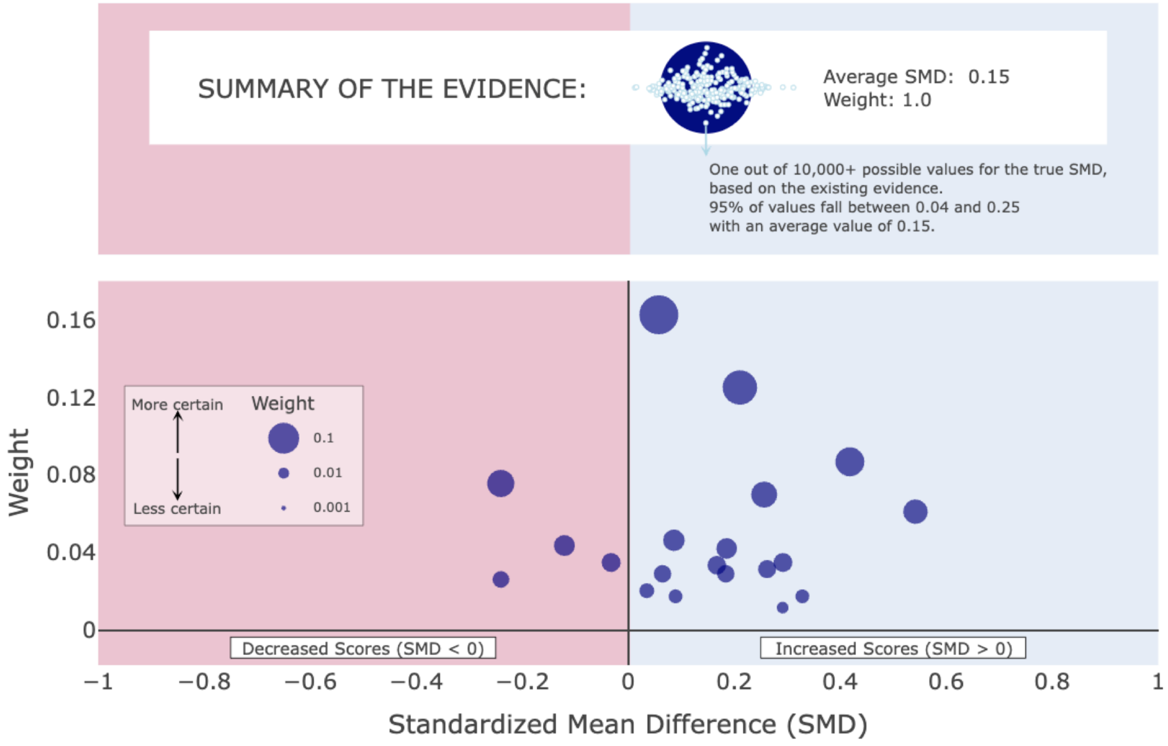
	Contrast	Difference Estimate	Lower	Upper	Adjusted P-value
	BP-FP	0.368	-0.084	0.819	0.153
	RFP-FP	0.006	-0.448	0.460	1.000
d = 0.76	MARC-FP	0.854	0.401	1.306	0.000
	RFP-BP	-0.362	-0.814	0.091	0.167
d = 0.43	MARC-BP	0.486	0.035	0.937	0.029
d = 0.76	MARC-RFP	0.848	0.394	1.302	0.000

