IMPROVING TRUST AND USE OF EDUCATION RESEARCH

by

Claire Margaret Chuter

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

May 2023

Abstract

Despite over two decades of steady improvement in research on what works in education policy and practice, the evidence on *how* best to deploy these findings is still very weak. While anecdotal data and professional expertise are inherently valuable – use of evidence *and* professional expertise are not mutually exclusive. On the contrary, by relying on high quality evidence for decisions where the data exists, education practitioners can be freed to refocus their mental energy to other day-to-day decisions. The purpose of this present triad of studies is (1) to better understand the landscape of obstacles that impede research use in education, (2) to test the impact of a strategy to mitigate one such obstacle, and (3) to identify how publication bias inflates the effects of education outcomes. The findings herein indicate that trust is an underlying issue in research use, and that interventions may support improvements in research use if meta-analysts work to insure more reliable, uninflated results. This work may be used to guide research translation efforts, identify ways to increase replicability of effects, and ultimately to improve uptake of high-quality research in education.

Primary Reader and Advisor: Dr. Hunter Gehlbach

Secondary Readers: Drs. Amanda Neitzel; Robert Balfanz; Alan Cheung; Rebecca Cruz

Dedication

To Aaron – my love and partner through all of life's adventures.

Table of Contents

List of Tables

List of Figures

Chapter 1: Literature Review

Introduction

To improve the state of education, rigorous evidence must inform practice and policy. This is not a new or controversial perspective. In fact, the evidence-based reform movement has become very popular among policymakers themselves over the past twenty years. In his presidential campaign, George W. Bush made the following claim:

> We will start by funding only what works in education, only those methods and ideas that prove their power to close the achievement gap. . . .My Administration will require every Federal program -- in teacher training, curriculum research, school safety -- to prove results. If it cannot, we will shift the money into a program that is using it wisely.
>
> Gov. George Bush
>
> September 2nd, 1999

Delivering on these campaign promises, President Bush pushed forward one of his signature pieces of legislation: the No Child Left Behind Act (2001). NCLB prioritized evidence for federal education funding – cementing its place in education practice. This policy was subsequently echoed and made even more stringent in terms of how evidence was quantified, through NCLB's reauthorization in the Every Student Succeeds Act (2015). This emphasis on rigorous study designs standards as tied to federal education funding is still in effect today.

While the need for knowledge utilization in policy development has become greater than ever in an era of evidence-based policymaking, there has been considerable disagreement over what and *whose* knowledge should have an impact on policy decisions. On the one hand, a focus on rigorous experiments evaluating replicable programs and practices is essential to build confidence in educational research among policymakers and educators. The hope is that these policy reforms could set in motion a process of research and development on programs and practices affecting children everywhere. This process could create the kind of progressive, systematic improvement over time that has characterized successful parts of our economy and society throughout the 20th century, in fields such as medicine, agriculture, transportation, and technology.

Yet, there is still a need for correlational and descriptive research, as well as anecdotal experience in education. Critiques of evidence-based reform, in particular the emphasis on randomized controlled trials, rightly point out that experimentation in the social sciences takes place in imperfect conditions. In most designs in education, both the inputs and the outcomes are complex. Much of the important work in education is centered around theoretical reasoning, ethical considerations, and understanding of historical and demographic context. For this reason, Bhaskar (1989) and others argue that it is not possible to design the sorts of controlled experiments that decisively allow for the elimination of alternative explanations.

Similarly, Hammersley argued that teaching practice cannot be based directly on research evidence because it needs to be filtered through teachers' experiences and understandings. Models of evidence-based reform direct attention and value away from teachers' funds of knowledge. Such informal knowledge is invaluable in daily decisions as teachers evaluate student understanding and struggles. Thus, as emphasis is drawn towards evidence-based

decision-making, teachers' professional expertise in devalued. Each of these critiques has merit – teachers have an invaluable wealth of knowledge, and understand the complex and unique contexts in which they work better than any outside researcher. This subsequent work outlined in this thesis does not seek to undervalue these contributions. On the contrary – given the high amounts of stress and low reserves of free time afforded to educators, it is more important than ever to offload some of the labor of figuring out what works in education. For this reason, this thesis seeks to delve deeper into the understudied field of research use in education, to better serve educators and education leaders.

Despite the critiques described here, on the surface, evidence-based reform seems to have widespread support. When asked about evidence-based reform, policymakers and practitioners are largely in favor of using research to guide decision-making (Coburn & Talbert, 2006; Datnow & Hubbard, 2016; Rickinson et al., 2021). Yet in practice, the situation becomes much less clear. Although the evidence-based reform movement has gained considerable support in the political arena over the past two decades (No Child Left Behind [NCLB] 2002; Slavin, 2002), high quality evidence remains an afterthought in most education practice and policy decisions (Cousins & Walker, 2000b; Williams & Coles, 2007). As is often the case, in the real world, idealized ambitions seem to have given way to more pressing concerns. When research *is* used, education stakeholders are regularly disappointed by lower-than-expected effects (Gehlbach & Robinson, 2021; Ioannidis, 2005a; Makel & Plucker, 2014a). This understandably contributes to a vicious cycle of furthering even greater hesitancy to depend on research in the future. Some have suggested that publication bias may contribute to inflated results in the published literature. Thus, mitigating such bias should be a critical goal for those trying to repair the research-practice relationship.

The purpose of this present triad of studies is (1) to better understand the landscape of obstacles that impede research use in education, (2) to test the impact of a strategy to mitigate one such obstacle, and (3) to identify how publication bias inflates the effects of education outcomes. This work may be used to guide research translation efforts, identify ways to increase replicability of effects, and ultimately to improve uptake of high-quality research in education.

Advances in Education Research

It is not controversial that high-quality research should be used to inform education practice and policy. Over the past twenty years, supporting evidence-based reform has become a staple of legislation in education (NCLB, 2002; ESSA, 2015). In the early 2000s, two pieces of legislation ushered in a sea change in education. First, the U.S. Congress passed the No Child Left Behind (NCLB) Act in 2001, followed by the Education Sciences Reform Act (2002). NCLB required robust scientific evidence to justify expenditures of federal funds on education interventions, products, and services, thereby generating demand for high-quality education research (NCLB, 2002; Congress, 2002). To meet this demand, The Education Sciences Reform Act established the Institute of Education Sciences (IES) - an agency deliberately shielded from political influence, and provided with the authority to fund independent research (Whitehurst, 2008). IES subsequently established the What Works Clearinghouse (WWC) to help practitioners and policymakers in education distinguish effective from ineffective programs and strong from weak evidence of effectiveness (Kerstyn, 2004). The WWC established strict standards of evidence for program evaluations(Polanin et al., 2021a; Kerstyn, 2004). These standards assessed methodological and analytical criteria of each program evaluation, including thresholds for attrition, baseline equivalence, and randomization. Each study thus evaluated and publicly labeled according to these standards. Finally, IES created funding streams dedicated to

producing high-quality education research through the National Center for Education Research and the National Center for Special Education Research (Connor et al., 2014).With these two pieces of legislation, both a demand for rigorous research, a process for conducting and identifying such research, and a stable source of funding for high-quality research forged a path for evidence-based practice and policy in education.

To meet this new challenge of high-quality evidence-based reform, many education researchers shifted their emphasis towards experimental and quasi-experimental research methods. Consequently, these methods became more widely used by the academic community. During this period, the WWC also generated a broad scientific knowledge base, identifying dozens of education products, interventions and services as effective or ineffective (J. Polanin et al., 2021b). More recently, other organizations have followed suit. the Center for Research and Reform in Education (CRRE) developed the site Evidence for ESSA (year), and the American Institute of Research (AIR) similarly created the Evidence Support Center (ESC).

These efforts have allowed for much stronger and larger evidence-based for education policy and practice to rest on (Hedges, 2018; Slavin, 2002, 2008). Since 2002, IES has supported over 350 randomized field trials in an effort to establish strong evidence to definitively guide education leaders' decision-making (Howells, 2018). Moreover, a growing focus on the generalizability of program evaluation provides for greater confidence in the effectiveness of an intervention to specific contexts. To this end, methods for adjusting experimental treatment effect estimates for known population characteristics have been developed by Stuart et al., (2015) and Tipton (2014). Other web tools such as "The Generalizer" allow practitioners and policymakers to assess how well the results of a study are likely to apply to their own setting (Tipton & Matlen, 2019).

Use of Research in Education Remains Low

Yet, despite a renewed focus on rigorous methods in education, those seeking to increase the use of research evidence (URE) in education still continue to face challenges at the practice and policy levels. Bérubé (2005) found despite high valuing of education research, 60% to 90% of teachers reported using research either 'never' to 'sometimes.' Williams & Coles (2007) similarly reported that 60% to 80% of respondents in schools used research related to teaching and learning 'infrequently.' In a selective review of the on grammar curricula and pedagogy, Larsen-Freeman found, "grammar instruction has been relatively unaltered by research findings" (p. 273, 2015).

Policymakers' use of education research is similarly low. Natow (2020) found that while research is sometimes used in the policymaking process, factors other than research were discussed more frequently. Instead, policymakers regularly relied on personal anecdotes, colleague recommendations, and their own intuition to justify their position for certain policies (Cousins & Walker, 2000a). Peer-reviewed research was seldom mentioned in policy debates, and when it was it served to substantiate rather than guide decision-making (Rubin et al., 2022).

Movement to Improve Use of Research Evidence

This research-practice gap has not gone unnoticed. URE in education as a field of study has grown considerably in tandem with legislative pressure to include evidence in education decision-making (Shewchuk & Cooper, 2018). Also gaining widespread agreement is the view that intermediary organizations have an important role to play in facilitating multi-stakeholder partnerships between researchers, practitioners and policymakers in order to increase the knowledge mobilization of research and its impact in public service sectors (Cooper, 2014; Levin, 2013) . Knowledge mobilization includes efforts to increase the use of research evidence

in policy and practice in education, and tends to occur through iterative, social processes involving interaction among two or more different groups or contexts (researchers, policymakers, practitioners, third party agencies, community members). Because practitioners rarely come into contact with primary research directly, intermediaries often take on the work of KMb (Neal et al., 2022).

The primary brokering function of intermediary organizations is to increase accessibility and engagement with research by simplifying academic work (Neal et al., 2022). Intermediaries also take part in advocacy and relationship-building with education decision-makers. Intermediaries come in many forms, from vendors (Bodilly et al., 1998), professional development associations (e.g., National Science Teaching Association; Camino, 1998), community/professional development coaches (Urban Strategies Council, 1996), universities (Harkavy & Puckett, 1991), to non-profit reform organizations (e.g., AIR; Kronley & Handley, 2003).

The expanding role of URE has provided an abundance of qualitative data around educators' and legislators' thoughts regarding why, how, and when they use (or not use) evidence in their work. This evidence can be more useful if it were gathered together in the form of a systematic review. The first study seeks to organize the findings across included studies into coherent themes to support knowledge brokers. Specifically, this study poses the following research questions:

1. *How do education leaders and policymakers understand and use research in their decision-making?*
2. *Why do education leaders and policymakers not use research, and how might knowledge brokers support increased use?*

Study 1 – The Use of Research in K-12 Decision-Making

In order to better understand the path forward for URE and its efforts, my first study systematically reviewed the way research is used and understood in K-12 education decision-making, and what education leaders perceive to be the main barriers to its use. This review informs current efforts by knowledge brokers, policymakers and researchers to encourage use of evidence in education. This review provides insight to knowledge brokers of how and when they can support URE by education leaders and policymakers. By reviewing patterns across many high-quality studies, this review seeks to address the above questions with greater confidence than any one study can alone. Finally, the review also highlights patterns across how URE may be moderated by contextual factors.

This review found, in part, that decision-makers mainly fail to use research because they find the presented research irrelevant or not available when a decision needs to be made. Furthermore, beyond these reported barriers to research use, we discovered a widespread lack of trust of intermediary organizations among research users. (See the full paper for a full list of findings).

Study 2 – Motivating Use of Research in Education

Shifting focus to teacher contexts, my second study focused on testing the impact of one promising instructional strategy on future use of education research. Qualitative research in teacher contexts has found that one of the largest barriers to research use is comprehension of the research itself. While education leaders and policymakers often have staffers and knowledge brokers working closely with them to translate research findings, teachers are rarely so lucky. This experiment drew from the instructional design literature to identify one way to potentially

support teacher research comprehension. The study evaluates whether including examples and/or images can support understanding and consequently future use of research evidence in practice.

Despite widespread non-experimental research in the field of research use in education, there is - to my knowledge – no experimental work yet on this topic. This is to be expected of an emerging field, but in order to move forward with increasingly confident conclusions, experimentation is the logical next step. The valuable contributions of non-experimental, largely qualitative researchers have given shape to areas that are ripe for further focused evaluation. This study seeks to provide this focused evaluation, and in doing so make an important first step for the field of URE into the world of experimental research.

In this study, I build on previously conducted descriptive research that identifies a major barrier to teachers' use of research to be low comprehension of the research itself. To conceptualize the impact of improving comprehension on a piece of research on use of that research in the classroom, I relied on the expectancy-value theory. This framework posits that valuing of an outcome works in tandem with expectations of success towards that outcome, to moderate motivation (Wigfield & Eccles, 2000). Thus, *both* high valuing and high expectations must be present for motivation towards an outcome to also be high. Applying this framework to the current state of educational research use: teachers seem to struggle with low expectations of success to use research despite their high valuing of success. Thus, an intervention to support teachers' expectations of success through increased capacity may support motivation to use research. The current study seeks to develop and test such an intervention.

In developing such an intervention, I found the Mayer's work on learning strategies 2003) particularly relevant. Across multiple studies, Mayer and colleagues have found that by adding dimension to presented information individuals perform better on "knowledge

acquisition" and "knowledge understanding" (Mayer, 1988, 2002, 2019; Quilici & Mayer, 1996). Mayer has found that by perceiving the same information presented in a variety of ways, readers are better able to identify the relevant information, and integrate the new content with previously learned knowledge (Mayer et al., 2005).

This study examined the impact of images and/or examples alongside a description of research, on *(1) understanding of education research, (2) research use self-efficacy, and (3) future use of research*. Teacher participants were randomly presented with either (1) a description of research only, (2) a description alongside an example of the research being used, or (3) a description alongside an example and a relevant image. Though weak, the results indicated positive effect sizes for every focal outcome in the treatment conditions (compared to the control condition). I hope to continue this work through a larger participant pool to better identify the efficacy of these findings.

Study 3 – Publication Bias in Social-Emotional Learning Interventions

The final study in this triad examines a second key barrier to research use in education. This barrier was identified within the previous review of education leaders' use of research. In this review, I found that policymakers and administrators repeatedly mentioned that they are hesitant to use research because they do not trust that reported findings will replicate in practice (Canfield-Davis et al., 2010; Canfield-Davis & Jain, 2010; Jabbar et al., 2014). Therefore, in order for federal, state, and local governments to confidently direct funding toward educational policies and/or practices that are effective, researchers must address the untrustworthiness of results in our meta-analytic reviews. There is a notable drop in effects between initial published findings, and results that occur in practice (Canfield-Davis & Jain, 2010; Jabbar et al., 2014).

Among the chief concerns that cause this drop in effects is that of publication bias. Thus, in order to improve trust in research use, publication bias must be identified and adjusted for in education reviews.

In this third paper, I conducted an extensive publication bias analysis, comprised of (1) a visual inspection of funnel plot asymmetry, using contour-enhancement, (2) the trim-and-fill method, (3) Egger's regression significance tests, and (4) Vevea and Wood's (2005) stepwise selection model to assess and adjust for publication bias. The results indicated that selection bias by outcome but not by overall studies was the root cause of inflated findings in social-emotional learning interventions. These findings reinforce the value of using multiple publication bias methods of varying theoretical backgrounds in evaluating publication bias.

The first study seeks to better understand the larger issue in weaving evidence into education policy and practice. This review uncovers several barriers, with perhaps trust being the most pervasive. The second study points to one of potentially many interventions that may improve use of research, and the third gets at a key problem with research use – replicability. Together, this collection of studies hopes to guide those in the research use field by identifying key gaps and potential solutions.

References

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*, 182–200.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088–1101.

Bérubé, B. (2005). *L'accès à la recherche en enseignement et son utilisation dans la pratique*. desLibris.

Bodilly, S. J., Keltner, B., Purnell, S., Reichardt, R., & Schuyler, G. (1998). Lessons from New American Schools' scale-up phase. *Santa Monica, CA: Rand*, 79–84.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Bransford, J. D., Stipek, D. J., Vye, N. J., Gomez, L. M., & Lam, D. (2009). The Role of Research in Educational Improvement. In *Harvard Education Press*. Harvard Education Press.

Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, *13*(3), 203–220.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. *Frontiers in Sociology of Education*, 127–162.

Bush, George W. (1999, September 2). *Speech to Latin Business Association in Los Angeles.*

Camino, L. (1998). Building local infrastructure for youth development: The added value of capacity-building intermediary organizations. *Washington, DC: Center for Youth Development and Policy Research, Academy for Educational Development*.

Canfield-Davis, K., & Jain, S. (2010). Legislative Decision-Making on Education Issues: A Qualitative Study. *Qualitative Report*, *15*(3), 600-629.

Canfield-Davis, K., Jain, S., Wattam, D., McMurtry, J., & Johnson, M. (2010). Factors of Influence on Legislative Decision Making: A Descriptive Study. *Journal of Legal, Ethical & Regulatory Issues*, *13*(2).

Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the" fail-safe" statistic in meta-analysis. *Educational and Psychological Measurement*, *50*(2), 233–243.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292.

Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, *30*, 727–744.

Cipriano, C., Naples, L. H., Eveleigh, A., Cook, A., Funaro, M., Cassidy, C., McCarthy, M. F., & Rappolt-Schlichtmann, G. (2023). A systematic review of student disability and race representation in universal school-based social and emotional learning interventions for elementary school students. *Review of Educational Research*, *93*(1), 73–102.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, *112*(4), 469–495.

Coburn, C. E., & Turner, E. O. (2011). Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Congress, U. S. (2002). *Education Sciences Reform Act*. Washington, DC: Author.

Connor, C. M., Alberto, P. A., Compton, D. L., & O'Connor, R. E. (2014). Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: A Synthesis of the Contributions from the Institute of Education Sciences Research Centers. NCSER 2014-3000. *National Center for Special Education Research*.

Cooper, A. (2014). Knowledge mobilisation in education across Canada: A cross-case analysis of 44 research brokering organisations. *Evidence & Policy*, *10*(1), 29–59.

Council, U. S. (1996). What is a community building intermediary. *What Work Does It Do*.

Cousins, J. B., & Walker, C. A. (2000a). Predictors of educators' valuing of systematic inquiry in schools. *Canadian Journal of Program Evaluation*, 25–52.

Cousins, J. B., & Walker, C. A. (2000b). Predictors of educators' valuing of systematic inquiry in schools. *Canadian Journal of Program Evaluation*, 25–52.

Dagenais, C., Lysenko, L., C. Abrami, P., M. Bernard, R., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of empirical research. *Evidence & Policy*, *8*(3), 285–309.

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Palo Alto, CA: Learning Policy Institute.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2

Dear, K. B., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science*, 237–245.

Dermody, C. M., Dusenbury, L., Greenberg, M., Godek, D., Connor, P., Cross, R., Martinez-Black, T., Solberg, S., Kroyer-Kubicek, R., & Atwell, M. (2022). A Developmental Framework for the Integration of Social and Emotional Learning and Career and Workforce Development. *Collaborative for Academic, Social, and Emotional Learning*.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432.

Duval, S. (2005). The trim and fill method. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 127–144.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Dynarski, M. (2015). Using research to improve education under the Every Student Succeeds Act. *Evidence Speaks Reports*, *1*(8), 1–6.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634.

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking Connections Between Research and Practice in Education: A Conceptual Framework. *Educational Researcher*, *47*(4), 235–245. https://doi.org/10.3102/0013189X18761042

Forness, S. R. (2001). Special education and related services: What have we learned from meta-analysis? *Exceptionality*, *9*(4), 185–197.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication Bias in Special Education Meta-Analyses. *Exceptional Children*, *83*(4), 428–445. https://doi.org/10.1177/0014402917691016

Gehlbach, H., Mu, N., Arcot, R. R., Chuter, C., Cornwall, K. J., Nehring, L., Robinson, C. D., & Vriesema, C. C. (2022). Addressing the vexing educational challenges of biodiversity loss: A photo-based intervention. *Contemporary Educational Psychology*, *71*, 102096.

Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296–315.

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79–89.

Gehlbach, H., Robinson, C. D., Vriesema, C. C., Bernal, E., & Heise, U. K. (2022). Worth more than 1000 words: How photographs can bolster viewers' valuing of biodiversity. *Environmental Conservation*, *49*(2), 99–104.

Ginsberg, Y. C., Hollands, F. M., Holmes, V. R., Shand, R., Evans, P., Blodgett, R., Wang, Y., & Head, L. (2022). Does ESSA Assure the Use of Evidence-based Educational Practices? *Educational Policy*, 08959048221127989. https://doi.org/10.1177/08959048221127989

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in research. Beverly Hills*. CA.: Sage Publications.

Group, J. (n.d.). APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology. Why do we need them? What might they be. *American Psychologist*, *63*(9), 839–851.

Harkavy, I., & Puckett, J. L. (1991). Toward effective university-public school partnerships: An analysis of a contemporary model. *Teachers College Record*, *92*(4), 556–581.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, *7*(2), 246–255.

Hedges, L. V. (2007). *Effect sizes in cluster-randomized designs*.

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*(1), 1–21.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, *21*(4), 299–332.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486.

Heene, M. (2010). *A brief history of the Fail Safe Number in Applied Research* (arXiv:1010.2326). arXiv. http://arxiv.org/abs/1010.2326

Herman, R., Gates, S. M., Arifkhanova, A., Bega, A., Chavez-Herrerias, E. R., Han, E., Harris, M., Tamargo, J., & Wrabel, S. L. (2017). *School leadership interventions under the every student succeeds act: Evidence review*.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Howells, K. (2018). *The future of education and skills: Education 2030: the future we want*.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, *106*(6), 1258–1287.

IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: A meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*(8), 860–869.

Ioannidis, J. P. (2005a). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Ioannidis, J. P. (2005b). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124.

Ioannidis, J. P. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice*, *14*(5), 951–957.

Ion, G., & Iucu, R. (2014). Professionals' perceptions about the use of research in educational practice. *European Journal of Higher Education*, *4*(4), 334–347.

Jabbar, H., La Londe, P. G., Debray, E., Scott, J., & Lubienski, C. (2014). How Policymakers Define 'Evidence': the politics of research use in New Orleans. *Policy Futures in Education*, *12*(8), 1013-1027.

Jin, Z.-C., Zhou, X.-H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, *34*(2), 343–360.

Kronley, R. A., & Handley, C. (2003). *Reforming Relationships: School Districts, External Organizations, and Systemic Change.*

Larsen-Freeman, D. (2015). Research into practice: Grammar learning and teaching. *Language Teaching*, *48*(2), 263–280.

Levin, B. (2013). To know is not enough: Research knowledge and its use. *Review of Education*, *1*(1), 2–31.

Light, R. J., & Pillemer, D. B. (1984). Quantitative procedures. *The Science of Reviewing Research*.

Lin, L., Chu, H., Murad, M. H., Hong, C., Qu, Z., Cole, S. R., & Chen, Y. (2018). Empirical comparison of publication bias tests in meta-analysis. *Journal of General Internal Medicine*, *33*, 1260–1267.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* SAGE publications, Inc.

Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational Research in Educational Practice: Predictors of Use. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *37*(2), 1–26.

Maier, M., VanderWeele, T. J., & Mathur, M. B. (2022). Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. *Campbell Systematic Reviews*, *18*(3), e1256.

Makel, M. C., & Plucker, J. A. (2014a). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316.

Makel, M. C., & Plucker, J. A. (2014b). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304–316.

Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods*, *11*(6), 725–742.

Mayer, R. E. (1988). Learning strategies: An overview. *Learning and Study Strategies*, 11–22.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71.

Mayer, R. E. (2019). Thirty years of research on online learning. *Applied Cognitive Psychology*, *33*(2), 152–159.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R., Veronikas, S., & Shaughnessy, M. F. (2005). An Interview with Richard Mayer. *Educational Psychology Review*, *17*(2), 179–189.

McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, *35*(3), 357–382.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749.

Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution*, *16*(10), 580–586.

Morrison, J. A., Raab, F., & Ingram, D. (2009). Factors influencing elementary and secondary teachers' views on the nature of science. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(4), 384–403.

Natow, R. S. (2020). Research Utilization in Higher Education Rulemaking: A Multi-Case Study of Research Prevalence, Sources, and Barriers. *Education Policy Analysis Archives*, *28*(95), n95.

Neal, J. W., Neal, Z. P., & Brutzman, B. (2022). Defining brokers, intermediaries, and boundary spanners: A systematic review. *Evidence & Policy*, *18*(1), 7–24.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906.

Palmer, T. M., Sutton, A. J., Peters, J. L., & Moreno, S. G. (2008). Contour-enhanced funnel plots for meta-analysis. *The STATA Journal*, *8*(2), 242–254.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., Rushton, L., & Moreno, S. G. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 575–591.

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, *42*(8), 424–432.

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90–100.

Polanin, J., Caverly, S., Pollard, E., Bzura, R., Davis, E., Goldston, C., Gnedko-Berry, N., Lindsay, J., Nolan, E., & Song, M. (2021a). A Reintroduction to the What Works Clearinghouse. *What Works Clearinghouse*.

Polanin, J., Caverly, S., Pollard, E., Bzura, R., Davis, E., Goldston, C., Gnedko-Berry, N., Lindsay, J., Nolan, E., & Song, M. (2021b). A Reintroduction to the What Works Clearinghouse. *What Works Clearinghouse*.

Polanin, J. R., & Espelage, D. L. (n.d.). *The ultimate haystack needle: Locating unreported outcome data for use in a systematic review and meta-analysis Objective*.

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144.

Kerstyn, C. (2004). *What Works Clearinghouse*.

Rickinson, M., Gleeson, J., Walsh, L., Cutler, B., Cirkony, C., & Salisbury, M. (2021). *Research and evidence use in Australian schools: Survey, analysis and key findings*.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 1–7.

Rubin, P. G., Ayscue, J. B., & Uzzell, E. M. (2022). Analyzing Congressional Research Service Reports on Education: How Is Information Used at Congress's Think Tank? *AERA Open*, *8*, 23328584221117170.

Sedgwick, P., & Marston, L. (2015). How to read a funnel plot in a meta-analysis. *Bmj*, *351*.

Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: Practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, *98*(23).

Silliman, N. P. (1997). Hierarchical selection models with applications in meta-analysis. *Journal of the American Statistical Association*, *92*(439), 926–936.

Simpson, R. L., Lacava, P. G., & Sampson Graner, P. (2004). The No Child Left Behind Act: Challenges and Implications for Educators. *Intervention in School and Clinic*, *40*(2), 67–75. https://doi.org/10.1177/10534512040400020101

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, *31*(7), 15–21.

Slavin, R. E. (2008). Evidence-based reform in education: What will it take? *European Educational Research Journal*, *7*(1), 124–128.

Stanovich, P. J. (2003). *Using research and reason in education: How teachers can use scientifically based research to make curricular & instructional decisions*. Partnership for Reading (Project), National Institute for Literacy, US ….

Sterne, J. A., Becker, B. J., & Egger, M. (2005). The funnel plot. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 73–98.

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99–110.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., & Schmid, C. H. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, *343*.

Stuart, E. A., Bradshaw, C. P., & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, *16*(3), 475–485.

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). Modelling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research*, *9*(5), 421–445.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113–2126.

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *Journal of Educational and Behavioral Statistics*, *39*(6), 478–501.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375.

Tipton, E., & Matlen, B. J. (2019). Improved generalizability through improved recruitment: Lessons learned from a large-scale randomized trial. *American Journal of Evaluation*, *40*(3), 414–430.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634.

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, *36*(2), 299–316.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Whitehurst, G. J. (2008). Rigor and Relevance Redux: Director's Biennial Report to Congress. IES 2009-6010. *Institute of Education Sciences*.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.

Williams, D., & Coles, L. (2007). Teachers' approaches to finding and using research evidence: An information literacy perspective. *Educational Research*, *49*(2), 185–206.

Yoshizawa, L. (2020). *Research use under pressure: State and district implementation of the ESSA evidence requirements* [PhD Thesis]. Harvard University.

Zhao, Y. (2020). Another education war? The coming debates over social and emotional learning. *Phi Delta Kappan*, *101*(8), 42–48.

Zwetsloot, P.-P., Van Der Naald, M., Sena, E. S., Howells, D. W., IntHout, J., De Groot, J. A., Chamuleau, S. A., MacLeod, M. R., & Wever, K. E. (2017). Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife*, *6*, e24260.

Chapter 2: The Use of Research in K-12 Decision-Making

Claire Chuter

Johns Hopkins University

School of Education

Abstract

The use of research evidence by education decision-makers has the potential to markedly improve student outcomes. However, without an understanding of how, why research is used, and what barriers prevent its use, those seeking to broker knowledge to policymakers and practitioners may misdirect their efforts. This systematic review seeks to uncover the way research is used and understood in K-12 education decision-making, and what education leaders and policymakers perceive to be the main barriers to its use. Our findings may be used to inform and direct current knowledge mobilization efforts to encourage use of evidence in education decision-making.

Introduction

The use of research evidence (URE) refers to the incorporation of research evidence to make decisions, think about problems and potential solutions, and justify the resolution of problems (Weiss, 1979). URE can and should be closely integrated into K-12 decision-making at the school, district, and statewide level. Yet inherent in improving the use of research is the understanding of the users' concerns, constraints, and motivations as they form policy. Below are two central reasons to closely integrate evidence into education practice and policy decision-making.

First, URE in education produces better student outcomes, and protects students from popular yet ineffective programs. The case of zero-tolerance versus restorative justice policies illustrates this point well. Zero-tolerance policies refer to school discipline that mandates predetermined consequences - typically severe and exclusionary - in response to specific types of student misbehavior, regardless of the context or rationale for the behavior (Casella, 2003). Despite the popularity of zero-tolerance policies, the evidence base supporting such discipline is lacking (Martinez, 2009). Students returning to school following suspension due to zero-tolerance policies tend to display the same or more severe behaviors than prior to the suspension (Cartledge et al., 2001; Christie et al., 2004; Ryan & Goodram, 2013). On the other hand, restorative justice policies have a much stronger evidence base across valued student outcomes. These include reduction of suspensions (Darling-Hammond et al., 2020; Gonzalez, 2015), and absenteeism (Latimer et al., 2005; Schiff, 2013) as well as improvements in school climate (Augustine et al., 2018) and academic achievement (Miller-Jones & Rubin, 2020; Gonzalez, 2012). While zero-tolerance policies are attractive to policymakers and practitioners, they in fact do not align with the evidence. The idea of cracking down on violent offences in schools may

seem enticing on an intuitive level, but by following the evidence school leaders may be better able to serve their students.

A second important reason to use evidence in education decision-making is purely economic. Investments in education are expensive. The federal government spends just under $55 billion per year on K-12 education, in addition to outlays for early childhood education and post-secondary programs like loans and grants for college tuition (Hanson, 2022). In 2019, state and local governments spent $717 billion on elementary and secondary education (Cornman et al., 2021). Given this enormous investment in education it behooves education leaders to choose the programs and policies that show the strongest evidence of effectiveness. Otherwise, such decisions inevitably result in massive waste of financial resources.

The Need for a Systematic Review of Use of Research Evidence

Many funders (e.g., William T. Grant Foundation) and prominent scholars (Cooper & Levin, 2010; Lingard, 2013; Gorard et al., 2020) in the field of research use have emphasized the need for more empirical work on third parties in the knowledge mobilization process (e.g., Davies & Nutley, 2008; Honig, 2004; Nutley et al., 2007; Levin, 2004, 2008). The number of educational intermediaries has been increasing in the past twenty years; in the U.S. alone, intermediary organizations have quadrupled from fewer than 70 in the 1970's to at least 300 by the year 2000 (Cooper, 2013). As these intermediary organizations become more involved in mediating research use in education, understanding this process – and what works to improve research use – is critical.

In this review we used Rowan's (2001) definition of intermediary organization (sometimes referred to as research brokering organizations). Rowan distinguishes three classes of intermediary organizations that play a role in education: (1) for-profit firms (e.g., textbook

publishers, instructional program vendors); (2) membership organizations (e.g., professional associations, advocacy groups, and networks such as the Council of State Science Supervisors), (3) nonprofit organizations (e.g., including universities, research firms, quasi-governmental agencies, and philanthropies such as the Gates Foundation).

In the past, lack of knowledge around how to best motivate research use has led to ineffective knowledge mobilization strategies (Deville & Chalhoub-Deville, 2011). As one example, during the No Child Left Behind (NCLB) era, the federal government passed down harsh measures to mandate use of research in programmatic decision-making and school reform plans (Darling-Hammond, 2007; Deville & Chalhoub-Deville, 2011). This strategy led to broad frustration among school, district, and state leaders (Fusarelli, 2007; Fast & Erpenbach, 2004;). Ultimately, the punitive measures of NCLB were discarded and replaced due to the strong reaction from education practitioners and policymakers (Johnson, 2016; McGuinn, 2016). To improve the use of research in education, and to avoid repeating the same mistakes, it is critical to understand what education leaders think about using research, where they may be skeptical and what barriers they face to using it.

This systematic review will examine the way research is used in K-12 education decision-making, and what education leaders perceive to be the main barriers to its use. This review will be useful in informing current efforts being made to encourage research use in education. The research questions addressed in this systematic review are as follows:

1. *How, and at what point in the decision-making process, is evidence most used by K-12 school, district, state, and federal leadership?*
2. *What are the main perceived barriers to the use of research in decision-making by K-12 school, district, state, and federal leadership?*

3. *What efforts are being made in education address the research-practice gap, and which are successful?*

Methods

Search and Screening Process

Figure 1 illustrates the process used to comprehensively search the published and unpublished available research literature. The researcher will use the Sheridan Library's Education homepage, EBSCO, JSTOR, and ProQuest digital libraries as many databases can be combined in one search. All search, screening, and data extraction were preregistered prior to data collection and are publicly available on the Open Science Framework website (Chuter et al., 2022). Search terms were identified by reviewing the ERIC thesaurus, and consultation with subject matter experts (Table 1). Boolean operators will be used to provide more focused and productive results. Quotation marks will be used around phrases if the database accepts this and truncation symbols (*) will be used for applicable terms where appropriate (e.g., "Decision-mak*"). Database searches will be limited to title, abstract, and subject heading fields. To determine whether new research is being released from journals and organizations, the researcher signed up for alerts through Sheridan library.

Table 1

Search Terms

Category	Search Terms
Context	K-12 "Elementary school" "Primary school" "Secondary school" School "School district" Education
Target Population	"School board" "State leaders" "Education leaders" "Education administrators" "Superintendent" "Principal"
Measures	Interview Survey Document* Self-report Observ* Qualitative Quantitative Mixed methods

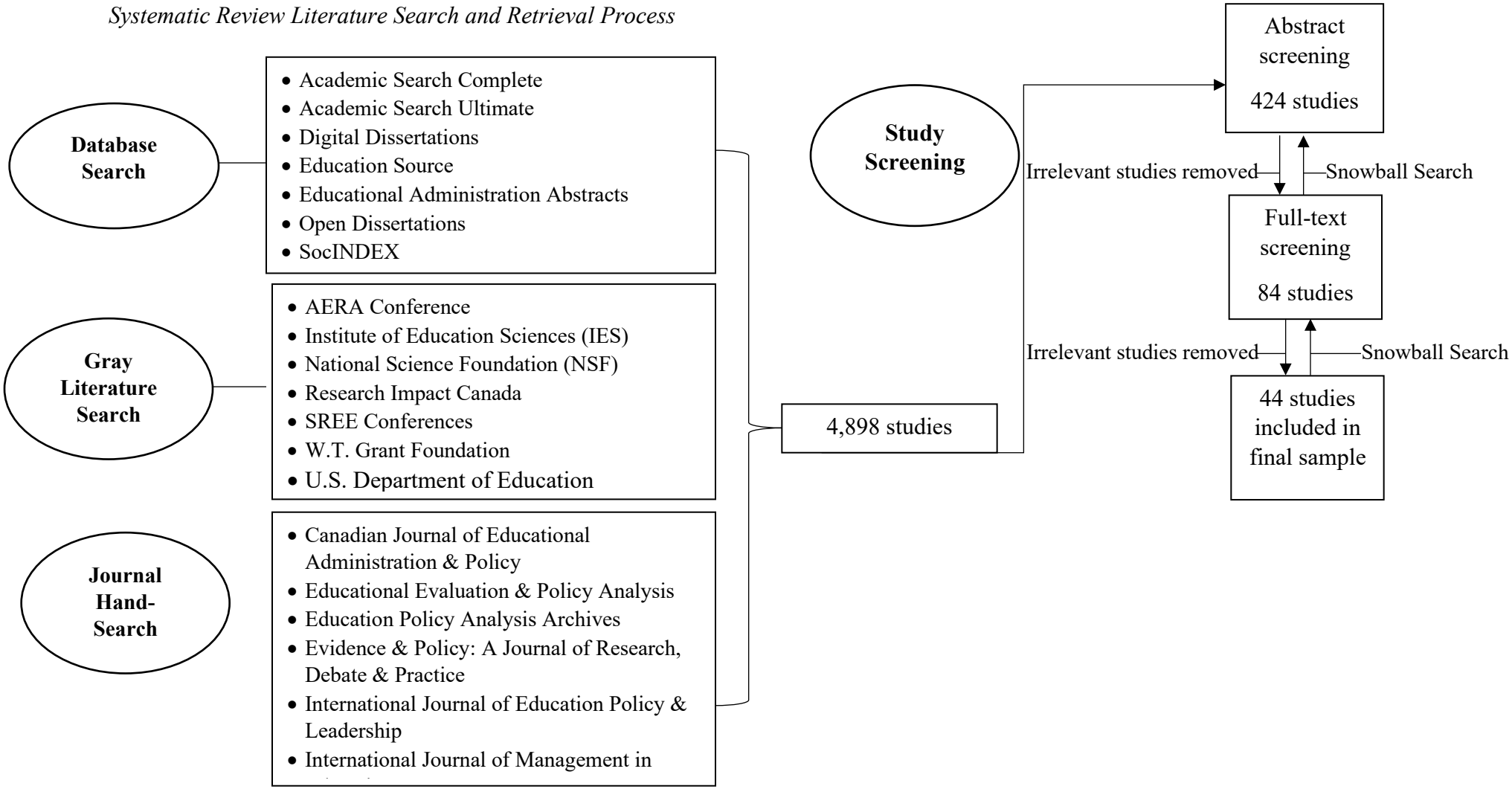
We also conducted a search of the for gray literature (Figure 1). In addition, articles from the journals that publish high volumes of URE literature were 'hand-searched' from 2000 to the present using the Paperfetcher tool (Pallath & Zhang, 2022; Figure 1).

Snowball searches of included studies were also conducted. For the forward snowball search the researcher examined reference lists of literature reviews, meta-analyses, book chapters, and included papers found in the initial data collection. These studies underwent the same two-stage screening process as before. For the backwards snowball search, studies that have since cited the included studies were collected and also underwent the two-stage screening. This cycle continued until saturation was reached and no new studies were found.

Studies gathered during the search process underwent two stages of screening. Each study was screened by two independent researchers at the title and abstract level against eligibility criteria usually found in the title or abstract. These are the following: (1) takes place outside of the established time frame (>2000), (2) does not describe URE; (3) takes place outside of the included countries (i.e., United Kingdom, European Union, Australia, New Zealand, Canada, or the United States); (4) does not examine K-12 education policymaking contexts; or (5) are not in English. Any conflicts between reviewers were discussed until consensus was reached. Studies that were not excluded during the abstract and title screening stage, were reviewed for their full text by two independent reviewers. At this stage, studies were excluded if they failed any of the previous criteria, or (1) failed to address the research questions or (2) were not empirical. Again, any conflicts were discussed until consensus was reached. (A full list of included studies can be found in Appendix B).

Figure 1

Systematic Review Literature Search and Retrieval Process



Data Extraction

Literature was uploaded into Atlas.io (a qualitative coding software package) and coded using a priori and open coding techniques. Data was then extracted by two independent reviewers through a priori codes developed from the research questions. These included the following covariate codes: (1) type of study (e.g., mixed methods); (2) form of the data (e.g., document analysis of meeting notes); (3) context (e.g., secondary school leadership); (4) country; (5) peer-reviewed; and (6) year. To examine the research questions, we used the following substantive codes: (1) definition/s of research evidence by education leaders; (2) interpretation processes of research evidence by education leaders; (3) forms of research used (e.g., instrumental, political, conceptual); (4) timeline of decision-making process; (5) point at which evidence is used; and (6) barriers to URE in K-12 decision-making contexts. The extracted data was then compared for inconsistencies and discussed until the reviewers come to consensus about any differences.

Open coding was used to capture emergent concepts from the literature that were not captured by a priori codes. Again, two reviewers undertook the process of open coding and theming. In this process, the reviewers independently assigned codes that described the content of the extracted dimensions. Reviewers solicited feedback from each other to refine and re-apply codes, reiterating this process until they reached consensus.

Results

All 44 articles were published between 2000 and 2020, with the most prolific period taking place over the last 5 years (n = 17). Six countries were represented, with articles originating from the United States (n = 29), Canada (n = 13), Australia (n = 1), England (n = 1),

Wales (*n* = 1), and New Zealand (*n* = 1)[1]. The majority of articles used qualitative research designs (*n* = 25), with fewer articles using mixed methods (*n* = 16) and quantitative designs (*n* = 3). All quantitative data included in this review was descriptive, with no study using experimental data. Most studies included a target population of school leaders (*n* = 23), followed by knowledge brokers (*n* = 18), district leaders (*n* = 14), and policymakers (*n* = 9). Finally, most articles were in peer-reviewed journal form (n = 30), with several dissertations (n =7), technical reports (n =4), and conference papers (n =2). The results of this review are structured first by research question and second by content theme (Table 2). Gaps in the literature prevented us from addressing some parts of the research questions (e.g., "at what point in the decision-making process most useful").

Table 2

Content Map

Research Question	**Theme**	**No. of Studies**
How, and at what point in the decision-making process, is evidence most used by K-12 school, district, state, and federal leadership?	Decisions Commonly Based on Intuition and Personal References	16
	High Interest in Context-Dependent Knowledge	8
What are the main perceived barriers to the use of research in decision-making by K-12 school, district, state, and federal leadership?	Reported Barriers to Research Use Include Irrelevance to Agenda and Timing	13
	Low Trust of Intermediary Organizations	4
What efforts are being made in education address the research-practice gap, and which are successful?	Active – Not Passive - Research Dissemination Supports Research Use	10
	Long-Term Exposure to Research is Important for Use	4

[1] Some studies represented more than country and/or target population.

Decisions Commonly Based on Intuition and Personal References

Kirst (2000) found that policymakers primarily used personal anecdotes to justify their position and explain the success of reforms. Peer-reviewed research was seldom used, but when it was used, it was passed through personal contacts or key partners. Jabbar et al., (2014) also found that state-level policymakers rely heavily on their personal and professional networks. For example, State Superintendent John White noted when consulting researchers, he called up those whom he had established relationships with for different policy issues.

