Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Polanin, Joshua R.	American Institutes for Research	0000-0001-5100-0164
Zhang, Qi	American Institutes for Research	
Taylor, Joseph	American Institutes for Research	0000-0002-3753-4888
Williams, Ryan T.	American Institutes for Research	0000-0002-5377-7492
Joshi, Megha	American Institutes for Research	0000-0001-7936-076X
Burr, Lauren	American Institutes for Research	

Publication/Completion Date—(if *In Press,* enter year accepted or completed) 2022

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of Research on Educational Effectiveness

DOI or URL to published work (if available) https://doi.org/10.1080/19345747.2022.2139312

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** Institute of Education Scie through **[Grant number]** R305A1801 to **Institution]** Development S.The opinions expressed are those of the authors and do not represent views of the **[Office name]** Institute of Education Scie or the U.S. Department of Education.

Evidence Gap Maps in Education Research

Joshua R. Polanin, Ph.D.

Qi Zhang, Ph.D.

Joe Taylor, Ph.D.

Ryan Williams, Ph.D.

Megha Joshi, Ph.D.

Lauren Burr, M.P.P.

American Institutes for Research

Author note: Joshua R. Polanin, Ph.D., Principal Researcher, American Institutes for Research, 1000 Thomas Jefferson St. NW, Washington, DC, 20007. Phone: 202-403-5509. Email: jpolanin@air.org. This project was supported by Award No. R305A180102, awarded by the Institute of Education Sciences, U.S. Department of Education and the National Science Foundation (1937412); additional support was received through the Methods of Synthesis and Integration Center (MOSAIC) at AIR. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Education.

Acknowledgements: The authors greatly thank Robin LaSota, Laura Perna, and Megan Austin for their thoughtful critique and helpful insights on earlier versions of the manuscript.

Abstract

Systematic reviews and meta-analyses are important techniques because they synthesize results from multiple primary studies on a similar topic. To influence policy, practice, and research, however, synthesis researchers must translate the results for various audiences. Ideally, the translation drives future research agendas, informs policymaking, or assists in practical decision-making. An Evidence Gap Map (EGM), a graphical or tabular visualization of systematic review and meta-analysis results, is one ideal translation technique because it provides a structured framework to assess contexts for which primary evidence is available or to determine whether the effectiveness of an intervention or a program differs across populations, conditions, and settings. To bolster the field and promote the use of EGMs, we provide an overview of what constitutes an informative EGM, detail multiple examples of EGMs using extant meta-analytic results, and present a free R Shiny application we created to easily generate EGMs from typical meta-analytic datasets. We conclude by reviewing education-based systematic reviews that included an EGM to describe the current state of the field.

Evidence Gap Maps in Education Research

Systematic reviews and meta-analyses[1] synthesize results from multiple primary studies on a similar topic. A systematic review's results can inform our understanding of a program, policy, or practices' overall effectiveness as well as indicate how effects vary by certain study or sample characteristics. It is no surprise, therefore, that use of systematic reviews in education, the social sciences, and sciences broadly has risen steadily over the past twenty years (PubMed, 2020) and that systematic review results help to guide decision-making (U.S. Department of Education, 2020).

To continue to be used and useful, however, synthesists should consider alternative forms of presentations of results and explanations of heterogeneity in effects. Synthesists should translate their work for diverse audiences like education practitioners or administrators, instructional leaders, and government-level policy-makers (Farley-Ripple, May, Karpyn, Tilley, & McDonough, 2018). The translation should allow decision-makers to understand how program effectiveness varies for whom, under what conditions, and in which settings, while conveying those same findings to practitioners and lay audiences in an intuitive manner.

An Evidence Gap Map[2] (EGM), incorporated as a part of systematic review, is an ideal translation and presentation technique because it can provide audiences with an understanding of the state of the field's literature in an intuitive fashion. It can elucidate where funders need to support more research, where researchers should focus their attention, and where practitioners should look for research given their contexts. An EGM is a visual or tabular representation of

---

[1]We refer to "systematic review" as the process of conducting a comprehensive evidence synthesis that does not include the quantitative synthesis of effect sizes. Some EGM authors do not intend to conduct a meta-analysis. We use the term "meta-analysis" only to refer to the occasions when EGM authors intend to synthesis effect sizes. In general, to reduce reader burden, we use the general term "systematic review".