Researchers vs. Practitioners

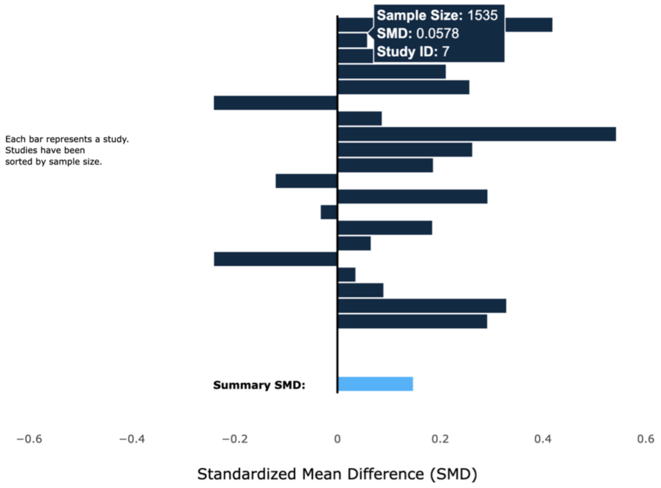


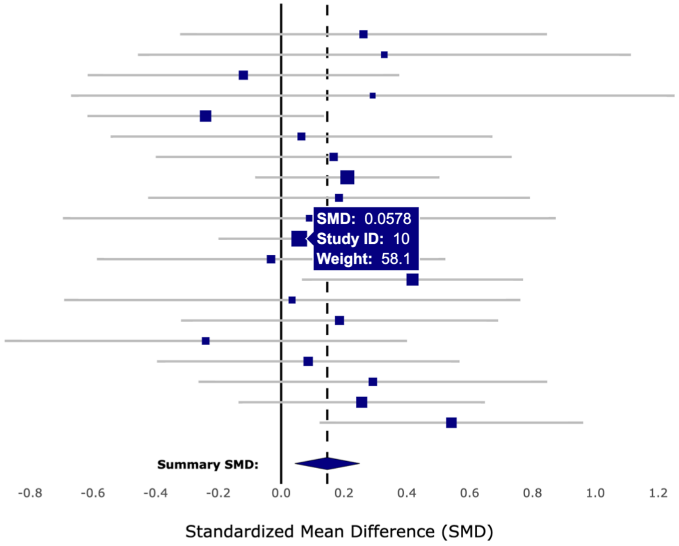
Concerns about large k