> "I'll call Kati Haycock [of Education Trust] a lot and ask her for what's the best study on this. Tim Daly [of The New Teacher Project], Anne Weisberg [of the Families and Work Institute], Andy Rotherham [of Eduwonk and Bellwether Education Partners], Tom Kane [of Harvard University and the Gates Foundation] – those are people I call a lot and ask for particular studies, and they tend to be able to point me toward the right stop." (Jabbar et al., 2014, pp. 1021)

When choosing educational technology, Morrison et al., (2019) also found that district leaders most often consulted peer references.

High Interest in Context-Dependent Knowledge

Education leaders and policymakers place high value on student data relevant to their own particular school or district. In a survey of school principals, Cooper (2013) found that respondents consulted internal data such as high school graduation rates, elementary school literacy levels, and achievement data by ethnicity and socioeconomic status regularly, while peer-reviewed journal articles were used rarely. School leaders interviewed by Honig (2004) report hesitancy around using research presented by intermediary organizations because these

groups are out of touch with the specific needs of their context. The following example is an interchange between an intermediary administrator and the school principal:

> Administrator: [arguing that particular sites can weather a recent central office decision not to provide promised funding]: I think depending on your vision and what you are doing costs can vary greatly. I know some places would die for your budgets. . . . I'm saying it varies widely....
>
> School Principal: You have never even been to my site. You don't even know our budgets. You make us crazy every time you talk like this. I don't care what my total budget looks like to you. If I can't make my payroll because I may not receive the check you promised this is going to hit the fan. We need to follow up on this. This is really contentious. (p. 77)

School administrators interviewed by Yoshizawa (2020) indicated that they would put their greatest confidence in locally generated evidence, even when that involved temporarily putting aside ESSA's expectations for rigorous design. Their reasons for this were varied: administrators vouched for the local judgment and expertise of their district and school leaders; they perceived the current research base to have limited generalizability for particular demographics and contexts; and for some, reviewing internal data was an established part of their routines. Jabbar et al., (2014) also found that state-level policymakers relied most heavily on internally produced data conducted in their district, then independent research provided by outside sources.

Main Barriers to Research Use: Irrelevance to Agenda and Timing

While there is a wide range of perceived barriers to research use, the most commonly reported issues across education leadership and policymakers were: an irrelevance of the research to their own goals, a misalignment of timing, or of personal values. Personal values were mainly a concern of policymakers as opposed to education leadership (Campbell et al., 2017).

Misalignment of Goals and Values

State and local policymakers interviewed by Jabbar et al., (2014) reported that traditional academic research did not address the questions they needed answered. University-based knowledge brokers who were successful in research dissemination chose research that were important to policymakers (Ackman, 2013). One participant commented on his brokering style,

> I'm picking topics that I think relate to real problems. I know in the end I'm going to end up with a journal article version of this thing. I know that's also going to be relevant and so I'm going to write a policy brief for an outline or something that discusses the findings. (p. 98)

Galway and Sheppard (2015) found that decision makers at the school district level attend less to political and pragmatic influences relying more on personal beliefs, values and experiential factors - supplemented by the advice of staff and in-house research/indicators. Hopkins et al 2019 found that research-based ideas that were most likely to cross the research-practice barrier when they had shared values between both sides.

Timing

Policymakers interviewed by Campbell and Fulford (2009) advised attention to policymaker timelines was advised. Scott et al. (2017) also found that knowledge mobilization is more likely to be effective if broached to stakeholders at opportunistic times, with a straightforward, clear message. Experienced knowledge brokers argued that keeping tabs on dominant discourse at an institution and sudden shifts in prioritized objectives was key to their influence.

Low Trust of Intermediary Organizations

Policymakers communicated their widespread belief that research provided by intermediary organizations was not trustworthy (Jabbar et al., 2014). Trust in research was highly dependent on the people they know within that organization. For example, one policymaker explained, "I trust the folks at Cowen. I think what they do is good. Picard is sort of questionable. The folks at SEDL are also sort of questionable" (Jabbar et al., 2014, pp. 1019). However, there was little knowledge around *why* certain individuals were more trustworthy than others in the eyes of policymakers. Canfield-Davis (2010) also discovered a low level of trust in intermediary organizations at the policy level. As noted by one policymaker,

> Lobbyists are hired guns. They have a single point of view, and they don't give a damn about anybody else's point of view, the contrary point of view. When they slide around and try to tell you that they are, 'really trying to do the right thing here and this is the line that you should believe, because it's the right thing,' it's bullshit. (p. 614)

Honig (2004) interviewed education leaders and found similar results. Multiple respondents believed that intermediary organizations handpicked the schools they worked with to showcase the best results, without an interest in genuinely strengthening the schools.

Active – Not Passive - Research Dissemination Supports Research Use

Passive research dissemination here is defined as the publication or provision of access to research. Active research dissemination refers to a more hands-on approach – involving regular contact and support between knowledge brokers and users. Across the papers in this review, passive research dissemination was found to be ineffective, while active dissemination was effective at encouraging research use (Canfield-Davis & Jain, 2010; Ackman, 2013; Cooper, 2013).

Positive and trusting relationships, sustained over long periods of time were reported to be associated with success in research brokering efforts to both education policymakers and leadership. Canfield-Davis and Jain (2010) found that trust forms the foundation upon which other factors depend including bill sponsors, party leadership, lobbyists, fellow legislators, and constituents. Ackman (2013) reported that in order to work with politicians to help incorporate scholarly research into the policymaking process, participants actively built relationships with both politicians and their staffs. One knowledge broker interviewed by Cooper (2013) explained of a sample of Canadian school district leadership, "Capacity is not synonymous with actual use… [D]istricts appear to have relatively weak processes and systems for using relevant research" (p. 15).

Similarly, Honig (2009) found that implementation of research in decisions, successes, and pitfalls came through more active engagement such as phone calls, site visits, and other opportunities for direct, ongoing interactions. Interestingly, policymakers interviewed by, Brisco et al., (2015) also found that information was used more often if it was voiced, in conversation, rather than written.

Perspective-Taking is a Key Component of Active Dissemination

In addition, participants reported carefully curating a research agenda according to the current political climate and the policymaker's goals. Canfield-Davis & Jain, (2010) found that engaging in political activities is a necessary prerequisite for knowledge brokers to be successful. Harrison et al., (2017) found that interpersonal relationships and social contexts are key to shaping evidence use in practice, establishing ongoing personal relationships, which support the use of research, was seen as one of the most successful strategies in knowledge transfer.

Long-Term Exposure to Research is Important for Use

Kirst (2000) found that decision makers process recommendations that emerge from research reports and conversation over long periods of time. Similarly, Awad (2015) found that over time government officials who feel a sense of ownership of research over time are more likely to use the research in policy. University knowledge brokers, interviewed by Ackman (2013) argued that long-term exposure to research ensure that important information is available and understood when the appropriate legislative cycle comes. One knowledge broker commented:

> You never know when [research] is going to be used and sometimes it is stuff that I have written a really long time ago…there is a lag sometimes between research and policy and you just have to understand that when you do this kind of work. (Ackman, 2013, p.86)

Discussion

This qualitative review sheds light on the mechanisms in which knowledge is mobilized or not, in K-12 decision-making contexts. A broad search of the published and unpublished

literature lead to the identification of six conclusions surrounding the field of research use in education. First, that decisions at the school, district, state, and federal level rely most often on intuition and personal references. Second, that contextual knowledge is valued much higher than the strength of the research design. Third, that when asked about barriers to research use, policymakers and practitioners most commonly cite an irrelevance of presented research to their agenda, and/or a misalignment between the timing of when decisions need to be made, and when the research becomes available. Fourth, beyond these *reported* barriers to research use, we discovered a widespread lack of trust of intermediary organizations among research users. Fifth, we found that active knowledge mobilization, such as relationship building, and continuous support are far more effective than passive dissemination. Finally, long-term exposure to research allows for increased conceptual understanding on the research user's side as well as being logistically supportive given the unpredictable priorities of each legislative cycle.

This research strongly indicates the value of the social nature of research use. Others have called attention to this factor in the past (Levin, 2013; Daly, 2012; Cooper, 2012). For a long time, and in some quarters still today, the application of research was thought of as a straightforward process in which researchers would discover knowledge which would then be unproblematically adopted into policy and practice. Although that view has long been dismissed by those who have studied the matter (Nutley et al., 2003; Bhattacharyya et al., 2009), it remains quite strongly embedded in the ongoing practice of those attempting to mobilize research knowledge. This body of work reaffirms the conclusion that increased knowledge is not enough to improve practice in education. This review uncovered one particularly compelling factor of research use: trust. Trust in colleagues, knowledge intermediaries, and academics, that the research will deliver the expected outcomes regularly determines policy decisions. Psychology

has long understood the value of trust in building strong interpersonal relationships (Simpson, 2007; Rempel et al., 1985). The field of research use should put these lessons to use in shifting away from an instrumental view of the research-practice pipeline, and towards a trusting, communicative one.

Limitations

The literature in the field of research use is handicapped by its paucity of experimental research. In answering each of these questions, we relied on stakeholder's perception, collected through surveys or interviews. However, individuals' assessment of their own priorities and experiences may be flawed. Individuals tend to view their own group in a more positive light and remember their own actions as matching up more closely with their idealized version of themselves than reality (Molenberghs, 2013; Harmon-Jones & Mills, 2019). The introduction of experimental research would advance the field of research use in education and provide knowledge brokers with greater confidence in their findings.

Conclusion

This review uncovered an unexpected paradox within the field of research use. That is, on the one hand, policymakers and education administrators understand the competing interests of intermediary organizations and are thus apprehensive to put their faith in the information they receive from these organizations. These knowledge users may be afraid of wasting valuable time and resources by using faulty research, and therefore rely instead on information based on their experience, colleague's advice, or their own intuition. However, by not following evidence, policymakers and practitioners are *more* likely to waste time and resources.

One common thread throughout this review is the importance of trust between knowledge brokers and users in the successful transfer of research. This understanding opens two avenues

for future research. First, the need to better understand what characteristics of knowledge brokers tend to encourage trust. And second, what identifiable characteristics distinguish trustworthy intermediary organizations from untrustworthy organizations to research users? These two areas of research are critical in developing and maintaining strong conduits of knowledge mobilization in K-12 education.

References

Ackman, E. R. (2013). Getting scholarship into policy: Lessons from university-based bipartisan scholarship brokers. Arizona State University

Amara, N., Ouimet, M., & Landry, R. (2004). New evidence on instrumental, conceptual, and symbolic utilization of university research in government agencies. Science communication, 26(1), 75-106.

Asmonga, J. A. (2019). The Space Between: Using Inquiry Teams to Encourage Multilevel Boundary Crossing Between Educational Research and Practice (Doctoral dissertation, The Johns Hopkins University).

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., & Joseph, A. A. (2018). Can restorative practices improve school climate and curb suspensions. An evaluation of the impact of restorative practices in a mid-sized urban school district, 1-112.

Awad, N. C. (2015). Knowledge transfer intervention theory: a model grounded in the strategies used by intermediate agents in the context of education.

Bhattacharyya, O., Reeves, S., & Zwarenstein, M. (2009). What is implementation research? Rationale, concepts, and practices. Research on Social Work Practice, 19(5), 491-502.

Briscoe, P., Pollock, K. E., Campbell, C., & Carr-Harris, S. (2015). Finding the sweet spot: Network structures and processes for increased KMb. Brock Education Journal, 25(1), 19.

Campbell, C., & Fulford, D. (2009, April). From knowledge generation to knowledge integration: Analysis of how a government uses research. In American Educational Research Association Annual Meeting, San Diego, CA. Retrieved from http://www. edu. gov. on. ca/eng/research/aera2009_kipaper. Pdf.

Campbell, C., Pollock, K., Briscoe, P., Carr-Harris, S., & Tuters, S. (2017). Developing a knowledge network for applied education research to mobilise evidence in and for educational practice. Educational Research, 59(2), 209-227.

Canfield-Davis, K., & Jain, S. (2010). Legislative Decision-Making on Education Issues: A Qualitative Study. Qualitative Report, 15(3), 600-629.

Cartledge, G., Tillman, L. C., & Talbert Johnson, C. (2001). Professional ethics within the context of student discipline and diversity. *Teacher Education and Special Education*, *24*(1), 25-37.

Casella, R. (2003). Zero tolerance policy in schools: Rationale, consequences, and alternatives. *Teachers College Record*, *105*(5), 872-892.

Christie, C. A., Nelson, C. M., & Jolivette, K. (2004). School characteristics related to the use of suspension. Education and treatment of Children, 509-526.

Chuter, C., & Neitzel, A. J. (2022, April 7). Research Use and Knowledge Mobilization in K-12 Decision-Making. Retrieved from osf.io/bkwam

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. American journal of Education, 112(4), 469-495.

Coburn, C. E., Spillane, J. P., Bohannon, A. X., Allen, A. R., Ceperich, R., Beneke, A., & Wong, L. S. (2020). The Role of Organizational Routines in Research Use in Four Large Urban School Districts. Technical Report No. 5. National Center for Research in Policy and Practice.

Cornman, S. Q., Ampadu, O., Hanak, K., Howell, M., & Wheeler, S. (2021). Revenues and Expenditures for Public Elementary and Secondary School Districts: FY 19. Finance Tables. NCES 2021-304. National Center for Education Statistics.

Cooper, A., Edelstein, H., Levin, B., & Leung, J. (2010). Use of web-based research materials in education: Is uptake occurring. In *Canadian Society for the Study of Education (CSSE) Conference, Montreal, Canada. Retrieved from http://www. oise. utoronto. ca/rspe/UserFiles/File/CSSEWebtrackerFinal. doc*.

Cooper, A. (2014). The use of online strategies and social media for research dissemination in education. Education Policy Analysis Archives, 22, 1-24.

Cooper, A., & Levin, B. (2013). Research Use by Leaders in Canadian School Districts. International Journal of Education Policy and Leadership, 8(7), n7.

Cooper, A. (2015). A tool to assess and compare knowledge mobilization efforts of faculties of education, research brokering organizations, ministries of education, and school districts. Brock Education Journal, 25(1).

Daly, A. J. (2012). Data, dyads, and dynamics: Exploring data use and social networks in educational improvement. Teachers College Record, 114(11), 1-38.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. Race ethnicity and education, 10(3), 245-260.

Darling-Hammond, S., Fronius, T. A., Sutherland, H., Guckenburg, S., Petrosino, A., & Hurley, N. (2020). Effectiveness of restorative justice in US K-12 schools: A review of quantitative research. Contemporary School Psychology, 24(3), 295-308.

Deville, C., & Chalhoub-Deville, M. (2011). Accountability-assessment under no child left behind: Agenda, practice, and future. Language Testing, 28(3), 307-321.

Farley-Ripple, E., & Jones, A. R. (2015). Educational Contracting and the Translation of Research into Practice: The Case of Data Coach Vendors in Delaware. International Journal of Education Policy and Leadership, 10(2), n2.

Farrell, C. C., Davidson, K. L., Repko-Erwin, M., Penuel, W. R., Quantz, M., Wong, H., Reidy, R., & Brink, Z. (2018). A Descriptive Study of the IES Researcher-Practitioner Partnerships in Education Research Program: Final Report. Technical Report No. 3. National Center for Research in Policy and Practice.

Fast, E. F., & Erpenbach, W. J. (2004). Revisiting statewide accountability under NCLB: A summary of state requests in 2003-2004 for amendments to state accountability plans. Washington, DC: Council of Chief State School Officers.

Fusarelli, L. D. (2007). Restricted choices, limited options: Implementing choice and supplemental educational services in No Child Left Behind. Educational Policy, 21(1), 132-154.

Galway, G., & Sheppard, B. (2015). Research and Evidence in Education Decision-Making: A Comparison of Results from Two Pan-Canadian Studies. education policy analysis archives, 23(109), n109.

González, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. JL & Educ., 41, 281.voight 2013

González, T. (2015). Socializing schools: Addressing racial disparities in discipline through restorative justice. Thalia González, Socializing Schools: Addressing Racial Disparities in Discipline Through Restorative Justice, in Closing the School Discipline Gap: Equitable Remedies for Excessive Exclusion (Daniel J. Losen ed., 2014).

Gorard, S., See, B. H., & Siddiqui, N. (2020). What is the evidence on the best way to get evidence into use in education?. *Review of Education*, *8*(2), 570-610.

Hanson, Melanie. "U.S. Public Education Spending Statistics" EducationData.org, March 15, 2022, https://educationdata.org/public-education-spending-statistics.

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory.

Harrison, C., Davidson, K., & Farrell, C. (2017). Building Productive Relationships: District Leaders' Advice to Researchers. International Journal of Education Policy and Leadership, 12(4), n4.

Hedges, S. (2020). Intermediary Organizations, KMb, and Education Policy: Cross-National Comparative Case Studies of How Political Environments Shape Evidence Mobilization (Doctoral dissertation, Indiana University).

Hemsley-Brown, J. (2005). Using research to support management decision making within the field of education. Management Decision.

Hollands, F., Yan, B., Leach, S., Shand, R., Dossett, D., Chang, F., Pan, Y., Wang, Y., & Head, L. (2019). SREE 2020 Abstract October 2019.

Hollar, J. (2018). The Inequitable Access of Knowledge: The Use of Federally Funded Intermediary Organizations as Knowledge Brokers (Doctoral dissertation).

Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. Educational Evaluation and Policy Analysis, 26(1), 65-87.

Honig, M. I. (2009). "External" organizations and the politics of urban educational leadership: The case of new small autonomous schools initiatives. Peabody Journal of Education, 84(3), 394-413.

Hopkins, M., Wiley, K. E., Penuel, W. R., & Farrell, C. C. (2018). Brokering research in science education policy implementation: the case of a professional association. Evidence & Policy: A Journal of Research, Debate and Practice, 14(3), 459-476.

Hopkins, M., Weddle, H., Gluckman, M., & Gautsch, L. (2019). Boundary crossing in a professional association: The dynamics of research use among state leaders and researchers in a research-practice partnership. AERA Open, 5(4), 2332858419891964.

Hopkins, M. (2016). Findings from a Survey of State Science Leaders. National Center for Research in Policy and Practice.

Jabbar, H., La Londe, P. G., Debray, E., Scott, J., & Lubienski, C. (2014). How Policymakers Define 'Evidence': the politics of research use in New Orleans. Policy Futures in Education, 12(8), 1013-1027.

Johnson, M. (2016). The Every Student Succeeds Act: Opportunities and Responsibilities. re: VISION. Hunt Institute.

Kaplan, R. G., Riedy, R., Van Horne, K., & Penuel, W. (2019). Going on a statewide listening tour: involving education leaders in the process of research to enhance the practical value of qualitative research. Evidence & Policy, 15(2), 179-196.

Kirst, M. W. (2000). Bridging education research and education policymaking. Oxford Review of Education, 26(3-4), 379-391.

Kochanek, J. R., Scholz, C., Monahan, B., & Pardo, M. (2020). An Exploratory Study of how to Use RPPs to Build Trust and Support the Use of Early Warning Systems. Teachers College Record, 122(14), 1-28.

Latimer, J., Dowden, C., & Muise, D. (2005). The effectiveness of restorative justice practices: A meta-analysis. The prison journal, 85(2), 127-144.mcmorris 2013

Lazaridou, A. (2009). The Kinds of Knowledge Principals Use: Implications for Training. International journal of education policy and leadership, 4(10), 1-15.

Laitsch, D., & Younghusband, C. H. (2019). British Columbia School Trustees' Use of Research and Information Seeking in Decision Making. Canadian Journal of Educational Administration and Policy, (188).

Levin, B. (2013). To know is not enough: Research knowledge and its use. Review of education, 1(1), 2-31.

Lingard, B. (2013). The impact of research on education policy in an era of evidence-based policy. *Critical Studies in Education*, *54*(2), 113-131.

MacGregor, S., & Phipps, D. (2020). How a Networked Approach to Building Capacity in KMb Supports Research Impact. International Journal of Education Policy and Leadership, 16(6).

Malik, S. (2020). KMb for Impact: A Multi-Case Study of Education Organizations. International Journal of Education Policy and Leadership, 16(7).

Malin, J. R., & Paralkar, V. K. (2017). Educational Knowledge Brokerage and Mobilization: The" Marshall Memo" Case. International Journal of Education Policy and Leadership, 12(7), n7.

Martinez, S. (2009). A system gone berserk: How are zero-tolerance policies really affecting schools?. *Preventing school failure: alternative education for children and youth*, *53*(3), 153-158.

McGuinn, P. (2016). From no child left behind to the every student succeeds act: Federalism and the education legacy of the Obama administration. Publius: The Journal of Federalism, 46(3), 392-415.

Miller-Jones, D., & Rubin, M. M. (2020). Achieving Equity in Education: A Restorative Justice Approach. Journal of Public Management & Social Policy, 27(1), 22.

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience & Biobehavioral Reviews*, *37*(8), 1530-1536.

Morrison, J. R., Ross, S. M., & Cheung, A. C. (2019). From the market to the classroom: How ed-tech products are procured by school districts interacting with vendors. Educational Technology Research and Development, 67, 389-421.

Naff, D. B. (2020). Podcasting as a Dissemination Method for a Researcher-Practitioner Partnership. International Journal of Education Policy and Leadership, 16(13), n13.Nelson, S. R., Leffler, J. C., & Hansen, B. A. (2009). Toward a research agenda for understanding and improving the use of research evidence. Northwest Regional Educational Laboratory (NWREL).

Nutley, S., Walter, I., & Davies, H. T. (2003). From knowing to doing: a framework for understanding the evidence-into-practice agenda. Evaluation, 9(2), 125-148.

Pallath, A., & Zhang, Q. (2021). PaperFetcher: A tool to automate handsearch for systematic reviews. *arXiv preprint arXiv:2110.12490*.

Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A. R. (2016). Findings from a National Study on Research Use among School and District Leaders. Technical Report No. 1. National Center for Research in Policy and Practice.

Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A. R. (2017). How school and district leaders access, perceive, and use research. AERA Open, 3(2), 2332858417705370.

Prendergast, S., & Rickinson, M. (2019). Understanding school engagement in and with research. The Australian Educational Researcher, 46, 17-39.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. Journal of personality and social psychology, 49(1), 95.

Ryan, T. G., & Goodram, B. (2013). The Impact of Exclusionary Discipline on Students. International Journal of Progressive Education, 9(3), 2014

Schiff, M. (2013). Dignity, Disparity and Desistance: Effective Restorative Justice Strategies to Plug the" School-to-Prison Pipeline".

Simpson, J. A. (2007). Foundations of interpersonal trust. Social psychology: Handbook of basic principles, 2, 587-607.

Weiss, C. H. (1979). The many meanings of research utilization. Public administration review, 39(5), 426-431.

Wentworth, L., Mazzeo, C., & Connolly, F. (2017). Research practice partnerships: A strategy for promoting evidence-based decision-making in education. Educational Research, 59(2), 241-255.

Witherow, K. (2011). Research use and the impact in secondary schools. University of Toronto.

Yoshizawa, L. (2020). Research use under pressure: State and district implementation of the ESSA evidence requirements (Doctoral dissertation, Harvard University).

Appendix A

Table 3

Inclusion Criteria

Topic – The included studies must focus on decision-making and/or research use in K-12 education contexts. In addition, the study must address one of the following research questions:

1. What forms of research and at what point in the decision-making process, is evidence most useful to practitioners and policymakers?
2. What are the main perceived barriers to the use of research by K-12 school, district, and state-wide leaders?
3. What efforts are being made in education to address the research-practice gap?

Location – The primary study must be in a K-12 decision-making education context. To ensure that replicability of findings to U.S. education contexts, studies must take place in the United Kingdom, European Union, Australia, New Zealand, Canada, or the United States.

Study Design – Studies must be empirical in nature. More specifically, studies can be qualitative, quantitative, or mixed-methods designs used to measure decision-makers (e.g., teachers and school board members) opinions of research use, as well as their actual use. Methodological approaches may include self-report surveys, interviews, observation, and document analysis. Measurement studies (i.e., studies to develop and validate instruments for assessing research use) will also be included.

Time frame – The study must use data collected in 2000 or later.

Language – The study must be available in English.

Appendix B

Table 4

Included Articles

Reference	Country	Research Design	Form of Report	Target Population
Ackman, E. R. (2013). *Getting scholarship into policy: Lessons from university-based bipartisan scholarship brokers*. Arizona State University	USA	Qualitative	Dissertation	Knowledge brokers
Amara, N., Ouimet, M., & Landry, R. (2004). New evidence on instrumental, conceptual, and symbolic utilization of university research in government agencies. *Science communication*, *26*(1), 75-106.	Canada	Quantitative	Journal article	Policymakers
Asmonga, J. A. (2019). *The Space Between: Using Inquiry Teams to Encourage Multilevel Boundary Crossing Between Educational Research and Practice* (Doctoral dissertation, The Johns Hopkins University).	USA	Mixed methods	Dissertation	School leaders District leaders
Awad, N. C. (2015). Knowledge transfer intervention theory: a model grounded in the strategies used by intermediate agents in the context of education.	Canada	Mixed methods	Dissertation	Knowledge brokers

Briscoe, P., Pollock, K. E., Campbell, C., & Carr-Harris, S. (2015). Finding the sweet spot: Network structures and processes for increased KMb. *Brock Education Journal*, *25*(1), 19.	Canada	Qualitative	Journal article	Knowledge brokers
Campbell, C., & Fulford, D. (2009, April). From knowledge generation to knowledge integration: Analysis of how a government uses research. In *American Educational Research Association Annual Meeting, San Diego, CA. Retrieved from http://www.edu. gov. on. ca/eng/research/aera2009_kipaper. Pdf*.	USA	Qualitative	Conference paper	Policymakers
Campbell, C., Pollock, K., Briscoe, P., Carr-Harris, S., & Tuters, S. (2017). Developing a knowledge network for applied education research to mobilise evidence in and for educational practice. *Educational Research*, *59*(2), 209-227.	Canada	Qualitative	Journal article	Knowledge brokers
Canfield-Davis, K., & Jain, S. (2010). Legislative Decision-Making on Education Issues: A Qualitative Study. *Qualitative Report*, *15*(3), 600-629.	USA	Qualitative	Journal article	Policymakers
Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the	USA	Qualitative	Journal article	School leaders District leaders

terrain. *American journal of Education*, *112*(4), 469-495.				
Coburn, C. E., Spillane, J. P., Bohannon, A. X., Allen, A. R., Ceperich, R., Beneke, A., & Wong, L. S. (2020). The Role of Organizational Routines in Research Use in Four Large Urban School Districts. Technical Report No. 5. *National Center for Research in Policy and Practice*.	USA	Qualitative	Technical report	School leaders District leaders
Cooper, A. (2014). The use of online strategies and social media for research dissemination in education. *Education Policy Analysis Archives, 22*, 1-24.	Canada	Mixed methods	Journal article	Knowledge brokers
Cooper, A., & Levin, B. (2013). Research Use by Leaders in Canadian School Districts. *International Journal of Education Policy and Leadership*, *8*(7), n7.	Canada	Quantitative	Journal article	School leaders District leaders
Cooper, A. (2015). A tool to assess and compare knowledge mobilization efforts of faculties of education, research brokering organizations, ministries of education, and school districts. *Brock Education Journal*, *25*(1).	Canada	Mixed methods	Journal article	Knowledge brokers
Farley-Ripple, E., & Jones, A. R. (2015). Educational Contracting and the Translation of Research into Practice: The Case of Data Coach Vendors in Delaware. *International*	USA	Qualitative	Journal article	Knowledge brokers

Journal of Education Policy and Leadership, *10*(2), n2.				
Farrell, C. C., Davidson, K. L., Repko-Erwin, M., Penuel, W. R., Quantz, M., Wong, H., Reidy, R., & Brink, Z. (2018). A Descriptive Study of the IES Researcher-Practitioner Partnerships in Education Research Program: Final Report. Technical Report No. 3. *National Center for Research in Policy and Practice*.	USA	Mixed methods	Technical report	School leaders District leaders Knowledge brokers
Galway, G., & Sheppard, B. (2015). Research and Evidence in Education Decision-Making: A Comparison of Results from Two Pan-Canadian Studies. *education policy analysis archives*, *23*(109), n109.	Canada	Mixed methods	Journal article	School leaders District leaders Policymakers
Harrison, C., Davidson, K., & Farrell, C. (2017). Building Productive Relationships: District Leaders' Advice to Researchers. *International Journal of Education Policy and Leadership*, *12*(4), n4.	USA	Qualitative	Journal article	School leaders District leaders
Hedges, S. (2020). *Intermediary Organizations, KMb, and Education Policy: Cross-National Comparative Case Studies of How Political Environments Shape Evidence Mobilization* (Doctoral dissertation, Indiana University).	New Zealand & USA	Qualitative	Dissertation	Knowledge brokers
Hemsley-Brown, J. (2005). Using research to support management decision	England & Wales	Mixed methods	Journal article	School leaders

making within the field of education. *Management Decision*.				
Hollands, F., Yan, B., Leach, S., Shand, R., Dossett, D., Chang, F., Pan, Y., Wang, Y., & Head, L. (2019). SREE 2020 Abstract October 2019.	USA	Qualitative	Conference paper	School leaders
Hollar, J. (2018). *The Inequitable Access of Knowledge: The Use of Federally Funded Intermediary Organizations as Knowledge Brokers* (Doctoral dissertation).	USA	Qualitative	Dissertation	Knowledge brokers
Honig, M. I. (2004). The new middle management: Intermediary organizations in education policy implementation. *Educational Evaluation and Policy Analysis*, *26*(1), 65-87.	USA	Qualitative	Journal article	Knowledge brokers
Honig, M. I. (2009). "External" organizations and the politics of urban educational leadership: The case of new small autonomous school initiatives. *Peabody Journal of Education*, *84*(3), 394-413.	USA	Qualitative	Journal article	Knowledge brokers
Hopkins, M., Wiley, K. E., Penuel, W. R., & Farrell, C. C. (2018). Brokering research in science education policy implementation: the case of a professional association. *Evidence & Policy: A Journal of Research, Debate and Practice*, *14*(3), 459-476.	USA	Quantitative	Journal article	Knowledge brokers
Hopkins, M., Weddle, H., Gluckman, M., & Gautsch, L. (2019). Boundary	USA	Mixed methods	Journal article	Policymakers Knowledge brokers

crossing in a professional association: The dynamics of research use among state leaders and researchers in a research-practice partnership. *AERA Open*, *5*(4), 2332858419891964.				
Hopkins, M. (2016). Findings from a Survey of State Science Leaders. *National Center for Research in Policy and Practice*.	USA	Mixed methods	Technical report	Policymakers Knowledge brokers
Jabbar, H., La Londe, P. G., Debray, E., Scott, J., & Lubienski, C. (2014). How Policymakers Define 'Evidence': the politics of research use in New Orleans. *Policy Futures in Education*, *12*(8), 1013-1027.	USA	Qualitative	Journal article	Policymakers
Kaplan, R. G., Riedy, R., Van Horne, K., & Penuel, W. (2019). Going on a statewide listening tour: involving education leaders in the process of research to enhance the practical value of qualitative research. *Evidence & Policy*, *15*(2), 179-196.	USA	Qualitative	Journal article	School leaders District leaders
Kirst, M. W. (2000). Bridging education research and education policymaking. *Oxford Review of Education*, *26*(3-4), 379-391.	USA	Qualitative	Journal article	Policymakers
Kochanek, J. R., Scholz, C., Monahan, B., & Pardo, M. (2020). An Exploratory Study of how to Use RPPs to Build Trust and Support the Use of Early Warning Systems.	USA	Qualitative	Journal article	Knowledge brokers School leaders District leaders

Teachers College Record, *122*(14), 1-28.				
Laitsch, D., & Younghusband, C. H. (2019). British Columbia School Trustees' Use of Research and Information Seeking in Decision Making. *Canadian Journal of Educational Administration and Policy*, (188).	Canada	Qualitative	Journal article	School leaders
Lazaridou, A. (2009). The Kinds of Knowledge Principals Use: Implications for Training. *International journal of education policy and leadership*, *4*(10), 1-15.	Canada	Qualitative	Journal article	School leaders
MacGregor, S., & Phipps, D. (2020). How a Networked Approach to Building Capacity in KMb Supports Research Impact. *International Journal of Education Policy and Leadership*, *16*(6).	Canada	Mixed methods	Journal article	Knowledge brokers
Malik, S. (2020). KMb for Impact: A Multi-Case Study of Education Organizations. *International Journal of Education Policy and Leadership*, *16*(7).	Canada	Qualitative	Journal article	Knowledge brokers
Malin, J. R., & Paralkar, V. K. (2017). Educational Knowledge Brokerage and Mobilization: The" Marshall Memo" Case. *International Journal of Education Policy and Leadership*, *12*(7), n7.	USA	Qualitative	Journal article	School leaders

Morrison, J. R., Ross, S. M., & Cheung, A. C. (2019). From the market to the classroom: How ed-tech products are procured by school districts interacting with vendors. *Educational Technology Research and Development*, *67*, 389-421.	USA	Mixed methods	Journal article	School leaders
Naff, D. B. (2020). Podcasting as a Dissemination Method for a Researcher-Practitioner Partnership. *International Journal of Education Policy and Leadership*, *16*(13), n13.	USA	Mixed methods	Journal article	School leaders
Nelson, S. R., Leffler, J. C., & Hansen, B. A. (2009). Toward a research agenda for understanding and improving the use of research evidence. *Northwest Regional Educational Laboratory (NWREL)*.	USA	Qualitative	Technical report	School leaders Policymakers
Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A. R. (2016). Findings from a National Study on Research Use among School and District Leaders. Technical Report No. 1. *National Center for Research in Policy and Practice*.	USA	Mixed methods	Technical report	School leaders District leaders
Penuel, W. R., Briggs, D. C., Davidson, K. L., Herlihy, C., Sherer, D., Hill, H. C., Farrell, C., & Allen, A. R. (2017). How school and district leaders access, perceive, and use	USA	Mixed methods	Journal article	School leaders District leaders

research. *AERA Open*, *3*(2), 2332858417705370.				
Prendergast, S., & Rickinson, M. (2019). Understanding school engagement in and with research. *The Australian Educational Researcher*, *46*, 17-39.	Australia	Mixed methods	Journal article	School leaders District leaders
Wentworth, L., Mazzeo, C., & Connolly, F. (2017). Research practice partnerships: A strategy for promoting evidence-based decision-making in education. *Educational Research*, *59*(2), 241-255.	USA	Mixed methods	Journal article	School leaders District leaders
Witherow, K. (2011). *Research use and the impact in secondary schools*. University of Toronto.	Canada	Qualitative	Dissertation	School leaders
Yoshizawa, L. (2020). *Research use under pressure: State and district implementation of the ESSA evidence requirements* (Doctoral dissertation, Harvard University).	USA	Qualitative	Dissertation	School leaders District leaders

Chapter 3: Motivating Use of Research in Education

Claire Chuter

Johns Hopkins University

School of Education

Abstract

Despite widespread efforts to support teacher use of research, teachers rarely use education research in practice. Descriptive research has supported the idea that a major barrier to teachers' use of research is low comprehension of the research itself. In this experiment, I rely on Mayer's work around the use of examples and images to support understanding of a piece of research. In addition to research comprehension, I also measure the effect of examples and images on downstream effects (i.e., self-efficacy, and future use of research). Results indicate that examples and images may be effective at improving understanding, self-efficacy, and use of research evidence by teachers.

Introduction

The idea that science should contribute to practice is not new, but the last two decades have seen a resurgence of interest in the connection between research and practice in education (Darling-Hammond et al., 2016; Dynarski, 2015; Herman et al., 2017). In the political sphere, the bipartisan passage of the *No Child Left Behind Act* (NCLB; Simpson et al., 2004) signaled a seismic shift towards the use of effective educational practices based on scientifically-based research. For the first time, applications for Title I funds by schools, districts, and state departments of education were required to support their instructional and professional development decisions through "scientifically-based research." The act defined "scientifically-based research" as "(i) employing systematic, empirical methods; (ii) involving rigorous data analyses… (iii) relying on measurements or observational methods that provide reliable and valid data … and (iv) uses every opportunity to conduct experimental or quasi-experimental designs… with appropriate controls to evaluate the effects of the condition of interest" (20 U.S.C. § 6319, 2002)[2]. Subsequently, the U.S. Department of Education established the What Works Clearinghouse (WWC, 2002) to identify scientifically based research, and to promote its the use in practice. Open-access resources of evidence-based educational practices such as *Doing What Works* and the *Best Evidence in Brief* were also created for the purpose of increasing research use among educators. Finally, initiatives to promote the dissemination and use of research evidence in education have also expanded (e.g., the *Campbell Collaboration, the William T. Grant Foundation, Research for Action*).

[2] In 2015, the *Every Student Succeeds Act* reaffirmed NCLB's commitment to evidence-based practice, again mandating that K-12 educational agencies invest federal education funds in evidence-based practices (20 U.S.C. § 6301, 2015).

Yet despite the fervor surrounding use of evidence in education, there has been repeated disappointment with the progress that has been made thus far (Broekkamp & van Hout-Wolters, 2007; Bryk et al., 2011). Despite considerable effort (McIntyre, 2005; Vanderlinde & van Braak, 2010), the paucity of research use in educational practice has persisted (Dagenais et al., 2012; Lysenko et al., 2014). While school districts may maintain an expectation that teachers use research to inform their work (Stanovich, 2003), reported use by teachers in practice remains low (Rickinson et al., 2021). Bérubé (2005) found that the majority of teachers consulted academic research between "never" and "sometimes." Similarly, Cousins and Walker (2000) found that schoolteachers seldom consulted research in their work. Williams and Coles (2007) also reported that between 60% and 80% of school practitioners used research from "never" to "occasionally".

Based on the apparent lack of use, one might assume that teachers' do not see the value in education research. Yet multiple studies have found that educators repeatedly emphasize the importance of making research-informed decisions in their work (Coburn & Talbert, 2006; Datnow & Hubbard, 2016; Rickinson et al., 2021).[3] Other studies find that people in schools value forms of evidence other than the standardized tests emphasized by accountability policy (Ingram et al., 2004; Morrison et al., 2009).

This apparent contradiction of beliefs begs the question: *If teachers think evidence can improve education, then why don't they use it*? There are many immediate responses to this question backed by surveys of teacher experiences (e.g., lack of time, leadership, or resources; Coburn & Turner, 2011; Farley-Ripple et al., 2018). However, this study seeks to evaluate the impact of one barrier in particular that is often mentioned: the difficulty in interpreting education

[3] Despite reporting high valuing of research, it is possible that teachers in fact do not value education research. Teachers may feel unduly pressured to be in favor of evidence-based practice, even in anonymous responses.

research (Lysenko et al., 2014; Bransford et al., 2009; Ion & Iucu; 2014; Rickinson et al., 2021). The current work is especially potent given the abundance of knowledge mobilization organizations paired with the paucity of research in this area. To my knowledge, this is the first study to experimentally evaluate the impact of a knowledge mobilization strategy on research use.

Conceptual Framework

To conceptualize the impact of improving comprehension on a piece of research on use of that research in the classroom, I relied on the expectancy-value theory. This framework posits that valuing of an outcome works in tandem with expectations of success towards that outcome, to catalyze? motivation (Wigfield & Eccles, 2000). Wigfield and Eccles (2000) conducted studies around how children's ability-related beliefs and subjective values develop across the school years and relate to performance and choice. They found that children's beliefs regarding their ability and expectancies for success were the strongest predictors of subsequent grades in math, predicting those outcomes more strongly than either previous grades or achievement values (Wigfield & Eccles, 2000; Eccles et al., 1983; 1998). In addition, children's subjective task values are the strongest predictors of children's intentions to keep taking math and actual decisions to do so. They then extrapolated these findings to the larger learner population.

In other words, *both* high valuing and high expectations must be present for motivation towards an outcome to also be high. Without one or the other, motivation towards the goal is severely weakened. Applying this framework to the current state of educational research use: teachers seem to struggle with low expectations of success to use research despite their high valuing of success. Thus, an intervention to support teachers' expectations of success through

increased capacity may support motivation to use research. The current study seeks to develop and test such an intervention.

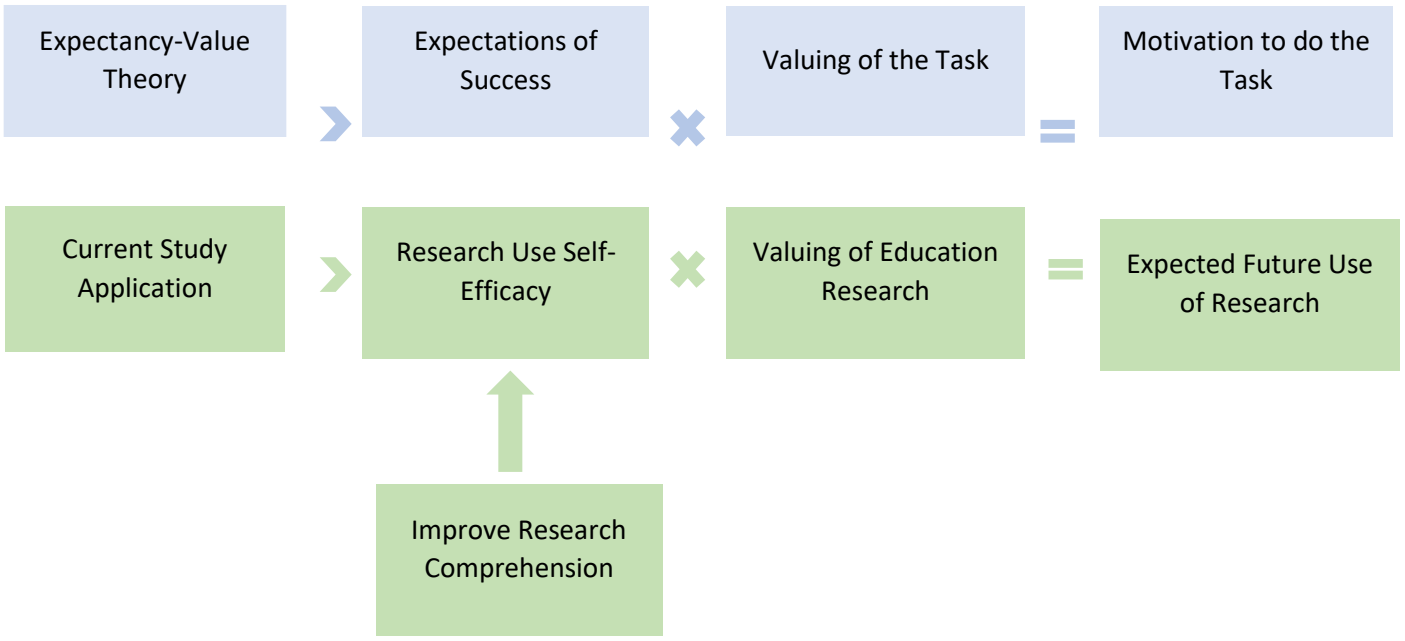
In developing such an intervention, I found the Mayer's work on learning strategies 2003) particularly relevant. Across multiple studies, Mayer and colleagues have found that by adding dimension to presented information individuals perform better on "knowledge acquisition" and "knowledge understanding" (Mayer, 1988, 2002, 2019; Quilici & Mayer, 1996). Mayer has found that by perceiving the same information presented in a variety of ways, readers are better able to identify the relevant information, and integrate the new content with previously learned knowledge (Mayer et al., 2005). Mayer has tested this approach using examples and images and found both to be effective at improving comprehension with students from middle school to postsecondary (Mayer, 2002; Mayer & Moreno, 2003; Quilici & Mayer, 1996). This study seeks to evaluate the efficacy of including examples and images alongside text, with a teacher population.