[2]Some authors and institutions refer to EGMs as Evidence and Gap Maps. In practice, we have found no substantive difference between Evidence and Gap Maps versus Evidence Gap Maps (without the "and"). We adopt the more parsimonious phrasing for this manuscript.

systematic review results that display the extent of research in a grid-like format; for graphic representations, the format is an intersection of *x*- and *y*- axes, and for tabular representations, the format is the intersection of rows and columns. When an EGM is represented graphically, a cell marker appears within each cell; the marker represents the number of studies, average effect size, or sample size of a particular cross-tabulation. If a 3rd dimension is required, for example, sample or design characteristics, then the cell markers may also reflect the range of those characteristics[3]. On the other hand, an empty cell indicates where research has *not* been conducted – a *gap* in the literature. Figure 1 is a simple example of a graphical EGM and part of our demonstration explained below.

Synthesists produce EGMs primarily for two reasons. In one circumstance, the researcher's primary goal is to produce a descriptive scan of the literature, often because a funder or policy-maker wants to understand the landscape of evidence. The intent is purely descriptive and therefore, the EGM presents no quantitative synthesis of program effectiveness. In a second circumstance, the researcher's goal is to quantitatively synthesize effect sizes using meta-analysis and attempt to explain variation among the effects. This second circumstance allows synthesists to present average effectiveness for each EGM cell. Both circumstances warrant an EGM, and each has the potential to guide evidence-based decision-making by, for example, elucidating future research priorities or highlighting promising practices under given contexts. In either circumstance, however, the EGM's results and decision-making utility are only as good as the quality of the systematic review from which the results derive.

EGMs were formalized less than a decade ago and have been used primarily by synthesists examining the effectiveness of international development programming. Snilsveit and

[3] For tabular representations, the information is simply conveyed numerically.

colleagues (2016) published the first methodological article on the topic, which outlined the technique, usage, and best practices for future use. Their colleagues at the *International Initiative for Impact Evaluation* (3ie) published the first branded EGM on the effectiveness of peacebuilding efforts in fragile and conflict-affected situations (Cameron et al., 2015). Since the official branding and launch of EGMs, at the time of this publication, 3ie has published 17 EGM articles; they also created a web application that produces online, interactive EGMs using proprietary software (Lacey & Stevenson, 2018). The *Campbell Collaboration*, an international organization dedicated to producing high-quality systematic reviews in the social sciences, has also recently prioritized publication of EGMs (Campbell Collaboration, 2022). White and colleagues (2020) published guidance on how *Campbell Collaboration* authors should produce EGMs, which provides a framework for creating EGMs to be published in the *Campbell Systematic Review* journal.

EGMs offer promise of an effective knowledge translation and decision-making tool but have not yet been solidified as an established part of the systematic review or meta-analysis practice in education research. The purpose of this paper, therefore, is to help researchers understand what EGMs are, what they contribute to programmatic decision-making and funding decisions, and how to conceptualize and build EGMs using systematic review data. First, we identify key tenets of an informative EGM. Second, we describe a free R Shiny application specifically designed to create EGMs. Third, we demonstrate the use of the application, in conjunction with a recently published meta-analysis (Lazowski & Hulleman, 2016), to highlight examples of informative EGMs. Finally, we review the education research literature to illustrate what kinds of EGMs synthesists have produced to date and provide some discussion on how EGMs can be used in the future.

What is an Informative Evidence Gap Map?

EGMs have several defining characteristics. An EGM can be represented as a table, graph, or chart where at least two variables, along the horizontal and vertical axes, intersect to create a grid; each intersection of the axes is referred to as a *cell*. The *axes* should represent characteristics of the systematic reviews' inclusion criteria with at least two levels. For example, the x-axis could contain the outcome variable, mathematics and reading achievement, and the y-axis could contain characteristics of the participants, primary- and secondary-aged students. Within the cells are markers, which represent the statistic (or statistics) displayed within the cell. The markers could represent the number of studies, number of effects, the size of the sample, or the meta-analytic average effect size. The cells themselves can also serve as a form of cell marker; for example, a synthesists could shade the cell to represent the number of studies or effects within it. The simplest EGM, therefore, has an x- and y-axis, each with two levels, where the markers delineate the number of studies within the four resulting cells. Although these characteristics define a basic EGM, we seek to advance the field by defining *informative* EGMs. What follows are tenets and characteristics of informative EGMs.