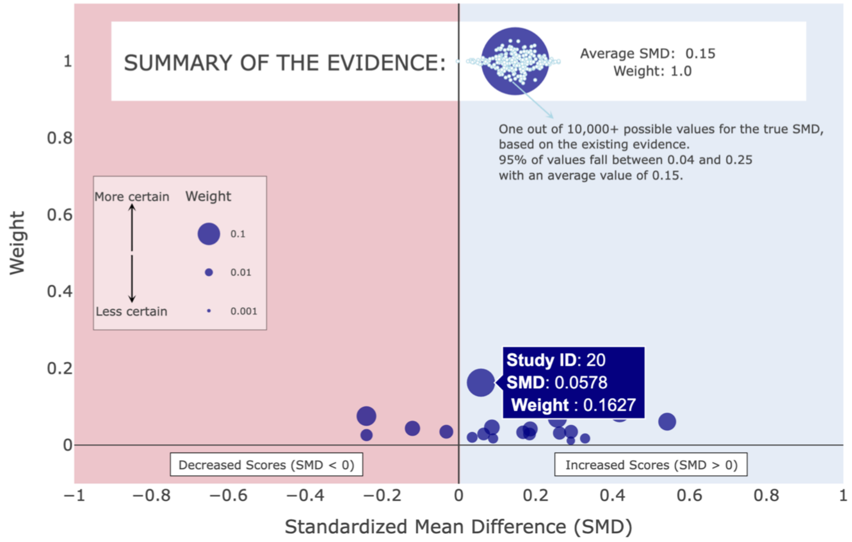


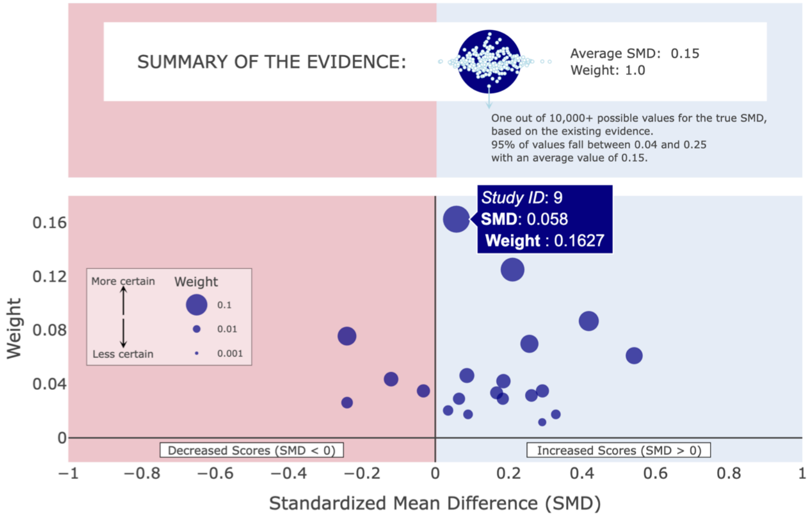


4*4 factorial design