Specifying Outcomes

The focal outcomes of this study are *(1) expected future use of research, (2) research use self-efficacy, (3) valuing of education research, and (4) understanding of education research.* Relying on expectancy-value theory, the outcomes are conceptualized according to Figure 1 below. The expectancy-value framework is visualized in the blue row, while the current study application is illustrated below in the green row.

Figure 1

Application of Expectancy-Value Framework



As illustrated above, motivation (operationalized here as *expected future use of research*) is a function of two requisite variables: *research use self-efficacy* and *valuing of education research.* This study seeks to improve understanding of a piece of research through images and/or examples, in order to reinforce self-efficacy to use that specific piece of research in their classroom.

Hypotheses

With these focal outcomes in mind, I designed a preregistered study (Chuter, 2023; Gehlbach & Robinson, 2021) with a sample of teachers using the survey platform, Prolific. I preregistered four confirmatory hypotheses that, relative to the control group:

1) …the example and example + image treatment conditions will perform better on the *comprehension of presented research* test.
2) …the example and example + image treatment conditions will score higher on *expected future use of education research*.
3) …the example and example + image treatment conditions will score higher on *research use self-efficacy*.
4) The higher individuals scored on the pre-intervention *valuing of education research* item, the greater the positive effect of the example and example + treatment conditions on the *expected use of education research.*

In addition, I explored the hypothesis that valuing of education research may improve in the example and example + image conditions in comparison to the control group. This hypothesis is less grounded in prior literature. However, in some studies (Gehlbach, Mu, et

al., 2022; Gehlbach, Robinson, et al., 2022), authors found that valuing of biodiversity was positively impacted by the presence of images, so this outcome may change as a result of the intervention.

Methods

This study examined the impact of images and/or examples alongside a description of research, on motivation to use education research by teachers. Teacher participants were randomly presented with either (1) a description of research only, (2) a description alongside an example of the research being used, or (3) a description alongside an example and a relevant image.

Participants

Participants (N = 164) were K-12 teachers recruited through the survey platform Prolific. Only participants that had listed their occupation as 'teacher' when joining the platform originally, were able to view this study. The final sample was 71% female; 86% identified as White, 5% Asian/Pacific Islander, 2% as Latino/a or Hispanic American, 2% as African American, 1% as Middle Eastern, and 2% as 'Other.' Fifty-seven percent of the sample identified as politically liberal, 25% as moderate, and 18% as conservative. The mean age of participants was 42.78 (SD = 11.12). Participants were paid $1.35 according to our lab's $15/hour payment policy (estimated survey time was 5.4 minutes).

Measures

Four response variables are examined in this experiment: *(1) valuing of education research, (2) understanding of education research, (3) research use self-efficacy, and (4) future*

use of research. Data fit and reliability were assessed using χ^2 and Cronbach's alpha (α) statistics (Klem, 2000; Bowen & Guo, 2011). Scales and test can be found in Appendix A.

Valuing of Education Research

The first construct was assessed through a self-report measure, adapted from the Trust in Science Scale (Nadelson et al., 2014). Participants rated their valuing of education research through one time pre-intervention, and two items post-intervention (with fully labeled response options). To assess the level that person-specific valuing of education research interacted with the effect of the intervention, one valuing item was asked prior to the intervention. In other words, I was interested in the extent to which an individual's pre-existing *valuing of education research* might impact their *expected future use of research* dependent on their treatment condition. The following pre-intervention was used to measure pre-existing *valuing of education research*: *'To what extent is education research useful to teachers?'.*

In addition, to assess how *valuing of education research* might change according to the treatment condition, two post-intervention valuing items were also included. The *valuing of education research scale* included the two following questions*: 'How much effort would you put into learning a new curriculum if it were backed by stronger evidence of student success than your current curriculum?'* and *'How much do you think teachers' should pay attention to what is happening in education research?'.* A composite score was generated by taking the mean of the two items. Cronbach's alpha showed good reliability ($\alpha = .75$), and the chi-squared test rejected the null hypothesis that the two post-intervention items were independent of one another ($\chi^2(12, N = 161) = 63.51$, p<.00).

Comprehension of Research

Participants' comprehension of the presented research was measured through three multiple-choice items. A composite score for understanding of research was created by taking the sum of all three items for each respondent. The test included questions such as, '*Which of the following are examples of how this instructional strategy might be used?*' Because this was a test and not a scale, a reliability coefficient was not collected. The full test can be found in Appendix A.

Research Use Self-Efficacy

Research use self-efficacy was assessed through a self-report measure of two items (with fully labeled response options). The items include: *'If you were to implement this strategy, how confident are you that you would be successful?'* and '*How confident do you feel that could implement the strategy presented with your own students?'.* A composite score was developed by taking the mean of both items. Cronbach's alpha shows good reliability (α = .74), and the chi-squared test rejected the null hypothesis that the two post-intervention items were independent of one another ($\chi^2(12, N = 162) = 79.46$, p<.00).

Expected Future Research Use

Expected future research use was assessed through a self-report scale of two items: '*How likely are you to use this research with your students?'* and *'How likely are you to share this technique with other teachers?'*. A composite score was developed by taking the mean of both items. Cronbach's alpha shows moderate reliability (α = .67), and the chi-squared test rejected the null hypothesis that the two post-intervention items were independent of one another ($\chi^2(16, N = 161) = 79.49$, p<.00).

Procedure

After completing the consent form, participants were asked an attention check item to ensure high study quality. Participants who failed this attention check were rerouted directly to the end of the survey, and informed that any re-entries to the survey would not be counted. All participants who passed the attention check were then given the first valuing of education research item. Then participants were randomly sorted into one of three conditions: (1) a control group which included only a description of education research, that may be used in classroom instruction, (2) an example-only treatment group, which includes the description and an example of how this research may be used, or (3) an example and image related to the research topic described. Participants in all groups then took the comprehension test, self-efficacy, future use of research, and second piece of the valuing scale. Finally, participants were asked a series of demographic items (e.g., gender, age, race, political orientation). Full treatment details can be found in Appendix B.

Analysis

To test our confirmatory and explanatory hypotheses, I fit OLS models that regressed each focal outcome on the treatment variables and any potentially relevant covariate as follows:

$$Outcome_i = \beta_0 + \beta_1 Example_i + \beta_1 Example{+}Image_i + \ \Gamma_1 X_i + \varepsilon_i$$

Here, $Outcome_i$ represents each of our four outcomes of interest: *comprehension of research,* (2) *expected future use of research,* (3) *research use self-efficacy,* and (4) *expected future use of education research*. $Example_i$ indicates that participant *i* was exposed to the example condition; $Example + Image_i$ indicates that participant *I* was exposed to the example and image condition; X_i is any participant-level covariate that was imbalanced across our experimental conditions as described in our preregistration (Chuter, 2023); and ε_i is a residual.

According to Cumming's (2014) recommendations, I evaluated these focal hypotheses by presenting and discussing confidence intervals and effect sizes rather than p-values. A Bonferroni correction will be applied given the number of models tested. This correction divides the original alpha level (i.e., .05) by the number of comparisons). Therefore, to achieve statistical significance of the treatment on each dependent variable, p-values must reach a level of .0125. As specified in the preregistration, I removed data from participants if they failed our initial attention check, engaged in straight-line responding, or sped through the survey. For the attention check, participants who did not read closely enough to follow our instructions on an initial item were immediately exited from the survey (and invited to participate in a future study instead).

Using Barge and Gehlbach's (2012) approach, I removed the data from any participants who gave identical sequential responses on the comprehension test, or on the entire last page of dependent variables (i.e., self-efficacy, expected future behaviors, and valuing). Our preregistration also described our approach for removing data from participants who sped through the survey without paying adequate attention. Specifically, I set our speeding criteria at 700 words per minute, the pace at which college educated speed readers can read (normal reading speed is thought to be about 200–250 words per minute, Seidenberg, 2017). Because our participants had to read at this rate *and* check off responses to survey items, I assumed that anything faster than 700 words per minute meant that our participants were not paying adequate attention.

To test whether the 164 participants were evenly balanced between control (n = 56), example (n =54), and example + image conditions (n = 58) participants based on their demographic characteristics, I fit logistic regressions predicting condition based on each demographic variable (i.e., age, gender, race, education, father's education, and political

orientation) to ensure that demographic characteristics were balanced across conditions. I also performed an additional logistic regression to assess differential attrition between individuals who were dropped from the sample due to speeding and those who remained in the sample.

Results

Selection Bias Assessment

There were 200 participants who entered the study. After data quality assessments outlined above, 36 individuals were removed from the study. Attrited individuals were on average 5.32 years younger (p <.001), 38% less likely to be White (p <.001). Across gender, education, and political orientation, attrited individuals were similar to non-attrited individuals. This may limit the generalizability of our results.

Randomization Assessment

Individuals included in the final sample differed across treatment conditions by level of education (p<.05). Therefore, education was included as a covariate in the following models.

Table 1

Descriptive Statistics and Correlations for Focal Outcomes

Variable	*n*	*M*	*SD*	1	2	3	4
1. Valuing of Education Research	161	1.64	0.56	-			
2. Understanding of Education Research	161	3.95	0.83	.06	-		
3. Research Use Self-Efficacy	161	3.41	0.87	.48*	.24*	-	
4. Expected Future Research Use	161	3.90	0.64	.57*	.15	.72*	-

Note. *p<.05; **p<.001; ***p<.001.

Confirmatory Results

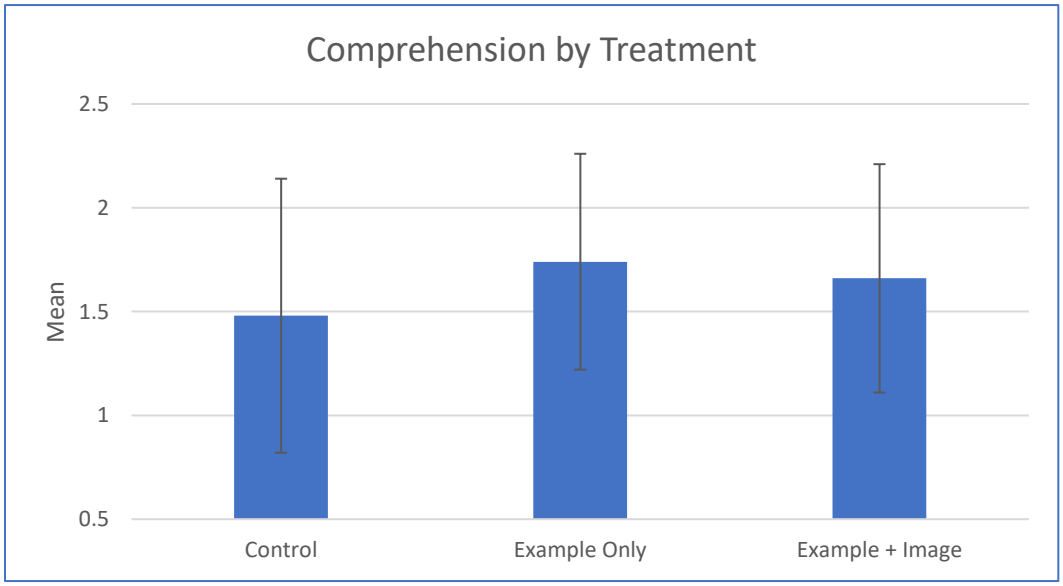
This study evaluated the impact of an image and/or example paired with a description of a piece of research on (1) *comprehension of research,* (2) *expected future use of research,* (3) *research use self-efficacy,* and (4) *expected future use of education research.*

Comprehension of Research

Individuals in the example treatment condition scored minimally higher on the comprehension of education research test than the control condition *($M_{Example}$* = 1.74, SD = 0.52; $M_{Control}$ = 1.48, SD = 0.66; $F_{(3\text{-}157)}$ = 2.21, $\beta_{Example}$ = .21; 98.33% CI = -0.01, 0.42; Cohen's *d* = 0.43). Individuals in the example + image treatment condition did not score higher on the comprehension of education research test *(M_{Image}* = 1.66, SD = 0.55; $M_{Control}$ = 1.48, SD = 0.66; $F_{(3\text{-}157)}$ = 2.21, β_{Image} = .10; 98.33% CI = -0.11, 0.31; Cohen's *d* = 0.29; see Figure 2).

Figure 2

Comprehension of Research by Treatment

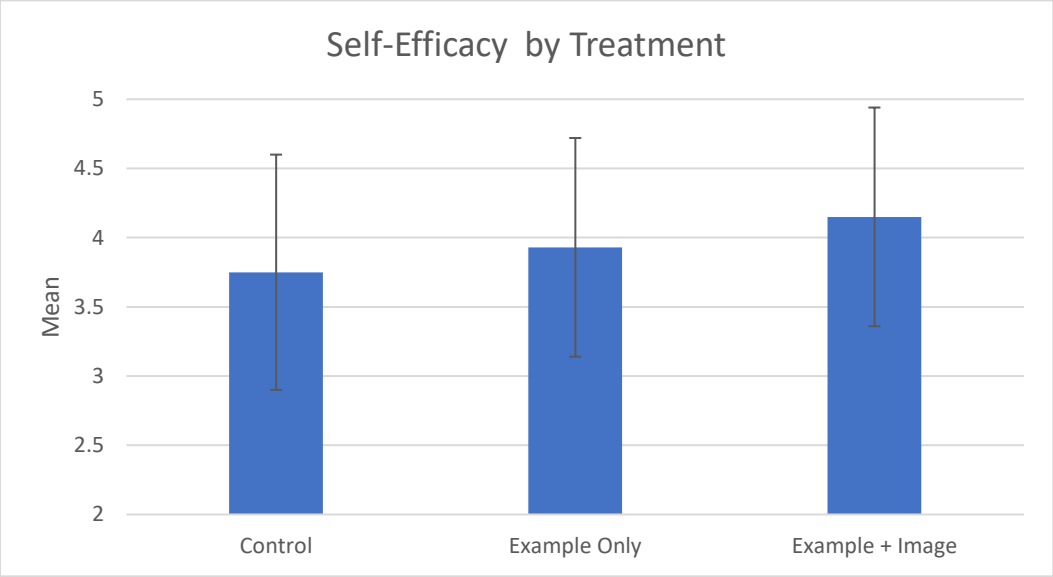


Research Use Self-Efficacy

Individuals in the example treatment condition did not score higher on the research use self-efficacy scale *(*M_{Example} = 3.93, SD = 0.79; M_{Control} = 3.75, SD = 0.85; $F_{(3\text{-}157)}$ = 2.39, β_{Example} = .20; 98.33% CI = -0.19, 0.60; Cohen's *d* = 0.23). Individuals in the example and image treatment condition scored higher on the research use self-efficacy scale *(*M_{Image} = 4.15, SD = 0.79; M_{Control} = 3.75, SD = 0.85; $F_{(3\text{-}157)}$ = 2.21, β_{Image} = .41; 98.33% CI = 0.03, 0.80; Cohen's *d* = 0.49; see Figure 3).

Figure 3

Research Use Self-Efficacy by Treatment



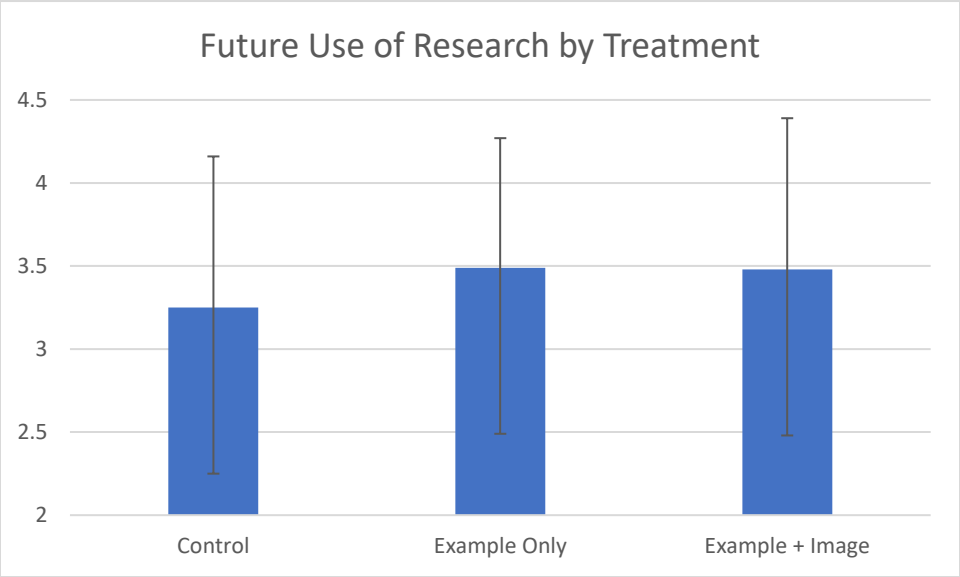
Expected Future Use of Research

Individuals in the example treatment condition did not score higher on the expected future use of research scale *(*M_{Example} = 3.59, SD = 0.78; M_{Control} = 3.25, SD = 0.91; $F_{(3\text{-}157)}$ = 2.39, β_{Example} = .25; 98.33% CI = -0.16, 0.67; Cohen's *d* = 0.29). Individuals in the example and image treatment condition also scored higher on the expected future use of research scale *(*M_{Image} =

3.48, SD = 0.91; $M_{Control}$ = 3.25, SD = 0.91; $F_{(3\text{-}157)}$ = 2.21, β_{Image} = .24; 98.33% CI = 0.17, 0.65; Cohen's d = 0.26; see Figure 4).

Figure 4

Expected Future Use of Research by Treatment



Expected Future Research Use* by *Valuing of Education Research

Participants reporting of future expected use of education research by example treatment was not affected by prior valuing of education research ($F_{(6\text{-}154)}$ = 14.75, $\beta_{Example}$ = -.28; 98.33% CI = -0.09, -0.67). This was contrary to our preregistered hypothesis. Participants reporting of future expected use of education research use by image treatment was not affected by prior valuing of education research ($F_{(4\text{-}156)}$ = 9.35, β_{Image} = .15; 98.33% CI = -0.26, 0.56).

Exploratory Hypotheses

Valuing of Education Research

Participants in the example condition did not value education research more highly as a result of the experiment than those in the control *(*M_{Example} = 3.86, SD = 0.62; M_{Control} = 3.93, SD = 0.64; $F_{(2\text{-}158)}$ = 0.18, β_{Example} = -.09; 98.33% CI = -.35, 0.17; Cohen's *d* = -0.12). Participants in the example + image condition did not value education research more highly as a result of the experiment than those in the control *(*M_{Image} = 3.78, SD = 0.73; M_{Control} = 3.93, SD = 0.64; $F_{(2\text{-}158)}$ = 0.73, β_{Image} = -.16; 98.33% CI = -0.41, 0.09; Cohen's *d* = 0.23).

Discussion

The results of this study are promising though weak. Both treatment conditions indicated positive effect sizes for every focal outcome (i.e., comprehension of research, self-efficacy to use research, and expected future use of research). It certainly seems unlikely that every main outcome shows a positive effect - though weak - due to chance alone. However, few of these outcomes demonstrated a strong or significant effect. Contrary to our hypothesis, expected future use of education research by treatment was not dependent on valuing of education research. This finding contrasts with the expectancy-value framework, which posits that higher valuing of a task motivates future action when combines with improved expectations of success.

The results here are promising given the light touch of the intervention, and the small sample size. The positive impact of the examples and images further support Mayer's supposition that providing additional ways to represent new information strengthens learning and retention. Furthermore, the positive impact of both treatment conditions on *expected future use of research* and *self-efficacy to use research* in education support the expectancy-value theory's conceptualization of task motivation. Interestingly, the impact of these interventions on *expected future use of motivation* was not modulated by individuals' *valuing of education research.* This

would have been expected is the expectancy-value theory's framework were at play in this work, because individuals with higher *valuing of education research* would have the most to gain from breaking down the barrier of poor research understanding. Therefore, this finding calls into question whether the expectancy-value framework may be functioning in the traditional way with this intervention.

The finding does not align with prior work in the expectancy-value literature in which increased self-efficacy in coordination with valuing of the task leads to greater motivation (Eccles, 1983; 1992). However, it does align with the underlying assumption of how expectancy-value theory works - that self-efficacy is an important cause of motivation in and of itself. This literature has also found that valuing is task-specific – valuing of one goal does not necessarily translate to valuing of a similar goal (Atkinson, 1957; Battle, 1965; Wigfield & Eccles, 1992). Therefore, it is possible that the items asking about valuing of research in general did not map onto valuing of the specific research presented within this study.

Limitations and Future Directions

This study suffered from small sample size with only approximately 50 participants per treatment condition. In addition, the study's results may be weakened by selection bias on two levels. First, individuals were recruited from the survey platform Prolific, which is not representative of the general teaching population. Second, our analyses found that individuals that attrited from our sample were younger (by approximately five years) and less likely to by White than our final sample, thus compromising the generalizability of this work. Finally, the comprehension test included a misleading item, which ultimately needed to be excluded from the study. Future research should mitigate these issues by expanding the participant sample to a larger, more representative pool.

References

Atkinson, J. W. (1957). Motivational determinants of risk taking behavior. Psychol. Rev. 64: 359-372.

Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education*, *53*(2), 182-200.

Battle, E. (1965). Motivational determinants of academic task persistence../. Person. Soc. Psychol. 2: 209-218.

Bérubé, B. (2005). *L'accès à la recherche en enseignement et son utilisation dans la pratique*. desLibris.

Bowen, N. K., & Guo, S. (2011). *Structural equation modeling*. Oxford University Press.

Bransford, J. D., Stipek, D. J., Vye, N. J., Gomez, L. M., & Lam, D. (2009). The Role of Research in Educational Improvement. In *Harvard Education Press*. Harvard Education Press.

Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation*, *13*(3), 203–220.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. *Frontiers in Sociology of Education*, 127–162.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2016). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?.

Chuter, C. (2023, February 14). Research Use among Teachers. Retrieved from osf.io/yfha7

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American Journal of Education*, *112*(4), 469–495.

Coburn, C. E., & Turner, E. O. (2011). Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research and Perspectives*, *9*(4), 173–206. https://doi.org/10.1080/15366367.2011.626729

Cousins, J. B., & Walker, C. A. (2000). Predictors of educators' valuing of systematic inquiry in schools. *Canadian Journal of Program Evaluation*, 25–52.

Cumming, G. (2014). The new statistics: Why and how. *Psychological science*, *25*(1), 7-29.

Dagenais, C., Lysenko, L., C. Abrami, P., M. Bernard, R., Ramde, J., & Janosz, M. (2012). Use of research-based information by school practitioners and determinants of use: A review of empirical research. *Evidence & Policy*, *8*(3), 285–309.

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Palo Alto, CA: Learning Policy Institute.

Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *17*(1), 7–28. https://doi.org/10.1007/s10833-015-9264-2

Dynarski, M. (2015). Using research to improve education under the Every Student Succeeds Act. *Evidence Speaks Reports*, *1*(8), 1–6.

Eccles J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.),

Achievement and achievement motivation (pp. 75–146). San Francisco, CA: W. H. Freeman.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) and N. Eisenberg (Vol. Ed.), Handbook of child psychology (5th ed., Vol. III, pp. 1017–1095). New York: Wiley.

Farley-Ripple, E., May, H., Karpyn, A., Tilley, K., & McDonough, K. (2018). Rethinking Connections Between Research and Practice in Education: A Conceptual Framework. *Educational Researcher*, *47*(4), 235–245. https://doi.org/10.3102/0013189X18761042

Gehlbach, H., Mu, N., Arcot, R. R., Chuter, C., Cornwall, K. J., Nehring, L., Robinson, C. D., & Vriesema, C. C. (2022). Addressing the vexing educational challenges of biodiversity loss: A photo-based intervention. *Contemporary Educational Psychology*, *71*, 102096.

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79–89.

Gehlbach, H., Robinson, C. D., Vriesema, C. C., Bernal, E., & Heise, U. K. (2022). Worth more than 1000 words: How photographs can bolster viewers' valuing of biodiversity. *Environmental Conservation*, *49*(2), 99–104.

Herman, R., Gates, S. M., Arifkhanova, A., Bega, A., Chavez-Herrerias, E. R., Han, E., Harris, M., Tamargo, J., & Wrabel, S. L. (2017). *School leadership interventions under the every student succeeds act: Evidence review*.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, *106*(6), 1258–1287.

Ion, G., & Iucu, R. (2014). Professionals' perceptions about the use of research in educational practice. *European Journal of Higher Education*, *4*(4), 334–347.

Klem, L. (2000). Structural equation modeling.

Levin, B. (2013). To know is not enough: Research knowledge and its use. *Review of education*, *1*(1), 2-31.

Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational Research in Educational Practice: Predictors of Use. *Canadian Journal of Education / Revue Canadienne de l'éducation*, *37*(2), 1–26.

Mayer, R. E. (1988). Learning strategies: An overview. *Learning and Study Strategies*, 11–22.

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71.

Mayer, R. E. (2019). Thirty years of research on online learning. *Applied Cognitive Psychology*, *33*(2), 152–159.

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R., Veronikas, S., & Shaughnessy, M. F. (2005). An Interview with Richard Mayer. *Educational Psychology Review*, *17*(2), 179–189.

McIntyre, D. (2005). Bridging the gap between research and practice. *Cambridge Journal of Education*, *35*(3), 357–382.

Morrison, J. A., Raab, F., & Ingram, D. (2009). Factors influencing elementary and secondary teachers' views on the nature of science. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(4), 384–403.

Nadelson, L., Jorcyk, C., Yang, D., Jarratt Smith, M., Matson, S., Cornell, K., & Husting, V. (2014). I just don't trust them: the development and validation of an assessment instrument to measure trust in science and scientists. *School Science and Mathematics*, *114*(2), 76-86.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144.

Rickinson, M., Gleeson, J., Walsh, L., Cutler, B., Cirkony, C., & Salisbury, M. (2021). *Research and evidence use in Australian schools: Survey, analysis and key findings*.

Seidenberg, M. (2017). *Language at the Speed of Sight: How we Read, Why so Many CanÕt, and what can be done about it*. Basic Books.

Simpson, R. L., Lacava, P. G., & Sampson Graner, P. (2004). The No Child Left Behind Act: Challenges and Implications for Educators. *Intervention in School and Clinic*, *40*(2), 67–75. https://doi.org/10.1177/10534512040400020101

Stanovich, P. J. (2003). *Using research and reason in education: How teachers can use scientifically based research to make curricular & instructional decisions*. Partnership for Reading (Project), National Institute for Literacy, US ….

Vanderlinde, R., & van Braak, J. (2010). The gap between educational research and practice: Views of teachers, school leaders, intermediaries and researchers. *British Educational Research Journal*, *36*(2), 299–316.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. Educational Psychology Review, 6, 49–78.

Wigfield, A., and Eccles, J. S. (1992). The development of achievement task values: A theoretical analysis. Devel. Rev. 12: 265-310.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81.

Appendix A
Research Use Scales and Test

Research Use Valuing Scale

1. To what extent is education research useful to teachers?
2. How much effort would you put into learning a new curriculum if it were backed by stronger evidence of student success than your current curriculum?
3. How much do you think teachers should pay attention to what is happening in education research?

Comprehension of Research Test

1. Which of the following are examples of how this strategy might be used?
2. What is the name of the strategy described previously?
3. How does the research show this strategy supports students' social-emotional growth?

Research Use Self-Efficacy Scale

1. How confident do you feel that could implement the strategy presented with your own students?
2. If you were to implement this strategy, how confident are you that you would be successful?

Expected Future Research Use Scale

1. How likely are you to use at least this strategy in your teaching?
2. How likely are you to share this technique with other teachers?

Appendix B

Study Conditions

Control Condition

"Why" Practice

Based on research by Stefanou et al., (2004), 'why' practice encourages student intentionality in their problem solving. It simply involves giving the students' a task with many potential ways to address the task and asking them to consider 'why' they made each choice. Stefanou and colleagues (2004) found that by regularly asking students to consider their reason for doing something, students were more likely to gain intentionality and autonomy in their work.

Reference: Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. Educational psychologist, 39(2), 97-110.

Example Condition

"Why" Practice

Based on research by Stefanou et al., (2004), 'why' practice encourages student intentionality in their problem solving. It simply involves giving the students' a task with many potential ways to address the task and asking them to consider 'why' they made each choice. Stefanou and colleagues (2004) found that by regularly asking students to consider their reason for doing something, students were more likely to gain intentionality and autonomy in their work.

For example: Students may be given a set of painting pots, and canvases. As you walk around the room, regularly ask students to explain why they chose a specific color or pattern. Answers like, "because it looks cool" do not pass the why check. Keep asking until you reach a more thoughtful response. This teaching strategy will encourage students' thoughtfulness around why they do what they do.

Reference: Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. Educational psychologist, 39(2), 97-110.

Example + Image Condition

"Why" Practice

Based on research by Stefanou et al., (2004), 'why' practice encourages student intentionality in their problem solving. It simply involves giving the students' a task with many potential ways to address the task and asking them to consider 'why' they made each choice. Stefanou and colleagues (2004) found that by regularly asking students to consider their reason for doing something, students were more likely to gain intentionality and autonomy in their work.

For example: Students may be given a set of painting pots, and canvases. As you walk around the room, regularly ask students to explain why they chose a specific color or pattern. Answers like, "because it looks cool" do not pass the why check. Keep asking until you reach a more thoughtful response. This teaching strategy will encourage students' thoughtfulness around why they do what they do.

Reference: Stefanou, C. R., Perencevich, K. C., DiCintio, M., & Turner, J. C. (2004). Supporting autonomy in the classroom: Ways teachers encourage student decision making and ownership. Educational psychologist, 39(2), 97-110.

Chapter 4: Publication Bias of Social-Emotional Learning Interventions

Claire Chuter

Johns Hopkins University

School of Education

Abstract

Education practitioners and policymakers are increasingly pressured to use evidence to inform their decision-making. Therefore, it is imperative that they can have confidence in the results that emerge from education research. Although publication bias negatively impacts the trustworthiness of results that emerge from education research, few meta-analytic reviews conduct assessments. This study assesses and adjusts for the impact of publication bias in social-emotional learning interventions and compares the conclusions of varying publication bias methods. I conducted (1) a visual inspection of funnel plot asymmetry, using contour-enhancement, (2) the trim-and-fill method, (3) Egger's regression significance tests, and (4) Vevea and Wood's stepwise selection model. Results indicated that publication bias was present as the result of clustering around significance levels but not as a result of small-study effects. These results emphasize the importance of conducting multiple forms of publication bias methods.

Introduction

Education practitioners and policymakers are increasingly pressured to use evidence to inform their decision-making (Ginsberg et al., 2022; Yoshizawa, 2020). Therefore, it is imperative that they can place confidence in the results that emerge from education research. This is particularly true among meta-analytics findings, as they are more likely to be used to influence policy (Forness, 2001; Glass et al., 1981). However, the replication crisis has raised concerns regarding the replicability of education research (Gehlbach & Robinson, 2021; Plucker & Makel, 2021). Traditional research practices such as p-hacking, paired with selective reporting, can lead to an inflated number of positive effect sizes in the published literature that fail to replicate in practice. If policymakers wish to engage in evidence-based practice and to direct funding toward educational policies and/or practices that are effective, researchers must address inflated findings in education meta-analytic reviews. Among the chief concerns is that of publication bias (Hunter & Schmidt, 2004; Rothstein et al., 2005).

Publication bias occurs when research with favorable or significant outcomes is published at higher rates than research with nonsignificant or unfavorable results (Marks-Anglin & Chen, 2020). In other words, the chance of a study being published does not depend entirely on its quality, but also dependent on the effect size, significance, and direction of the results. This bias comes in two forms: outcome-level and study-level bias. Firstly, at the outcome-level (sometimes called selective reporting) occurs when authors filter their results to appear more appealing – sometimes of their own accord, sometimes to appease editors, reviewers, funders, or program developers to publishers, developers, or funders. Thus, the final draft may have a much higher proportion of large, positive, or significant results, than the initial analysis. Secondly, at the study-level, authors may leave whole papers with overall small, null, or uninteresting effects

unpublished because they believe that they have a high chance of reviewer rejection. Both forms of exclusion lead to an incomplete and skewed representation of intervention impacts (Franco et al., 2014; Møller & Jennions, 2001).

Both outcome-level and study-level bias have been found to be prevalent in education research (Pigott et al., 2013; J. R. Polanin et al., 2016). Pigott et al. (2013) investigated publication bias at the outcome level by comparing the reports of educational interventions from dissertations to their published versions. The authors found that nonsignificant outcomes were 30% less likely to appear in publication compared to statistically significant outcomes. Similarly, Chow & Ekholm (2018) found that for 1,752 studies included in education and special education meta-analyses, published studies had significantly higher effect sizes than unpublished studies (d=+0.64), and studies with larger effect sizes were more likely to be published than those with smaller effect sizes. Exploratory moderator analysis revealed that publication status was a significant predictor of effect size magnitude. Together, these findings highlight a grave problem for the evidence-based movement in education. Namely, publication bias leads to inflated effects that do not represent the true distribution of effects in reality. Yet, thus far social-emotional learning (SEL) has not been assessed in terms of publication bias. This paper seeks to assess and adjust for the impact of publication bias in SEL. First, this paper describes several methods for identifying publication bias, then implements some of these methods using a SEL dataset (Cipriano et al., 2023), and finally discusses the implications of selection certain publication bias methods over others.

Publication Bias Methods

As publication bias is prevalent and by definition misrepresents the overall implication of the literature, various methods have been devised to identify and adjust for its presence (Jin et

al., 2015; McShane et al., 2016; Møller & Jennions, 2001). In an ideal world, researchers would retrieve all relevant high-quality, unpublished results, and include them in the results. One popular approach is to contact the author of a primary study and ask for the missing data (Polanin & Williams, 2016). However, this method is time-consuming, and generally returns a low response-rate (Polanin & Espelage, 2019). Even if known missing outcomes are generally retrievable, unknown missingness due to publication bias remains a problem.

As a result, various statistical methods have been proposed to identify and adjust for publication bias (Marks-Anglin & Chen, 2020; Møller & Jennions, 2001). This section describes five methods and their related strengths and limitations. I choose to exclude the Fail-Safe-N method here because it has been repeatedly found to be faulty, misleading, and obsolete as a way to detect and account for publication bias (Rosenthal, 1979). I follow the recommendations of the Cochrane Collaboration (2022), Becker (2005), and Coyne (2014) to abandon use of the Fail-Safe-N for publication bias assessment.

It should be noted that the data used in education meta-analysis is almost always clustered, yet there is no publication bias statistical package for any of these methods which accounts for data dependencies. As a result, there may be higher incidence of false positives in regression and correlation publication bias tests described below.

Small-Study Effects Methods

The most common way of identifying evidence of publication bias is through a phenomenon called, “small-study effects” (SSE). SSE occurs when studies with smaller sample sizes tend to exhibit larger effects than studies with larger sample sizes. The smaller the sample, the lower degree of precision, and thus greater standard error. Smaller samples are less likely to

be significant unless they have a large effect size. On the other hand, larger studies can reach significance at smaller effect sizes. Furthermore, large studies may be more costly and be more likely to be published regardless of the outcome than small studies. Thus, if authors and reviewers are biased towards significant, positive effects, the lower the sample size, the greater the likelihood of having a large, significant effect.

According to SSE, in the absence of publication bias, sample size should not be associated with the effect size. However, there are other reasons for sample size to be associated with larger effect sizes, outside of publication bias. For example, small samples have higher levels of heterogeneity than large samples (IntHout et al., 2015; Ioannidis, 2008). It is impossible to know for certain whether SSE occurs as the result of publication bias, or heterogeneity. In addition, meta-analysts in the social sciences generally assume that meta-analyses do not consist of a collection of outcomes with one true effect size (Borenstein et al., 2010; Hedges & Vevea, 1998). Thus, from a theoretical standpoint, small-sample effects methods may not be appropriate for random-effects models. Despite these limitations, SSE methods can be useful in combination with selection models for pointing to *possible* publication bias (Sterne et al., 2011).

Among the many statistical methods on this issue, funnel plots have been commonly used to study SSE. Statistical tests based on funnel plot symmetry have been developed, including the rank correlation test (Begg & Mazumdar, 1994) and regression-based tests (Egger et al., 1997; Sterne & Egger, 2005). Using the symmetry of funnel plots, Duval & Tweedie (2000) developed the nonparametric trim and fill method for imputing missing studies in a meta-analysis further (Duval, 2005; Shi & Lin, 2019). These methods are discussed in more detail below.

Funnel Plots.

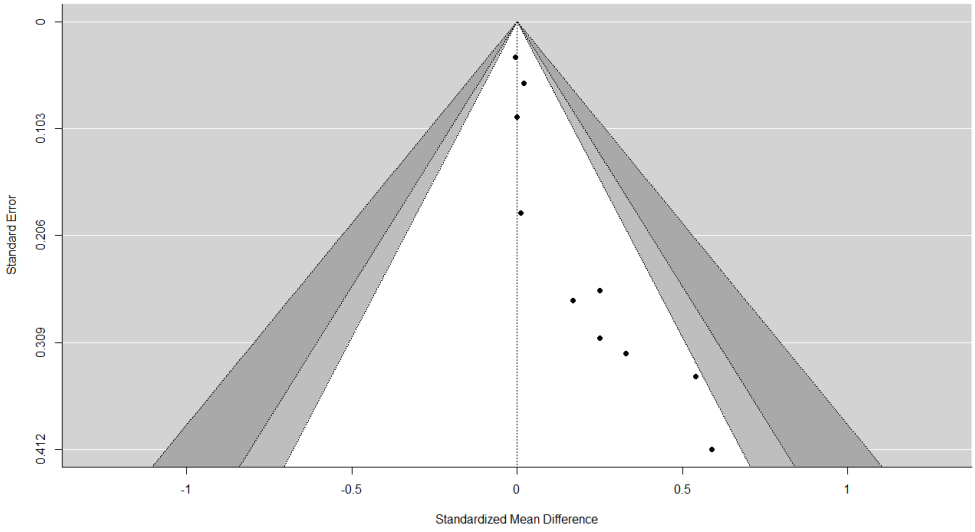
A funnel plot is a scatter plot of study sample size or variance on the vertical axis, and effect size on the horizontal axis (Sedgwick & Marston, 2015; Sterne et al., 2005, 2011). If no publication bias is present, effect sizes will resemble a symmetrical inverted funnel around the underlying true effect size, with more variability in the smaller studies than in the larger studies owing to the greater influence of sampling error. In reality, the true effect size is unknown, but it is often assumed that the top point of the funnel (indicating effect sizes for the largest studies) is a valid estimate of the true effect size. If a gap in one side of the lower extremities of the funnel is observed, publication bias may be suspected (Light & Pillemer, 1984). However, the direction of publication bias depends on the expectation from authors and reviewers. Therefore, gaps in the funnel plot may be detected on either side of the funnel plot and should be theoretically considered by the authors on a case-by-case basis.

Enhancing funnel plots is often recommended by adding contour lines to visualize areas of statistical significance within the plot (Palmer et al., 2008; Peters et al., 2008). Because it is generally assumed that the treatment effect of each study is normally distributed, then the significance of any effect size can be calculated from the effect size and the standard error. Contours representing conventional ‘‘milestone’’ levels of statistical significance (e.g., >0.01, >0.05, >0.1) can be defined and regions associated with these significance levels plotted. Contour enhancement is useful because the usual funnel plot may suggest asymmetry, but assessment of the contour-enhanced funnel plot indicates whether the areas where studies seem to be ‘‘missing’’ are areas where nonsignificant studies would be plotted (i.e., the area without shading). As described earlier, publishers prefer significant outcomes over non-significant outcomes. Thus, when missing studies fall in areas of non-significance, this adds further

credence to the possibility that the asymmetry observed is caused by publication bias. If, however, the ‘‘missing’’ studies were in areas of statistical significance -such as in Figure 1- this would lead one to suggest that the observed asymmetry is more likely to be due to factors other than publication bias based on statistical significance (e.g., variable study quality or even nonstatistical significance-based publication bias mechanisms). For example, in Figure 1, to mirror the existing studies on the right, missing studies should exist on the left side of the funnel. Because these missing studies appear in the nonsignificant region of the funnel, publication bias is particularly likely.

Figure 1

Contour-Enhanced Funnel Plot - Sample



It is important to note that inspection of the funnel plots assumes that a gap on the lower extremities of the funnel is due to publication bias. However, there are other reasons for an asymmetrical scatter plot. For example, covariate effects may distort the appearance of the funnel plot despite the absence of publication bias (Maier et al., 2022; Peters et al., 2010). Another reason for gaps in the funnel plot may be the result of small studies being prone to inflated effects (Cheung & Slavin, 2016). Most difficult to identify, plot asymmetry may also be the result of chance. Due to an inability to confirm that a gap in a funnel plot is definitively the result of publication bias, it is advisable to use this method as a complement to statistical tests for publication bias (Sterne et al., 2011).

Trim-and-Fill Method.

Trim-and-fill is an iterative method of formalizing assessment of the funnel plot (Duval, 2005; Duval & Tweedie, 2000; Shi & Lin, 2019). This method assumes that studies with extreme

effect sizes in an unfavorable direction are suppressed, causing funnel asymmetry. However, in reality most researchers assume that publication bias is largely based on p-values and sample size as well. Therefore, the expected direction of effect sizes preferred by reviewers and authors should be considered on a case-by-case basis. The idea of the trim-and-fill method is to first *trim* the studies that cause a funnel plot's asymmetry so that the overall effect estimate produced by the remaining studies can be considered minimally impacted by publication bias, and then to *fill* imputed missing studies in the funnel plot based on the bias-corrected overall estimate. First, the method calculates how many studies would need to be trimmed off the right side of the funnel, to leave a symmetric center. The adjusted overall effect size is then calculated. Then, the trimmed studies are added back in, together with the imputed studies on the opposite side of the funnel. The method not only indicates the existence of publication bias but also provide bias-adjusted results (Shi & Lin, 2019). Simulations suggest that it performs poorly in the presence of between-study heterogeneity when no publication bias is present (Terrin et al., 2003).

Rank Correlation Test and Egger's Test.

There are two significance tests to identify publication bias: the rank correlation test (Begg & Mazumdar, 1994), and Egger's test (Egger et al., 1997). The rank correlation test examines the correlation between effect sizes and their corresponding sampling variances. A strong correlation implies publication bias. Alternatively, Egger's test calculates the regression of the effect size on their level of precision. This method entails estimating a random-effects model that includes the standard errors of the effect size estimates as the predictor, weighted by the inverse sampling variances. This formula is shown below.

$$\frac{\widehat{g_{ik}}}{SE_{\widehat{gl_k}}} = \beta_0 + \beta_1 \frac{1}{SE_{\widehat{g_{ik}}}} + \beta_j + \omega_k + \varepsilon$$

In this model, *g* represents Hedge's *g, i* represents the individual outcome, *k* indicates the number of studies in the sample, *j* represents the number of additional covariates, and ω_k is between-study variance. Covariates associated with publication bias may be added to Egger's regression to improve the performance of the model in the presence of between-study heterogeneity (Ioannidis, 2008; Terrin et al., 2003). To evaluate funnel asymmetry using Egger's regression, one inspects the size of β_0, and if it differs significantly from zero. In the absence of publication bias, the regression intercept is expected to be zero. The predictor in this model indicates the precision of a particular study, so the intercept shows the expected effect size when precision is zero (i.e., when the standard error of the study is infinitely large). The expected effect size should be zero when there is no publication bias because extremely large standard errors will result in extremely large confidence intervals. However, when the funnel plot is asymmetric -potentially due to SSE – small studies with high effect sizes will be overrepresented in the data, leading to a high number of significant low-precision studies. As a result, the predicted effect size for zero precision will be larger than zero, resulting in a significant intercept.