1. EGMs should use the results from a high-quality, comprehensive systematic review.

An EGM is only as good as the systematic review from which the results are derived because one of the main purposes of an EGM is to illustrate contexts in which research has and has not been conducted. The implicit assumption is that all available published and unpublished eligible research is included in the EGM's results. Not finding previously conducted research will result in perceived gaps in the literature. These identified gaps could simply represent an incomplete search and are therefore erroneous. Thus, the quality of an EGM, and the

implications of its results, are directly related to the quality and comprehensiveness of the systematic review.

Numerous resources are available on how to conduct a comprehensive systematic review (Cooper, Hedges, & Valentin, 2019; Pigott & Polanin, 2020). A fuller description of a comprehensive systematic review is outside the scope of this paper, but briefly, several steps are involved: (1) defining a research question, (2) operationalizing inclusion criteria, (3) identifying all applicable search terms, (4) conducting online searches across multiple databases, (5) locating grey literature through reference harvesting, and (6) contacting authors for missing or unpublished studies. A synthesist should also consider publicly publishing a review protocol and transparently reporting each step in the review process, including the publication of data analysis scripts and datasets (Polanin, Hennessey, & Tsuji, 2020).

2. EGMs are most informative when presenting results of broad, heterogeneous systematic reviews.

Systematic reviews vary in size and scope. Some systematic reviews have specific research questions, focusing on very specific interventions for specific populations, with specific focal outcomes. Others cast a broader net to capture variants of interventions for different populations in different settings, and across different outcomes. EGMs are most informative for translating broadly focused systematic reviews with heterogeneous replications, meaning systematic reviews that include multiple types of related interventions or outcomes may not produce a meaningful EGM due to the limited variability in the dimensions available for the axes.

An important consideration at the beginning of the systematic review process is the scope. A review's scope is based on the research questions, availability of resources, previously

conducted reviews, and the consideration of inclusion criteria. We define scope, therefore, as the breadth of the project. Single-topic systematic reviews can vary widely in the included number of studies; for example, a synthesis of bullying prevention programs can yield many dozens of studies while a synthesis of a branded reading intervention would likely yield only a few.

Research questions and inclusion criteria impact the scope of a systematic review and the results of a resulting EGM. For instance, a researcher may limit the scope of a systematic review to only investigate *education* interventions; as a result, intervention type could not be used as an EGM axis because it has less than two levels. To use intervention type as an axis level, the synthesist must expand the scope and include additional intervention types or variations. The scope of a systematic review may be limited by the resources and data available. However, when possible, if a goal is to produce an EGM, then researchers should use a larger scope or a more refined scheme for coding study or intervention characteristics because a review with too small a scope will have limited dimensional information.

3. EGMs should use Cronbach's Units, Treatments, Outcomes, or Settings (UTOS) characteristics, or a similar framework, on their x- and y-axes.

An informative EGM represents UTOS characteristics, or a similar framework, on both the x- and y-axes[4]. The traditional EGM, produced by Snilstveit and colleagues (2016) and referenced by White and colleagues (2020), relied on treatment (or intervention) and outcome variables for the *x*- and *y*-axes. These EGMs continue to be useful for policymakers or funders especially interested in determining where research has and has not been conducted. Any of the UTOS characteristics, or other variables beyond UTOS, could be used as an *x*- or *y*-axis. The "units" or participant characteristics, for example, could be useful in visualizing the types of

[4] We suggest representing EGM results using a graphical format. However, we do not believe a graphical representation is necessarily more informative than a tabular representation.

samples that have been researched within a particular field. Similarly, EGMs could map the setting of studies, perhaps U.S. and non-U.S.-based locations, to visualize the geographic distribution of research. Other possible variables beyond UTOS include: publication status, research design, research quality, covariates (for correlational studies), comparison group or counterfactual type, economic evidence or cost, or original effect size type (e.g., standardized mean-difference versus odds ratio). The most informative combination of variables should address the proposed research question.