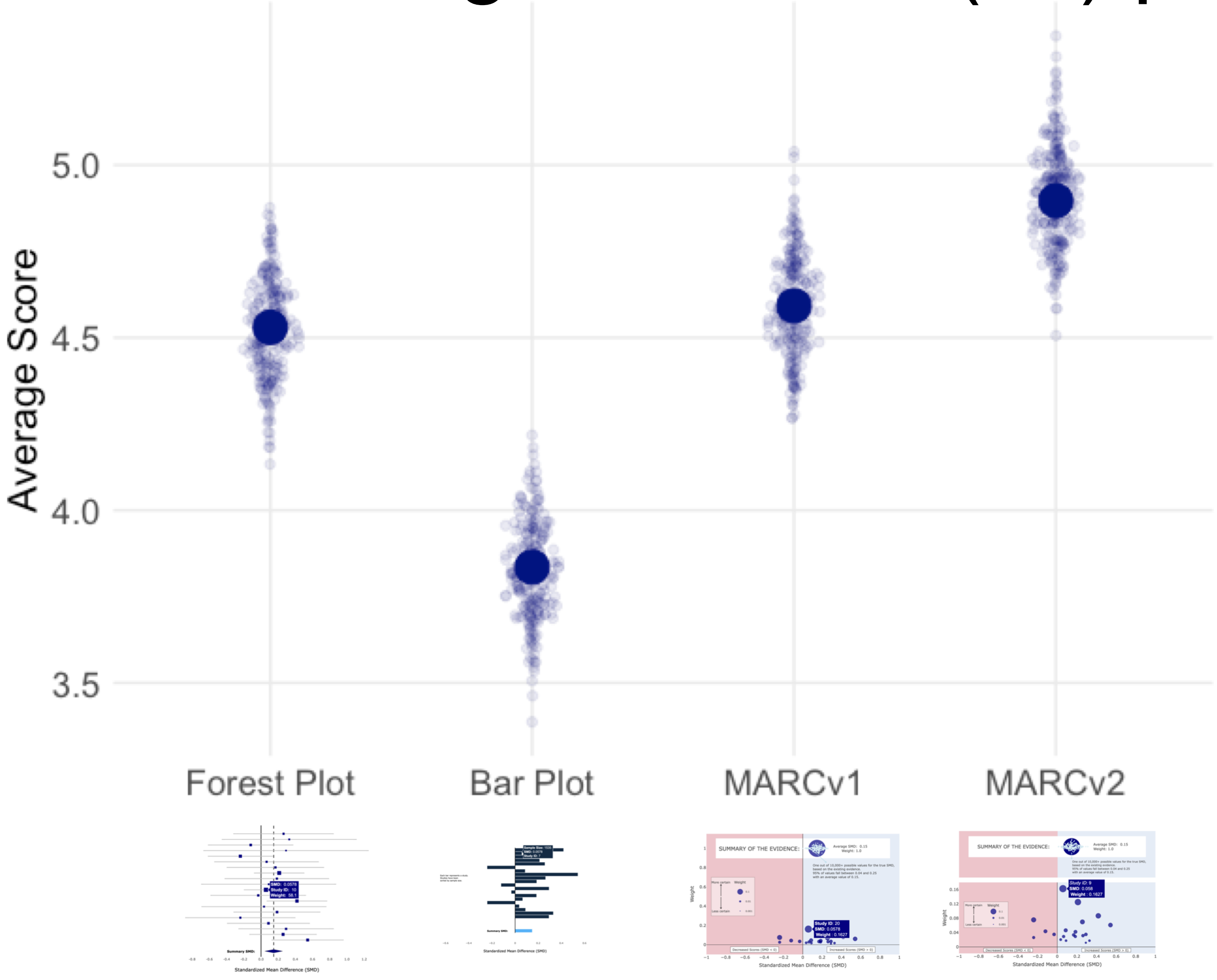






k
10
20
50
100

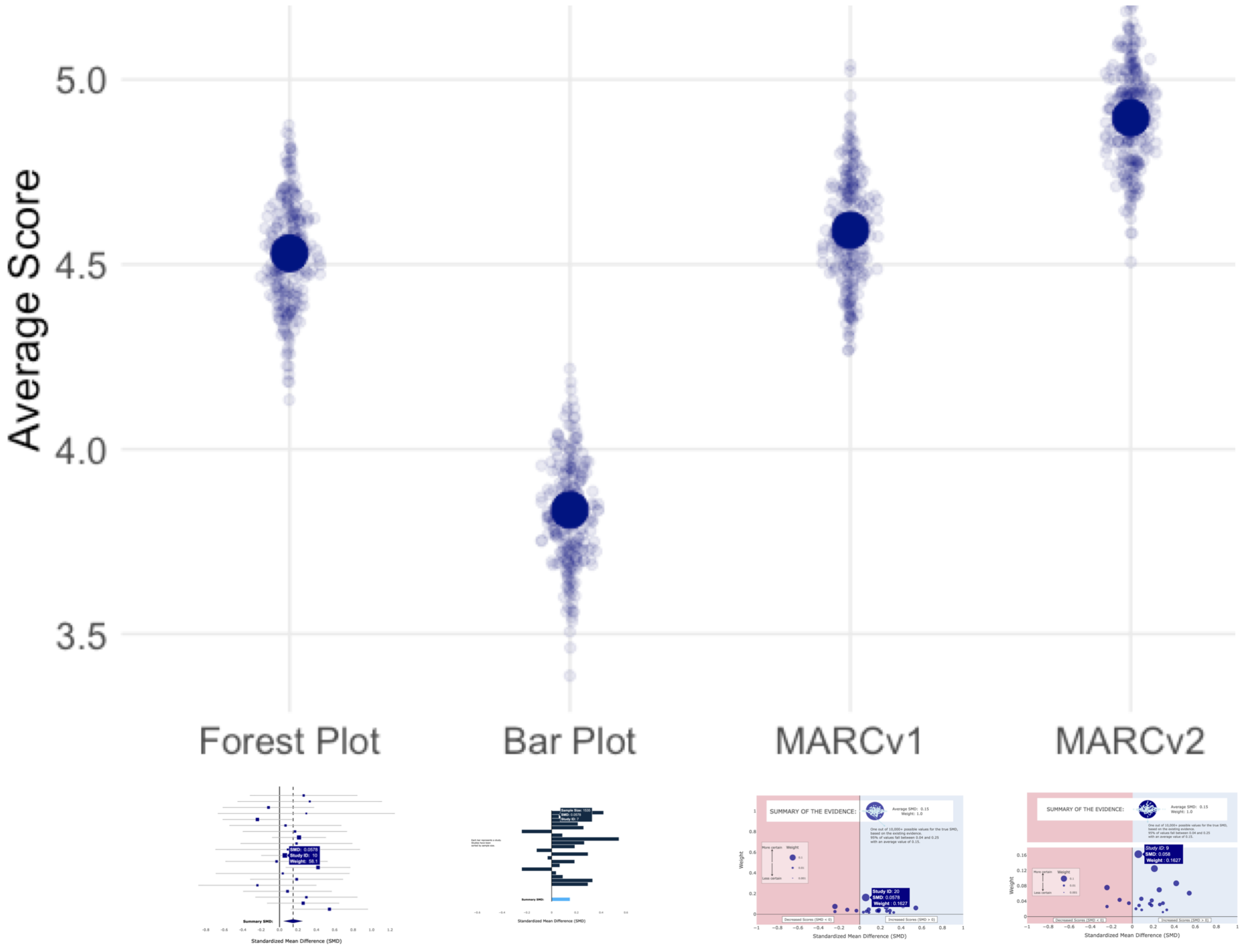
Advantage of MARC(v2) persists for large k