These tests suffer from the same limitations that plague the previous tests due to their reliance on SSE to indicate publication bias. If small-study effects exist for a reason other than publication bias, then publication bias will be falsely identified. In addition, Zwetsloot and colleagues (2017) highlighted that Egger's regression test overestimates false positives when applied to standardized mean difference effect size estimates. In a small simulation study, they

showed that due to correlation between the effect size estimate and its standard error, the test has inflated Type I errors.

Selection Models

A lesser-used approach to detecting publication bias is selection modeling (Hedges & Vevea, 1996; Vevea & Woods, 2005) Selection models are conceptually distinct from SSE methods. This approach typically uses weighted distribution to model the selection (i.e., publication) process and develop estimation procedures that account for the selection process (Dear & Begg, 1992; Hedges, 1992; Silliman, 1997; Sutton et al., 2000). The statistical model underlying selection methods consists of two components, a data model and a selection model. The data model describes how the data are generated in the absence of any publication bias, and it is generally chosen to be equivalent to the data models typically employed in education research. The data model, followed by the selection model are represented below.

$$\widehat{g_{\imath k}} = \beta_0 + \beta_j + \omega_k + \varepsilon$$

$$\widehat{g_{\imath k}} = \beta_0 + \beta S_j + \omega_k + \varepsilon$$

The selection model describes the process underlying publication bias, and it can take a wide variety of forms (Hedges, 1992; Vevea & Woods, 2005). For example, it might specify that (a) only studies with results that are statistically significant are published, (b) only studies with results that are statistically significant and directionally consistent are published, or (c) studies with results that are not statistically significant (or directionally consistent) are relatively less likely to be published than studies with results that are statistically significant (and directionally consistent). Vevea and Woods (2005) described the stepwise selection model, in which "cut points" are set to assess publication bias. Because p-values should exist on a continuum outside

of publication bias, this method presupposes that any clustering around traditional cut-points for significance (e.g., .1, .05., .01) may indicate bias in the selection process.

On strength of the selection models is that they can incorporate study level covariates in the same way that any random-effects linear would, so this method allows for between-study heterogeneity without breaking down. However, because the selection process is inherently unknown, researchers must resort to assumptions regarding the study characteristics associated with higher probability of publication. Jin et al., (2015) argue that selection processes in real life are far more complicated than those idealized by selection models. For example, studies seldom have a single effect of interest; typically, studies have multiple dependent effects, and selection is likely to be based on the size, direction, and statistical significance of these multiple dependent effects *jointly*. Further, reviewers and authors may have biases that interact or even contradict their publication biases in ways that do not map cleanly onto selection models. Because the use of weight functions is based on the hypothesized selection process, the validity of their adjusted results may be doubtful if taken at face value (Lin et al., 2018). As a result, these methods are advised to be employed as sensitivity analyses rather than stand-alone tests (McShane et al., 2016).

Current Study

In this study, I will test and adjust for publication bias in a dataset of social-emotional learning interventions (SEL). SEL intervention studies have not yet been assessed for this form of bias yet may be particularly vulnerable given the wide range of potential outcomes that may be included or selected out of a single study. In comparison to reading or math interventions, which typically have few common outcomes of interest (e.g., reading comprehension, math proficiency etc.), a single SEL study may cover a diverse field of outcomes, from drug use, to

attendance, to prosocial behavior (Dermody et al., 2022; Durlak et al., 2011; Zhao, 2020). As such, removing unfavorable outcomes from study report are more easily undetected. This may make outcome-bias more prevalent in SEL.

In this study, I (1) visually inspected the funnel plot for asymmetry, (2) used the trim-and-fill method, (3) conducted Egger's regression (Egger et al., 1997), and (4) conducted Vevea and Woods' step-function selection model (Vevea & Woods, 2005). As recommended by McShane et al., (2016), I employed multiple tests, as no individual test is a perfect assessment of publication bias.

Methods

Data

The data used for this publication bias assessment followed the contemporary Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Page et al., 2021), and the Reporting Standards for Research in Psychology (APA Publications and Communications Working Board Group, 2008). In this section I briefly outline the data collection methods used in the following subsections: inclusion criteria, search strategy, data screening and extraction, and risk for bias. To review the full list of inclusion and exclusion criteria, see the original paper (Cipriano et al., 2023).

Inclusion and Exclusion Criteria

This dataset included studies of SEL interventions in the United States published and unpublished in English from January 1, 2008, through September 1, 2020. The year 2008 was used as the start date to capture all studies available since the SEL field's seminal review paper (see Durlak et al., 2011). SEL interventions in this study were defined as interventions for students that are provided during school hours and within the school setting that target the

development of social and emotional skills among all students in the classroom or school. This dataset included studies examining the elementary school student population (K-5).

Search Strategy

Relevant studies were identified through electronic searches of bibliographic databases. An iterative process was used to translate and refine the searches, and to maximize sensitivity, the formal search used controlled vocabulary terms and synonymous free-text words to capture the concepts of "SEL programs" and "elementary school." The database searches were limited to articles available in English in 2008 and beyond. In addition, the original authors followed the recommendations outlined by Higgins et al., (2019) and manually searched for unpublished studies through several methods. The initial electronic search yielded (n = 9,676) and gray literature review yielded (n = 5,419). The final set of articles was uploaded to Covidence for screening. Following automatic duplicate removal on Covidence, the uploaded sample was 11,018.

Articles were double screened in two phases by four of the authors: first title and abstracts and then full text. Authors reached "almost perfect" interrater reliability for both stages of screening (0.93%–1% for title and abstract, and 0.81%–0.98% agreement for full text; McHugh, 2012), resulting in 236 articles screened for inclusion. Backward and forward hand searching was then conducted. After removing duplicates and double screening, the analytic total reached 174 studies.

Data was coded according to program components, study design, sample size, grade level, outcome measures (e.g., social relationships), form of measurement (e.g., teacher-report) and effect sizes.

Effect Size Calculations

The original authors calculated mean effect sizes across studies after assigning each study a weight based on inverse variance (Lipsey & Wilson, 2001), with the standardized mean difference as the outcome measure.). Mean effect sizes across studies were calculated after assigning each study a weight based on inverse variance (Lipsey & Wilson, 2001), with adjustments for clustered designs suggested by (Hedges, 2007). In combining across studies, the original authors used a random effects model as recommended by (Borenstein et al., 2010, 2021) when there is a belief that there was no single true effect size but a range of effect sizes that may have depended on other factors. To account for effect size dependencies (i.e., multiple effects per study), the authors used robust variance estimation to adjust the standard errors and degrees of freedom for regression coefficients, using the small-sample correction based on the Satterthwaite approximation (Tipton, 2015; Tipton & Pustejovsky, 2015).

Analysis

I used four conventional approaches to identify selective reporting bias: (1) visually inspect the funnel plot for asymmetry, (2) use the trim-and-fill method, (3) conduct Egger's regression significance test, and (4) run Vevea and Woods' step-function selection model. Given the limitations described above, these methods should be viewed as sensitivity analyses rather than definitive estimates of publication bias (Sterne et al., 2011). Each analysis will use the *metafor* package (Viechtbauer, 2010).

Contour-enhanced funnel plots were visually inspected, paying close attention to whether any missing studies are significant or not, with nonsignificant missing studies indicating higher likelihood of publication bias. Contours represented conventional levels of statistical significance (i.e., <0.1, <.05, and <.01).

Next, I conduct a trim-and-fill sensitivity analysis to estimate the number of studies missing from a meta-analysis due to suppression of the most extreme results on one side of the funnel plot Duval and Tweedie (2000; Duval, 2005). The method then augments the observed data so that the funnel plot is more symmetric and recomputes the summary estimate based on the complete data. Three different estimators for the number of missing studies were proposed by Duval and Tweedie (2000). Here, I used "R_0" because it provides a test of the null hypothesis that the number of missing studies is zero. The *trimfill* function in the *metafor* package was used to conduct this analysis. Then, I conducted Egger's significance test using the *regtest* function (Viechtbauer, 2010) to evaluate significance of the intercept.

Last, selection models were employed to identify and adjust for any underlying selection bias. I conducted the step-function Vevea and Woods selection model (1995) using the *selmodel* function, with cut-points set at .1, .05, .01 (Viechtbauer, 2010).

Results

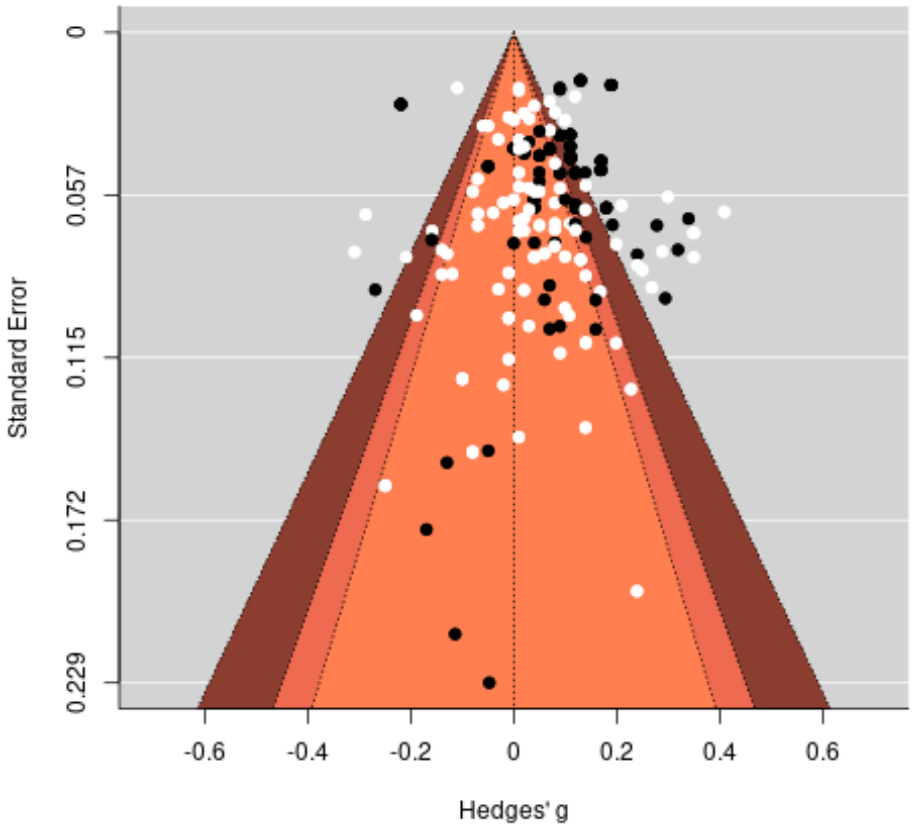
A total of $k = 174$ studies were included in the analysis. The observed outcomes ranged from -0.3092 to 0.4088, with the majority of estimates being positive (74%). The estimated average outcome based on the random-effects model was Hedges $\mathrm{g} = 0.06$ (95% CI: 0.05 to 0.08). Therefore, the average outcome differed significantly from zero ($z = 10.8602$, $p < 0.0001$). Published outcomes were slightly higher than non-published outcomes (g = .07; g = .04). A forest plot showing the observed outcomes and the estimate based on the random-effects model can be found in Appendix A.

According to the Q-test, the true outcomes appear to be heterogeneous ($Q(395) = 2387.86$, $p < 0.0001$, $\hat{\tau}^2 = 0.0095$, $I^2 = 82.8982\%$). A 95% prediction interval for the true

outcomes is given by -0.1277 to 0.2549. Hence, although the average outcome is estimated to be positive, in some studies the true outcome may in fact be negative. An examination of the studentized residuals revealed that none of the studies had a value larger than ± 3.8336 and hence there was no indication of outliers in the context of this model. A funnel plot of the estimates is shown in Figure 2.

Figure 2

Contour-Enhanced Funnel Plot



Note. Published studies represented by white dots, unpublished studies represented by black dots represent unpublished studies. Shaded layers indicate significance levels (light orange = $p < .1$, orange = $p < 05$, dark orange = $p < .01$).

Neither the rank correlation nor the regression test indicated any funnel plot asymmetry ($p = 0.3143$ and $p = 0.3965$, respectively). The funnel plot visualization indicates some evidence of publication bias, as demonstrated from the three unpublished studies (black dots) on the lower, left side of the funnel. The trim-and-fill method indicated no change in included studies, and can be found in Appendix B.

Step-function selection models assessed the potential for selection based on whether the outcome was just under key significance levels (.1, .05, .01; Vevea & Hedges). I found evidence of publication bias at the .01 significance level ($\beta = 0.73$, p<.0001, [95% CI: 0.03, 0.07]), the .05 level ($\beta = 0.49, p < .0001$, [95% CI: 0.22, 0.76]), and the .1 level ($\beta = 0.66, p < .0001$, [95% CI: 0.37, 0.96]). Therefore, the results indicate that while there is no small sample bias, there is publication bias according to conventional levels of significance. The overall effect size was adjusted according to the step-function weights; the adjusted estimate was slightly smaller (Hedges $g = 0.05$, [95% CI: 0.03, 0.07]).

Discussion

In this study I examined the impact that publication bias may have on SEL interventions. For this analysis, I employed a dataset which included published and unpublished SEL intervention studies from 2008-present in elementary schools. Because various publication bias methods complement one others' limitation, multiple tests were conducted (i.e., funnel plot, trim-and-fill, Egger's regression, and the step-function selection model). The results indicated that there is no publication bias as a result of SSE, but there is publication bias from clustering around key significance levels *(p <.1, .05, .01).* This finding points to selective outcome repression, potentially following p-hacking of results. Because small studies are just as likely to

be published as larger ones in this dataset, there is less evidence of the file-drawer problem (Franco et al., 2014; Tipton & Pustejovsky, 2015)

This study highlights the value of using both theoretical approaches to publication bias. Traditionally, SSE methods have been preferred – if publication bias is assessed at all (Gage et al., 2017). In a review of special education meta-analyses, Gage et al., (2017) found that only 33% (k = 36) described procedures to assess publication bias. Of the methods employed, 17 used the fail-safe *N,* a faulty method that has been widely discredited by the statistical literature (Carson et al., 1990; Heene, 2010). The remaining studies used methods based on SSE. In any one of the meta-analyses included in this review, publication bias as the result of significance clustering would have gone undetected.

Meta-analysis has made great strides in evolving publication bias methodology (Marks-Anglin & Chen, 2020). Since the earliest attempts to quantify the impact of selective publishing, publication bias methods have become more sensitive and flexible. From the first iteration of these methods in the form of the fail-safe n, to the more recent graph-based tools, and small-effects or selection models, publication bias analyses are constantly modified to incorporate increasingly complex assumptions (Sterne & Egger, 2005, Copas & Shi, 2000). Yet with this increasing complexity, the use of these methods has dropped considerably (Marks-Anglin, 2020). Given the impact of publication bias, use of multiple methods should become a staple of the field. Broader and simplified trainings of these methods should be disseminated throughout the meta-analytic community, to minimize the barrier to entry of these methods. The evidence-based movement rests on researchers' reliable assessment of intervention effects. Routine use of publication bias assessment in meta-analysis is key to producing these reliable assessments.

References

Becker, B. J. (2005). Failsafe N or file-drawer number. *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 111-125.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 1088–1101.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

Carson, K. P., Schriesheim, C. A., & Kinicki, A. J. (1990). The usefulness of the" fail-safe" statistic in meta-analysis. *Educational and Psychological Measurement*, *50*(2), 233–243.

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292.

Chow, J. C., & Ekholm, E. (2018). Do published studies yield larger effect sizes than unpublished studies in education and special education? A meta-review. *Educational Psychology Review*, *30*, 727–744.

Cipriano, C., Naples, L. H., Eveleigh, A., Cook, A., Funaro, M., Cassidy, C., McCarthy, M. F., & Rappolt-Schlichtmann, G. (2023). A systematic review of student disability and race representation in universal school-based social and emotional learning interventions for elementary school students. *Review of Educational Research*, *93*(1), 73–102.

Copas, J., & Shi, J. Q. (2000). Meta-analysis, funnel plots and sensitivity analysis. Biostatistics, 1(3), 247-262.

Coyne, J. C., & Kok, R. N. (2014). Salvaging psychotherapy research: A manifesto. *Journal of Evidence-Based Psychotherapies*, *14*(2).

Dear, K. B., & Begg, C. B. (1992). An approach for assessing publication bias prior to performing a meta-analysis. *Statistical Science*, 237–245.

Dermody, C. M., Dusenbury, L., Greenberg, M., Godek, D., Connor, P., Cross, R., Martinez-Black, T., Solberg, S., Kroyer-Kubicek, R., & Atwell, M. (2022). A Developmental Framework for the Integration of Social and Emotional Learning and Career and Workforce Development. *Collaborative for Academic, Social, and Emotional Learning*.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, *82*(1), 405–432.

Duval, S. (2005). The trim and fill method. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 127–144.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, *315*(7109), 629–634.

Forness, S. R. (2001). Special education and related services: What have we learned from meta-analysis? *Exceptionality*, *9*(4), 185–197.

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505.

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication Bias in Special Education Meta-Analyses. *Exceptional Children*, *83*(4), 428–445. https://doi.org/10.1177/0014402917691016

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79–89.

Ginsberg, Y. C., Hollands, F. M., Holmes, V. R., Shand, R., Evans, P., Blodgett, R., Wang, Y., & Head, L. (2022). Does ESSA Assure the Use of Evidence-based Educational Practices? *Educational Policy*, 08959048221127989. https://doi.org/10.1177/08959048221127989

Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in research. Beverly Hills*. CA.: Sage Publications.

Group, J. (n.d.). APA Publications and Communications Board Working Group on Journal Article Reporting Standards (2008). Reporting standards for research in psychology. Why do we need them? What might they be. *American Psychologist*, *63*(9), 839–851.

Hedges, L. V. (1992). Modeling publication selection effects in meta-analysis. *Statistical Science*, *7*(2), 246–255.

Hedges, L. V. (2007). *Effect sizes in cluster-randomized designs*.

Hedges, L. V., & Vevea, J. L. (1996). Estimating effect size under publication bias: Small sample properties and robustness of a random effects selection model. *Journal of Educational and Behavioral Statistics*, *21*(4), 299–332.

Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486.

Heene, M. (2010). *A brief history of the Fail Safe Number in Applied Research* (arXiv:1010.2326). arXiv. http://arxiv.org/abs/1010.2326

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.

IntHout, J., Ioannidis, J. P., Borm, G. F., & Goeman, J. J. (2015). Small studies are more heterogeneous than large ones: A meta-meta-analysis. *Journal of Clinical Epidemiology*, *68*(8), 860–869.

Ioannidis, J. P. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice*, *14*(5), 951–957.

Jin, Z.-C., Zhou, X.-H., & He, J. (2015). Statistical methods for dealing with publication bias in meta-analysis. *Statistics in Medicine*, *34*(2), 343–360.

Light, R. J., & Pillemer, D. B. (1984). Quantitative procedures. *The Science of Reviewing Research*.

Lin, L., Chu, H., Murad, M. H., Hong, C., Qu, Z., Cole, S. R., & Chen, Y. (2018). Empirical comparison of publication bias tests in meta-analysis. *Journal of General Internal Medicine*, *33*, 1260–1267.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* SAGE publications, Inc.

Maier, M., VanderWeele, T. J., & Mathur, M. B. (2022). Using selection models to assess sensitivity to publication bias: A tutorial and call for more routine use. *Campbell Systematic Reviews*, *18*(3), e1256.

Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods*, *11*(6), 725–742.

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, *11*(5), 730–749.

Møller, A. P., & Jennions, M. D. (2001). Testing and adjusting for publication bias. *Trends in Ecology & Evolution*, *16*(10), 580–586.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906.

Palmer, T. M., Sutton, A. J., Peters, J. L., & Moreno, S. G. (2008). Contour-enhanced funnel plots for meta-analysis. *The STATA Journal*, *8*(2), 242–254.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., Rushton, L., & Moreno, S. G. (2010). Assessing publication bias in meta-analyses in the presence of between-study heterogeneity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *173*(3), 575–591.

Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, *42*(8), 424–432.

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90–100.

Polanin, J. R., & Espelage, D. L. (n.d.). *The ultimate haystack needle: Locating unreported outcome data for use in a systematic review and meta-analysis Objective*.

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 1–7.

Rosenthal, R. The "file drawer problem" and tolerance for null results. Psychological Bulletin, 1979, &5, 638-641.

Sedgwick, P., & Marston, L. (2015). How to read a funnel plot in a meta-analysis. *Bmj*, *351*.

Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: Practical guidelines and recommendations based on a large database of meta-analyses. *Medicine*, *98*(23).

Silliman, N. P. (1997). Hierarchical selection models with applications in meta-analysis. *Journal of the American Statistical Association*, *92*(439), 926–936.

Sterne, J. A., Becker, B. J., & Egger, M. (2005). The funnel plot. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 73–98.

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, 99–110.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., & Schmid, C. H. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj*, *343*.

Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). Modelling publication bias in meta-analysis: A review. *Statistical Methods in Medical Research*, *9*(5), 421–445.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine*, *22*(13), 2113–2126.