After determining which variables to include on the EGM axes, another consideration is how to represent all the relevant information in the cells. As is commonplace in traditional systematic reviews, review teams must decide how to lump or split coded information (Weir, Grimshaw, Mayhew, & Ferguson, 2012). For example, a systematic review of math outcomes could collect effect sizes from algebra, geometry, calculus, or general math achievement. To analyze the findings, the review team has several options: split the findings into four separate analyses by outcome type, lump the specific outcomes into one analysis but keep general math achievement for separate analysis, or lump all outcomes into one analysis. An EGM creator is faced with similar considerations. The value of splitting the UTOS category by at least two categories is demonstrated in Figure 2, which displays two tutoring programs (A and B) and a lumped math outcome variable. Note how a programmatic decision-maker's conclusion about which tutoring program to select for high school could differ significantly based on whether they have access to disaggregated effect size data, such as in Figure 3, which has a math outcome variable with two levels. Our best advice on whether to lump or split UTOS categories is to follow the pre-determined protocol plan, which should articulate the research questions and the audiences the questions will address. If a protocol is not available, then the researchers could

apply lumping or splitting decisions that allow for consistent analyses throughout the paper or product.

4. Evidence markers should convey summary information in an easy to interpret and translatable manner.

The evidence markers, or the representation of evidence and lack thereof within each cell, constitute what is known about the landscape of evidence. Synthesists may choose among several evidence marker options. In some circumstances, the EGM's purpose is to identify the number of studies within a particular literature. Should this be the case, the least complex option – and one that requires little data extraction from primary studies – is simply to represent the number of studies or effects within studies found. The resulting EGM provides the reader with quick information about where research has or has not been conducted. Because of its simplicity, it is easy for many audiences to understand. The marker's simplicity, however, renders the information to a surface-level description that will not support more nuanced conclusions.

The review team should consider an evidence marker information that helps best inform audiences while making it easy for readers to understand the results. The choice of evidence marker should also help to clearly translate the information. For a simple data representation, using total studies or sample size and the size of the evidence marker may help draw attention to where research has - and has not - been conducted. If possible, using color to represent marker data helps to clarify various data, so long as it conforms to standards inclusive of the visually impaired. The examples we present below, and in our Shiny application, utilize an R package specifically designed to create color contrasts that can be viewed by individuals who have difficulty distinguishing between colors. An alternative to color, and one that may be less cost-prohibitive if printed in a journal, is to represent data with gray shading or different shading

variants. Combinations of color, shading, size, or other indicators are endless and true visualization best practices are outside the scope of this paper (see Yau, 2011 for informative examples). Synthesists should ensure that, regardless of the choices, the representation is clear and informative.

Synthesists could also split the cells further and create multiple markers, each representing a level within a UTOS category. 3ie's EGMs, for example, represent methodological design and quality with the markers. Creating and populating further subclassifications of the data can render the EGM difficult to interpret, so synthesists should balance the clarity of the findings with the added complexity.

Another more complex option is to represent the cell's meta-analytic effect size. Representing the average effect size along with the total number of studies, provide audiences with two critical pieces of information: the research that has been conducted and the corresponding findings. With only these two pieces of information, policy-makers or other decision-makers can determine where to fund missing research or further investigate - or implement - a certain category or condition. Should a synthesist decide to estimate the cell's meta-analytic average, we suggest using a continuous color scheme where darker shades denote larger average effect sizes and lighter shades denote smaller average effects.

A final consideration is the statistical comparison of average effectiveness between the cells using hypothesis testing, for example comparing the effectiveness of one intervention type across a variety of outcome types. Similar to traditional meta-analytic moderator analyses, the analyses require a sufficient number of studies per cell; indeed, our review of EGMs found no meta-analyses where moderator analyses had been conducted. We anticipate that, as meta-

analyses increase in study sample size and scope, the ability and desire to conduct moderator analyses between cells will likely increase as well.

Examples of EGMs Using Data from a Published Systematic Review and Meta-Analysis

To help promote the use of EGMs, we constructed an R Shiny application ("app") that will help review teams produce various EGMs. Our Shiny app relies on the statistical program *R* (*R* Core Team, 2020) as well as the *ggplot2* R package (Wickham, 2016) to plot the information. We created an Open Science Framework project page (OSF) to warehouse the reproducible *R* code and datasets (Polanin & Joshi, 2022). The app allows users to upload their dataset, use point-and-click options to summarize data for combinations of factors, and then create an EGM that is downloadable and ready for publication. The app also provides the R Syntax so that users can further modify the graphic using a local R instance.