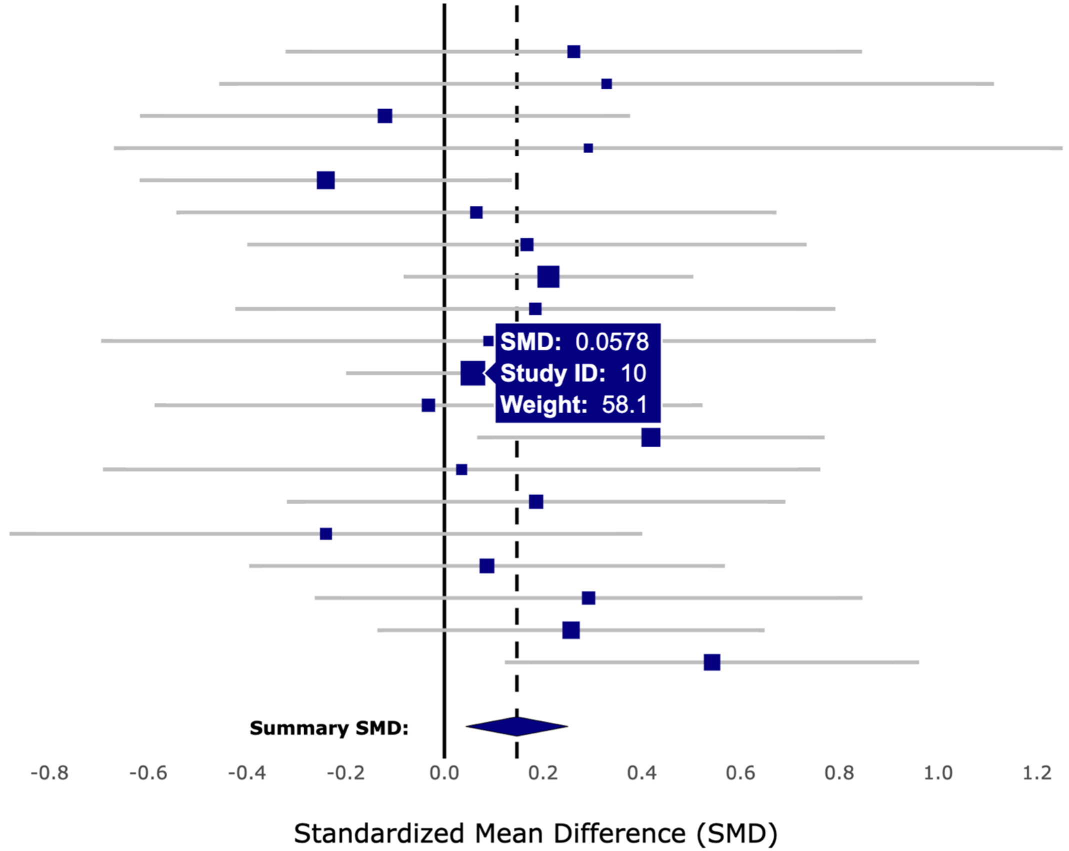


	Contrast	Difference Estimate	Lower	Upper	Adjusted P-value
	BP-FP	-0.694	-1.002	-0.387	0.000
	MARCv1-FP	0.063	-0.240	0.365	0.951
d = 0.36	MARCv2-FP	0.367	0.064	0.670	0.010
	MARCv1-BP	0.757	0.452	1.062	0.000
d = 1.03	MARCv2-BP	1.061	0.755	1.367	0.000
d = 0.30	MARCv2 -MARCv1	0.304	0.004	0.605	0.045