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods*, *20*(3), 375.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods*, *10*(4), 428.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

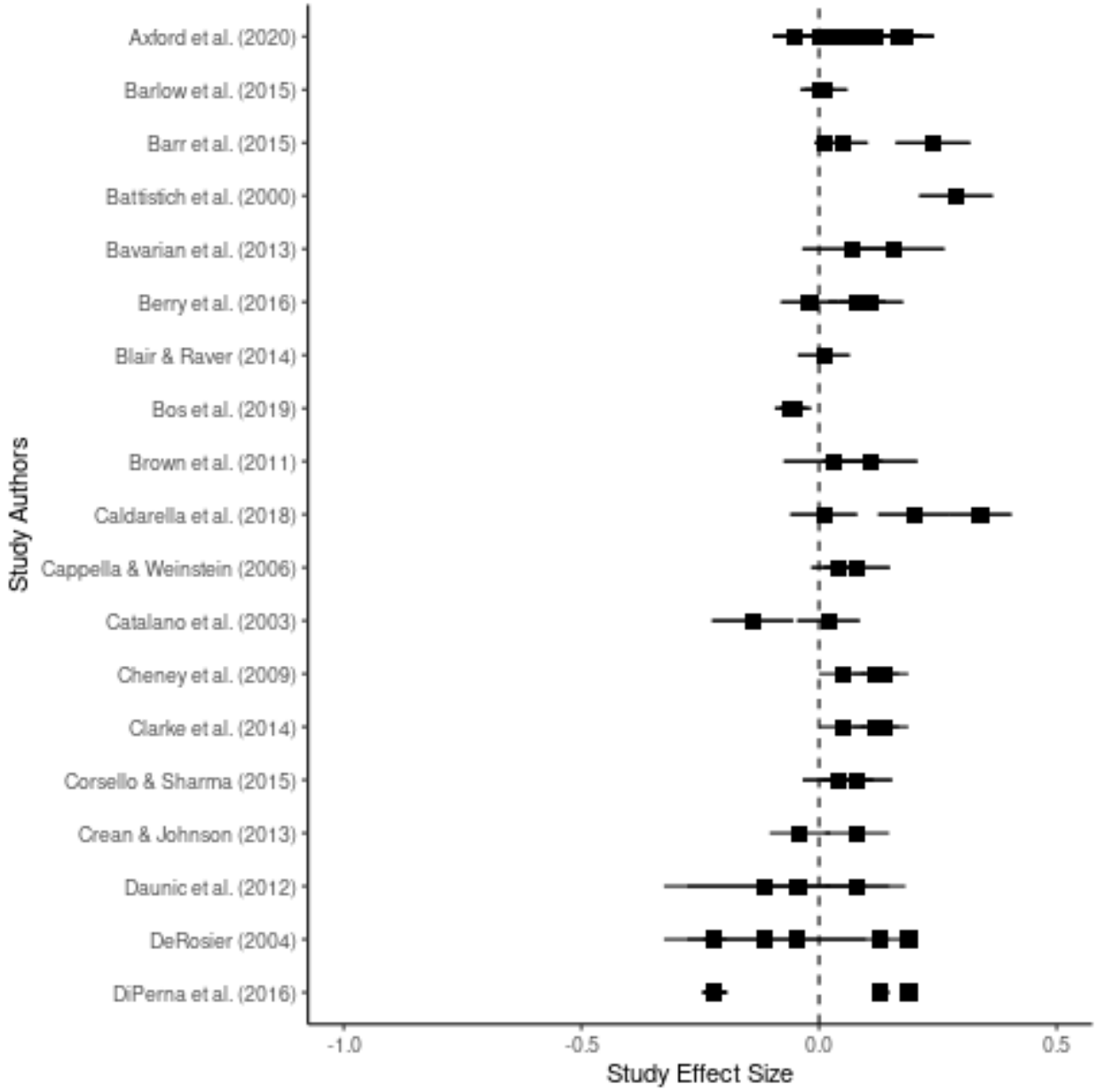
Yoshizawa, L. (2020). *Research use under pressure: State and district implementation of the ESSA evidence requirements* [PhD Thesis]. Harvard University.

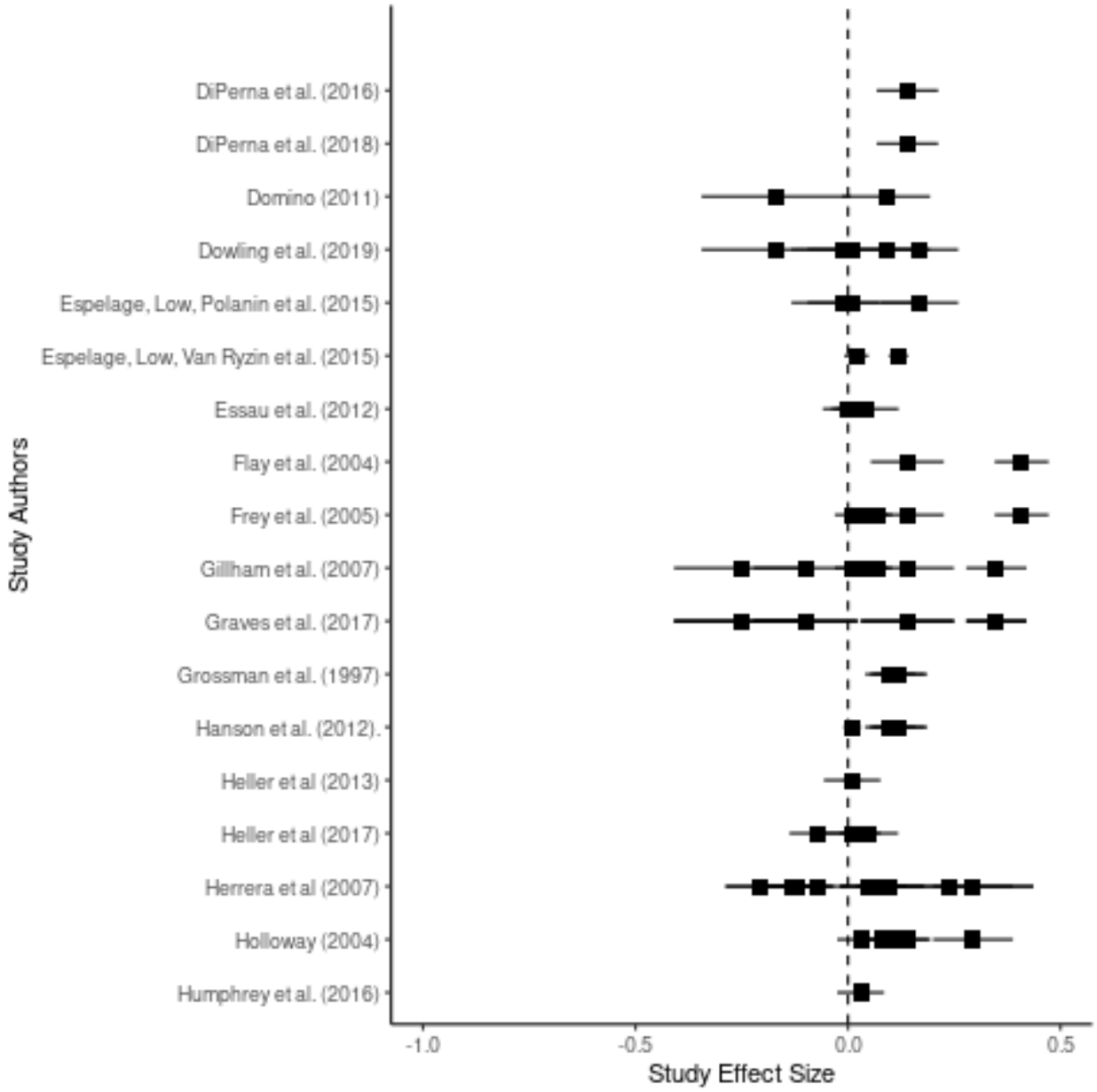
Zhao, Y. (2020). Another education war? The coming debates over social and emotional learning. *Phi Delta Kappan*, *101*(8), 42–48.

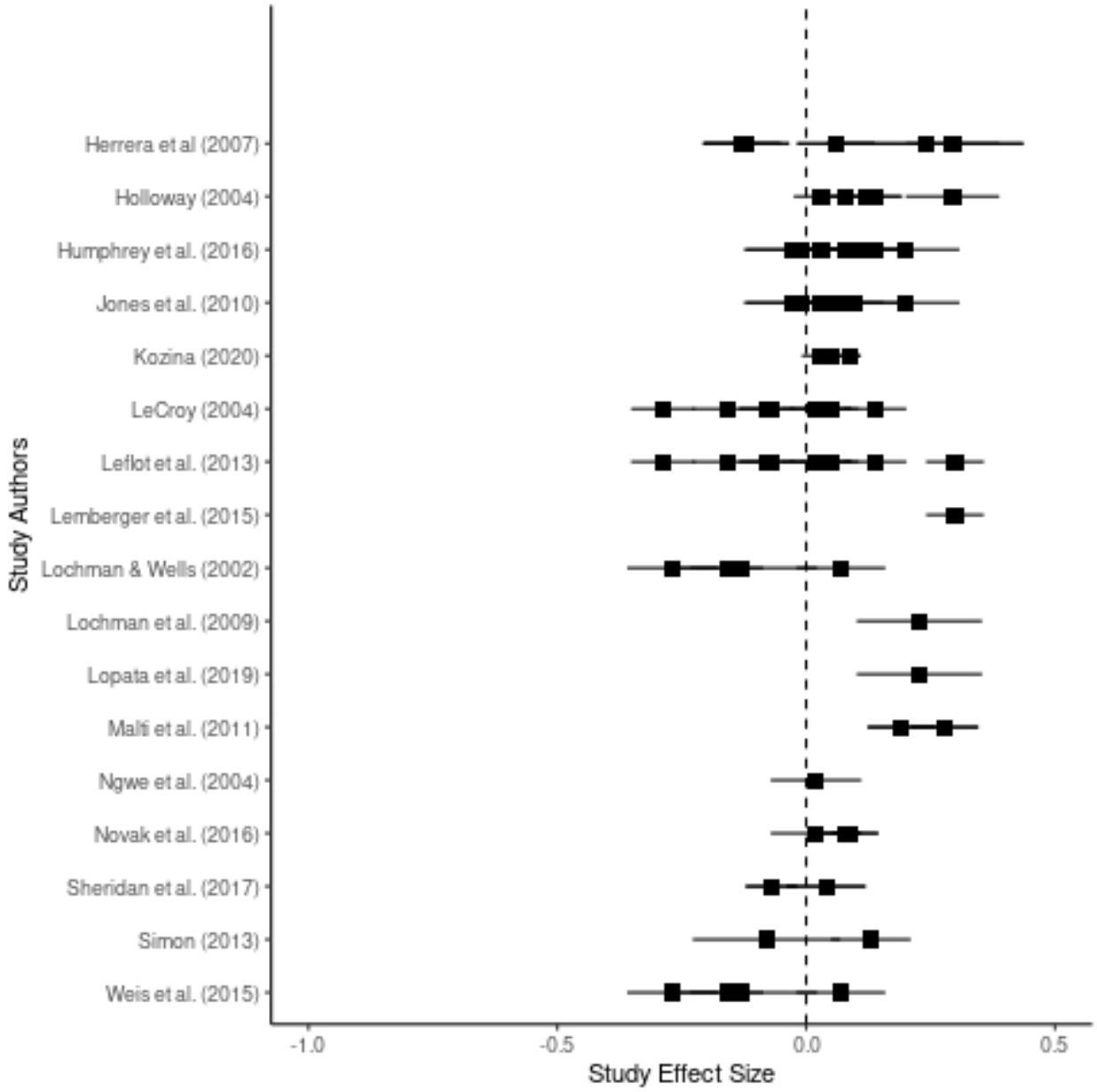
Zwetsloot, P.-P., Van Der Naald, M., Sena, E. S., Howells, D. W., IntHout, J., De Groot, J. A., Chamuleau, S. A., MacLeod, M. R., & Wever, K. E. (2017). Standardized mean differences cause funnel plot distortion in publication bias assessments. *Elife*, *6*, e24260.

Appendix A

Forest Plot of Included Studies



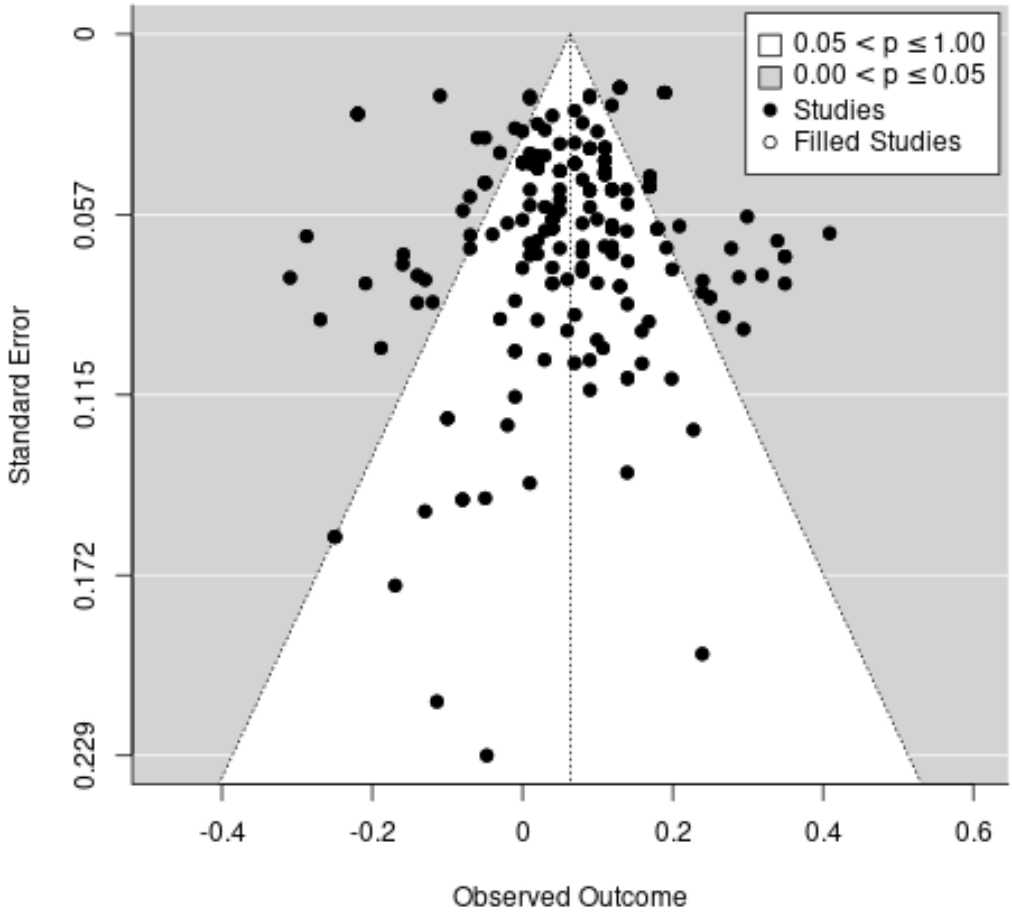

DiPerna et al. (2016)
DiPerna et al. (2018)
Domino (2011)
Dowling et al. (2019)
Espelage, Low, Polanin et al. (2015)
Espelage, Low, Van Ryzin et al. (2015)
Essau et al. (2012)
Flay et al. (2004)
Frey et al. (2005)
Gillham et al. (2007)
Graves et al. (2017)
Grossman et al. (1997)
Hanson et al. (2012).
Heller et al (2013)
Heller et al (2017)
Herrera et al (2007)
Holloway (2004)
Humphrey et al. (2016)
Study Authors
-1.0
-0.5
0.0
0.5
Study Effect Size


Herrera et al (2007)
Holloway (2004)
Humphrey et al. (2016)
Jones et al. (2010)
Kozina (2020)
LeCroy (2004)
Leflot et al. (2013)
Lemberger et al. (2015)
Lochman & Wells (2002)
Lochman et al. (2009)
Lopata et al. (2019)
Malti et al. (2011)
Ngwe et al. (2004)
Novak et al. (2016)
Sheridan et al. (2017)
Simon (2013)
Weis et al. (2015)
Study Authors
-1.0
-0.5
0.0
0.5
Study Effect Size

Appendix B

Figure 3

Trim-and-Fill Funnel Plot



Chapter 5: Conclusions

Despite over two decades of steady improvement in research on what works in education policy and practice, the evidence on *how* best to deploy these findings is still very weak. While anecdotal data and professional expertise are inherently valuable – use of evidence *and* professional expertise are not mutually exclusive. On the contrary, by relying on high quality evidence for decisions where the data exists, education practitioners can be freed to refocus their mental energy to other day-to-day decisions.

Education practitioners and policymaker have the opportunity to drastically improve student outcomes if they rely on high-quality evidence to guide their decision-making. However, there are currently many reasons that these stakeholders may understandably hesitate to use research evidence. Ironically, the evidence-based reform movement has no strong evidence around which research is used, when, and under what conditions. First, this collection of studies examined the current uses of research in K-12 leadership decision-making and outlines key obstacles to URE. This qualitative review should be used to guide future experiments around improving use of evidence in education. In my second study, I developed one such intervention derived from qualitative findings. This study evaluated the impact of one instructional strategy on teachers' understanding and use of research. However, the value of such interventions is questionable if the evidence used is flawed. Meta-analysis is the backbone of the evidence-based reform movement, and most commonly used by education leadership in forming policy. In my third study, I identify and adjust for publication bias in social-emotional learning interventions. Meta-analysts must include publication bias analyses as part of their regular best practices – and reviewers must begin to expect them. In particular, authors should include publication biases methodologies that rely on varying assumptions of the cause of such bias. These studies mapped

the field of research use in its current state as well as take some first steps in resolving its key pitfalls. As an evolving area of study, researchers should address these shortcomings by examining creative solutions to improving use of evidence in education.

Evolving the Field of Research Use in Education

The field of research use in education is currently under-researched and not well understood. Yet, given the value placed on evidence-based reform, evolving the methods and questions asked to better suit causal claims is necessary and urgent. For decades there have been calls by concerned stakeholders to improve the quality of education research in math, literacy, and social-emotional subject areas (NCLB, 2002; ESSA, 2015). The field of education has risen to the challenge, with improved study designs investigating instructional strategies across many subject areas (Hedges, 2018; Slavin, 2002). However, there has been no parallel improvement in secure knowledge about how best to get that evidence into use. The research presented here emphasized the absence of studies using the appropriate design and quality needed to make robust causal claims about evidence-into-use.

A good place to start would building up this base of studies would be to test interventions to change users' knowledge and behavior. As is emphasized in the systematic review included here, simply providing greater access to research evidence is not an effective manner to improve research use. Instead, interviews and surveys of education stakeholders indicate that ongoing, trusting relationships with researchers or intermediaries are much more likely to be effective. Interpersonal relationships and social contexts are considered key to bridging the research-practice divide. Testing related strategies within well-implemented randomized-controlled trials or quasi-experiments will be an important next step in both building greater confidence in these strategies' efficacy, and uncovering hidden nuances of how, when, and with whom they tend to

work best. For example, setting up relationships between school districts and librarians in searching an evidence database may be a much stronger strategy than simply providing the database for districts to navigate themselves. Yet this investigation may lead to other findings regarding the forms and frequency of communication between librarians and district representative that is most effective. Furthermore, it is possible that a central office employee may not be the best conduit for research information, but rather a teacher with high social capital among their employees. These and other questions cannot be answered as conclusively within descriptive research.

URE researchers should also borrow instructional and motivational strategies that have proven effective at increasing use of knowledge in other populations and assess the efficacy of such strategies with research users. While imperfect, the randomized-controlled trial detailed in this volume provides a good illustration of how such strategies might be evaluated to build upon understanding of improving research use in education.

Another critical area of growth concerns improving the reliability of results coming out of the field of education. The reputation of education research has been damaged by disappointing replications of promising results (Plucker & Makel, 2021; Wiliam, 2022). When this happens, users become more hesitant to depend on research in policy and practice decisions (Canfield-Davis & Jain, 2010; Jabbar et al., 2014). As further demonstrated in this thesis, one of the main threats to replication in education comes as a result of publication bias. Yet increasing use of publication bias methodology is not the only way that researchers can improving replication. Greater transparency through open science practices should be used in primary research. Not only will adoption of these practice increase trust in the results, but it also

facilitates replication studies, which might otherwise fail due to erroneous implementation (Gehlbach & Robinson, 2018, 2021; Makel & Plucker, 2014).

In my own work I hope to contribute to each of these areas, ultimately towards the abiding goal of improving outcomes for students. A mentor of mine once observed,

> "*None of us will live to see our structure completed, because education keeps growing in techniques and capability. But it's useful to stop from time to time and remember why we do what we do, and for whom." (Dr. Robert Slavin, 2021).*

As a mother and a former teacher, I am reminded of this quote often. As is often mentioned in interviews with practitioners, education research can easily become abstract and unattached from students. Moving forward in my career, when I choose what and how I conduct my research, I will strive to keep these students top of mind.

References

Canfield-Davis, K., & Jain, S. (2010). Legislative Decision-Making on Education Issues: A Qualitative Study. *Qualitative Report*, *15*(3), 600-629.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296-315.

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79-89.

Hedges, L. V. (2018). Challenges in building usable knowledge in education. *Journal of Research on Educational Effectiveness*, *11*(1), 1-21.

Jabbar, H., La Londe, P. G., Debray, E., Scott, J., & Lubienski, C. (2014). How Policymakers Define 'Evidence': the politics of research use in New Orleans. *Policy Futures in Education*, *12*(8), 1013-1027.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, *43*(6), 304-316.

No Child Left Behind Act of 2001, P.L. 107-110, 20 U.S.C. § 6319 (2002).

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90-100.

Robert Slavin's Blog. (2021, April 29). *What Bob Slavin would want us to remember*. Robert Slavin's Blog. Retrieved March 8, 2023, from https://robertslavinsblog.wordpress.com/2021/04/29/what-bob-slavin-would-want-us-to-remember/

Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational researcher*, *31*(7), 15-21.

Wiliam, D. (2022). How should educational research respond to the replication "crisis" in the social sciences? Reflections on the papers in the Special Issue. *Educational Research and Evaluation*, *27*(1-2), 208-214.