We used the Shiny app to create example EGMs using the results of meta-analytic data conducted by Lazowski and Hulleman (2016). The authors synthesized the effects of educational interventions grounded in motivational theory on various types of educational outcomes. We chose this study because it included a large number of studies (92 studies), the authors provided clear UTOS characteristics for each included study, the authors provided each extracted effect size and effect size variance, and they did not previously create an EGM. We removed the "multiple perspective" theory because the large number of studies associated with the theory skewed the data and hindered the EGM's pedagogical appeal. For purposes of illustration only, we also truncated the effect sizes to be less than or equal to 1.00, which evenly distributed the figure's colors. We do not consider these modifications best practice and suggest presenting all data retrieved via the systematic review process, regardless of its appearance. Our OSF project page includes the dataset used to produce the example EGMs.

[Insert Table 1 Here]

We used Table 1 as a guide and created progressively more complex EGMs to illustrate various options. We start by illustrating an EGM that we do not consider informative because it uses a single axis and answers a basic question: "How much research has been conducted on educational outcomes?" Figure 4 displays the three large outcome domains that the meta-analytic authors used: (1) self-report, (2) performance indicator, and (3) behavioral indicator. The size of the marker represents the total sample size from the individual studies; the marker text is the number of studies within the cell. In this case, the cell with the largest sample size – behavioral indicator – has the smallest number of studies. Although the information delineated in this EGM might be useful to some readers, we do not consider it optimally informative because the size of the cells hides the distribution of research across the potential axes. We present it here for illustrative purposes only.

Figure 1 (referenced above) represents what we consider the bare minimum information represented for an informative EGM. The y-axis is, again, the three outcome domains, while the x-axis is the intervention type or "theory" as the meta-analytic authors suggested. Four such theoretical domains are represented, (1) mindset oriented, (2) self-perception oriented, (3) value oriented, and (4) goal-oriented. Crossed with the three outcome domains, this EGM has 12 cells. Within the cells, the markers' size represents the number of studies. One can see that the self-report ~ mindset-oriented cell has the most studies (k = 28) while the behavioral indicator ~ goal-oriented cell has none (k = 0).

The next two EGMs create further complex visualizations. Figure 5 splits the intervention theories into 14 categories while keeping the original three outcome domain categories intact; this results in 42 possible outcomes ~ intervention cells. We also displayed the number of studies

using the size of the marker, as well as the marker's text, and added in the average effect size as the marker's color. Markers with a more purple color represent smaller effects while markers with a green-yellow color represent larger effects. An even more complex representation of this data is in Figure 6, where the cells are themselves split by the grade level (elementary, middle, high school, and postsecondary). The marker shape types also represent the grade levels. Should a reader be interested in this third UTOS category, one could look across the cells and observe a difference in the types of participants. For example, the self-report ~ attribution cell has seven studies, but all seven sampled students from postsecondary institutions. The number of intervention ~ outcome ~ sample types, however, renders this EGM difficult to interpret and is not informative.

Our final EGM illustrates how a synthesist might visualize differences from a moderator analysis (Figure 7). We replaced the y-axis outcome domain data with the two included research designs, randomized controlled trials (RCT) and quasi-experimental designs (QED). The x-axis intervention data, splitting the four domains into 14 categories, remained the same as the previous two EGMs. Also similarly, we represent the number of studies as the size and text of the marker, and the average effect size as the markers' colors. Overall, from a standpoint of research design usage, one can see that RCTs appear to be used frequently in this literature; however, a policy-maker or research may be interested in investigating the "interest" intervention as only two QEDs have been conducted. In the very first column of self-determination, effects from QEDs are greater relative to effects from RCTs. This visualization highlights differences possibly found from moderator analyses.

We describe these examples to demonstrate the range of options for presenting meta-analytic findings using EGMs. Depending on the study objectives and intended audience, EGMs

vary in complexity and can be tailored to portray the most pertinent information. The process of selecting axes, markers, and visual complexity is similar to the process of generating other graphical representations of data. Synthesists should approach the process by being thoughtful about the intended messages from the EGM – such as to quickly see trends in program effectiveness or identify research gaps – and plot the relevant data accordingly, with enough information for readers to easily interpret the EGM.