Advantage of MARC(v2) persists for large k

Forest plot seemed to improve w/ the hover text

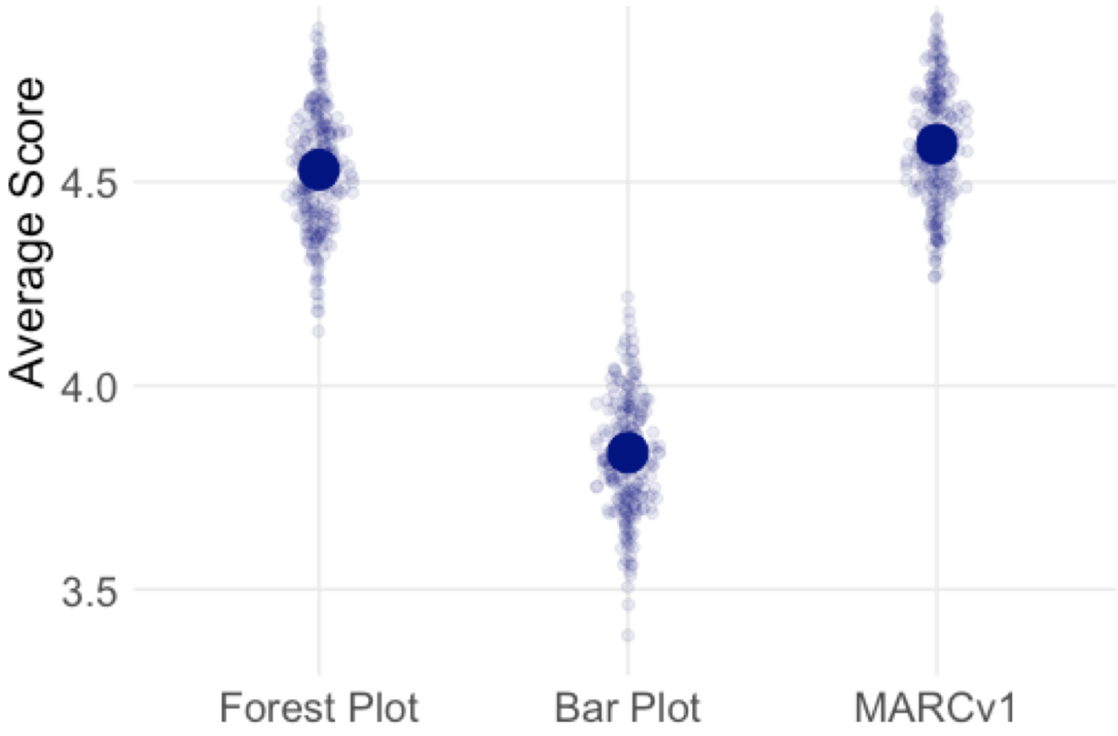


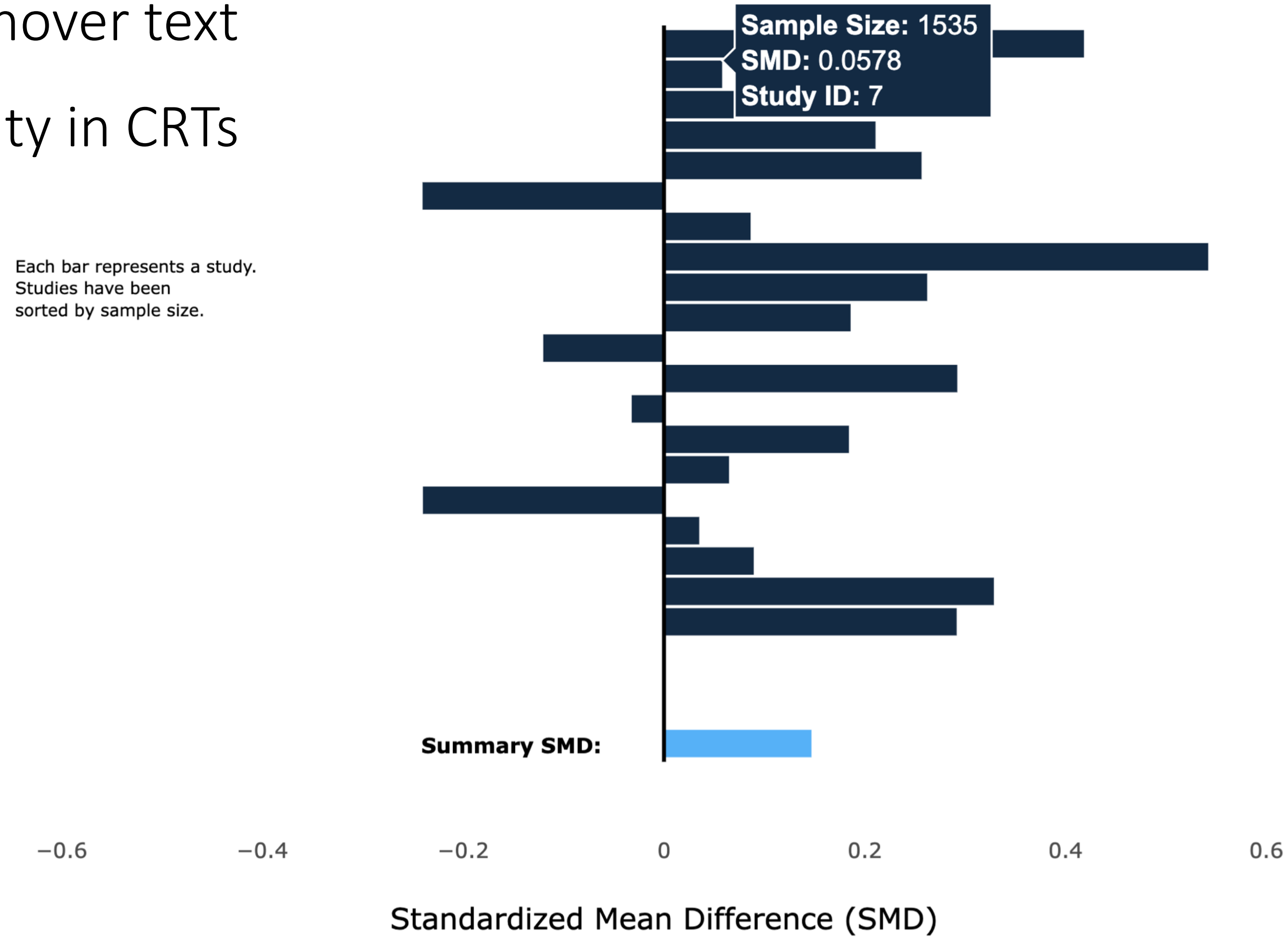


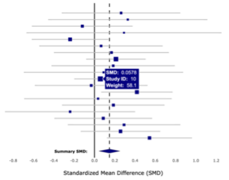
Advantage of MARC(v2) persists for large k

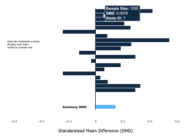
Forest plot seemed to improve w/ the hover text

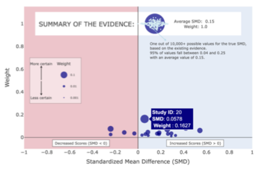
Sample size as poor proxy for uncertainty in CRTs

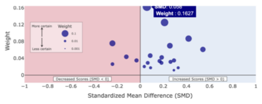












Future directions

- R package on CRAN
  - Bare bones version currently available on GitHub ☺
  - https://github.com/kgfitzgerald/MARCviz
- Encoding other study characteristics. How to help people reason about subgroup effects and moderators?
  - Need descriptive, normative, and prescriptive work here!
- Comparison of multiple interventions
  - More realistic to decision-making process

Takeaways

Beware of the curse of expertise

Let’s examine our own norms and evaluate our own practices

We need (more) evidence on how decision-makers engage with meta-analytic evidence & their decision-making needs

We need healthy feedback loops between normative, descriptive, prescriptive work – an integrated science – to establish best practices for mobilizing knowledge

Thanks!

Email: kfitzgerald@apu.edu
Twitter: @fitzgerald_kg
MARCv2 code: https://github.com/kgfitzgerald/MARCviz

What visualization / evidence communication scenarios would you like to see tackled?