What EGMs Have Been Produced in Education Research?

As a way of providing further EGM illustrations, we sought to identify previously produced EGMs in education research. A full description of our methodological review of the literature can be found in the Supplemental Materials. Briefly, we conducted a database search as well as hand-search of the primary meta-analytic journal in education, *Review of Educational Research*. We then screened the studies for EGM representations, where the table or graphic included two axes and a marker of the evidence within the cell. Eligible studies were coded for descriptive information about their axes and the type of analyses conducted.

Our database search yielded a total of 118 results; we removed duplicates (n = 7), books (n = 6), articles that were not in English (n = 2), and articles that we were not able to retrieve (n = 2). From our hand-search of the *Review of Educational Research*, we reviewed 364 additional articles. The combined search and screening process yielded 19 systematic reviews that included at least one EGM. Some of the systematic reviews produced multiple EGMs, and we therefore coded 41 total EGMs.

Table 2 summarizes the results of the search, screening, and coding. Considering the systematic reviews papers in their entirety, not specific EGMs within a paper, 12 of the 19 (63%) focused on grades PreK-12, 4 focused on professional or higher education (21%), and 5 focused

(26%) on adult education or educators. Across the 41 EGMs within the 19 systematic reviews included, we determined the most commonly used variables on one of the two axes[5]. The EGMs' axes commonly included a treatment variable (35%). Fifteen EGMs included an axis variable related to participants (18%), 13 included study design or methods (16%), 11 included outcomes (13%), 9 included settings (11%), and 5 included extrinsic characteristics or others (6%).

[Insert Table 2 Here]

We observed several key features of EGMs. The majority of the EGMs (90%) showed results in a table format, while only 6 EGMs (15%) depicted the results graphically (2 EGMs included a combined table and graph). All the EGMs reported the frequency of studies numerically within the cells. Four EGMs (10%) represented the number of studies within a cell by the size of the evidence marker (i.e., the larger the marker, the greater the number of studies within that cell). Finally, 8 EGMs (20%) from one study included the cell's meta-analytic effect size. Of the 41 EGMs produced, the largest *x*-*y* axis combination was treatment ~ units (*n* = 8, 20%) and treatment ~ design (*n* = 8, 20%), followed by treatment ~ outcome (*n* = 7, 17%).

The included EGMs each rendered at least one of the elements of an informative EGM, as described above, yet we recommend interested readers examine four articles, detailed below, that used all four elements. Readers interested in an EGM demonstrating the use of research design and sample characteristics should consult Major and colleagues' (2018) systematic review on classroom dialogue and digital technologies. The authors conducted the review with the explicit intent of producing an EGM as the primary product. Joksimović and colleagues (2018) created an EGM using outcome domains and intervention characteristics, investigating the use of Massive Open Online Courses. Hooper and colleagues (2013) produced two EGMs, each with

[5] Our 41 included EGMs have a total of 82 axes because there are 2 axes per EGM. Therefore, the denominator of the percentage is 82 and not 41.

the same y-axis – primary focus of the educational intervention – but with different x-axes – research design and outcomes (referred to as “impact levels”, pg. 13), respectively. Finally, Seidel and colleagues (2007) produced a table-based EGM where the cells represent the meta-analytic average effects.

Summary

Toward the goal of more effectively communicating and translating evidence syntheses to various audiences, we sought to clarify the defining characteristics of an informative EGM, provide examples of informative EGMs, introduce a new R Shiny application to easily create informative EGMs, and describe the characteristics of EGMs produced in the education research literature. We articulated four elements of informative EGMs, each of which should be considered necessary by future synthesists. Our R Shiny application applies each of the four elements by default, so that synthesists can focus on the EGM’s content instead of the details of the design.

The conceptualization of an EGM only emerged within the past decade, so it is no surprise that their use is relatively minimal. It is also of no surprise, therefore, that the EGMs created in the past decade, or previously without the EGM nomenclature, do not include all four elements that we recommend. We made these recommendations based on our understanding of evidence syntheses, our work translating synthesis results, and a basic understanding of data visualization. But more work must be done. Fitzgerald and Tipton’s (2022) research on another meta-analytic visualization plot, the meta-analytic rain cloud, illustrated how synthesists should collect empirical evidence to guide EGMs’ future developments. The authors insights aligned with the basic EGM structure, but more can and should be done in the future to test variations on the basic structure.