Funded by IES
Grant Award R305B140042

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. Psychological Methods, 10(4), 389–396. https://doi.org/10. 1037/1082-989X.10.4.389

Bell, D. E., Howard, R., Amos, T. (1988). Descriptive, normative, and prescriptive interactions in decision making. In D. E. Bell, H. Raiffa, and A. Tversky (Eds.), Decision making: Descriptive, normative, and prescriptive interactions. Cambridge University Press. https://doi.org/10.1017/CBO9780511598951.003

Beyth-Marom, R., Fidler, F., & Cumming, G. (2008). Statistical cognition: Towards evidence- based practice in statistics and statistics education. Statistics Education Research Journal, 7(2), 20–39. https://doi.org/10.52041/serj.v7i2.468

Bowers, A. J. (2021). Data visualization, dashboards, and evidence use in schools: Data collabora- tive workshop perspectives of educators, researchers, and data scientists. https://doi.org/10.7916/ d8-jj2g-e225.

Bowers, A. J., & Krumm, A. (2021). Supporting the initial work of evidence-based improvement cycles through a data-intensive partnership. Information and Learning Sciences, 122(9/10), 629– 650. https://doi.org/10.1108/ILS-09-2020-0212

Coburn, C. E., Penuel, W. R., & Farrell, C. C. (2021). Fostering educational improvement with research-practice partnerships. Phi Delta Kappan, 102(7), 14–19. https://doi.org/10.1177/ 00317217211007332

Conaway, C. (2021). Funding research that is useful and used. Education Next (blog). https:// www.educationnext.org/funding-research-that-is-useful-and-used-20-years-institute-education- sciences/

Correll, M., & Gleicher, M. (2014). Error bars considered harmful: exploring alternate encodings for mean and error. IEEE Transactions on Visualization and Computer Graphics, 20(12), 2142– 2151. https://doi.org/10.1109/TVCG.2014.2346298

Fitzgerald, K. G., & Tipton, E. (2022). The meta-analytic rain cloud plot: A new approach to visu- alizing clearinghouse data. Journal of Research on Educational Effectiveness, 15(4), 848–875. https://doi.org/10.1080/19345747.2022.2031366

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. Psychological Science in the Public Interest, 22(3), 110–161.

Lortie-Forgues, H., Sio, U. N., & Inglis, M. (2021). How should educational effects be communi- cated to teachers? Educational Researcher, 50(6), 345–354. https://doi.org/10.3102/0013189X2 0987856

Nakajima, N. (2021). Evidence-based decisions and education policymakers. Unpublished Paper. https://nozominakajima.github.io/files/nakajima_policymaker.pdf.

National Academies of Sciences, Engineering, and Medicine. (2022). The future of education research at IES: Advancing an equity-oriented science. The National Academies Press. https:// doi.org/10.17226/26428

Vivalt, E., Coville, A., & Kc, S. (2022). Weighing the evidence: Which studies count? https://evavivalt.com/wp-content/uploads/Weighing-the-Evidence.pdf

Descriptive – examples from education

Educational Researcher

AERA AMERICAN EDUCATIONAL RESEARCH ASSOCIATION FOUNDED 1916

Impact Factor: 8.2 /

Open access | Research article | First published online January 12, 2021

How Should Educational Effects Be Communicated to Teachers?

Hugues Lortie-Forgues, Ut Na Sio, and Matthew Inglis View all authors and affiliations

Volume 50, Issue 6 | https://doi.org/10.3102/0013189X20987856

Find: the metric on which evidence is presented greatly influences teachers' level of engagement with the evidence as well as their perception of the effectiveness of the intervention.

Caution against:

Months of progress as an effect size

Prescriptive – examples from education

Evidence-Based Decisions and Education Policymakers

Nozomi Nakajima *

November 2021

Find: Strong preference for external validity compared to internal validity. Policymakers do update their beliefs in response to research evidence, but that the effect is large and persistent only when the explanation provided for how the evidence was generated is brief and accessible.