What is clear based on our review, and our own use of EGMs, is that EGMs should be used more often and synthesists should also work to create the most informative EGMs to address their research questions. We believe in the usefulness of evidence syntheses, and especially so when the results can be conveyed in different ways that speak to multiple audiences. Synthesists should continue to find new and different ways of delivering on their promise.

Characterizing evidence accumulation for the purposes of driving future research agendas, informing actionable policymaking, or assisting practical decision-making requires the use of all available techniques. EGMs offer one such technique to education researchers and synthesists who strive to translate bodies of evidence for these audiences. However, EGMs in education research are currently underutilized or utilized in ways that do not capitalize on their full potential. Noting the persistent challenges of bridging research to practice and policy, we suggest that part of the solution is more widespread and purposeful use of EGMs in synthesis studies and our R Shiny app is one resource that we offer to support the field in this effort.

References

Studies marked with an asterisk denote reviews with evidence gap maps.

*Anstee, S., Price, A., Young, A., Barnard, K., Coates, B., Fraser, S., & Moran, R. (2011). Developing a matrix to identify and prioritise research recommendations in HIV Prevention. *BMC Public Health*, 11(381), 1-8.

*Berkovich, I., & Eyal, O. (2015). Educational leaders and emotions: An international review of empirical evidence 1992–2012. *Review of Educational Research*, *85*(1), 129–167. https://doi.org/10.3102/0034654314550046

*Buckley, E. J., & White, D. G. (2007). Systematic review of the role of external contributors in school substance use education. *Health Education*, 107(1), 42-62.

Campbell Collaboration. (2022). *Evidence and gap maps (EGMs)*. Retrieved from: https://www.campbellcollaboration.org/evidence-gap-maps.html

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach*. Thousand Oaks, CA: Sage Publications.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis*. New York, NY: Russell Sage Foundation.

*Davis, E. A., Petish, D., & Smithey, J. (2006). Challenges new science teachers face. *Review of Educational Research*, *76*(4), 607–651. https://doi.org/10.3102/00346543076004607

*de Araujo, Z., Roberts, S. A., Willey, C., & Zahner, W. (2018). English learners in K–12 mathematics education: A review of the literature. *Review of Educational Research*, *88*(6), 879–919. https://doi.org/10.3102/0034654318798093

*Edmonds, M. S., Vaughn, S., Wexler, J., Reutebuch, C., Cable, A., Tackett, K. K., & Schnakenberg, J. W. (2009). A synthesis of reading interventions and effects on reading

comprehension outcomes for older struggling readers. *Review of Educational Research, 79*(1), 262–300. https://doi.org/10.3102/0034654308325998Farley-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.

*Frampton, G.K., Harris, P., Cooper, K., Cooper, T., Cleland, J., Jones, J., Shepherd, J., Clegg, A., Graves, N., Welch, K., & Cuthbertson, B.H. (2014). Educational interventions for preventing vascular catheter bloodstream infections in critical care: Evidence map, systematic review and economic evaluation. *Health Technology Assessment*, 18(15), 1-365.

*Hooper, B., King, R., Wood, W., Bilics, A., & Gupta, J. (2013). An international systematic mapping review of educational approaches and teaching methods in occupational therapy. *The British Journal of Occupational Therapy*, 71(1), 9-22.

International Initiative for Impact Evaluation. (2020). *Evidence gap maps.* https://www.3ieimpact.org/evidence-hub/evidence-gap-maps

Jacob, R. T., Doolittle, F., Kemple, J., & Somers, M. A. (2019). A framework for learning from null results. *Educational Researcher*, *48*(9), 580-589.

*Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, *88*(1), 43–86. https://doi.org/10.3102/0034654317740335

Lacey, S. & Stevenson, J. (2018). *3ie evidence gap maps platform automatic importer guide*. International Initiative for Impact Evaluation. Retrieved from: https://www.3ieimpact.org/file/1785/download?token=47I75sp7

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: a meta-analytic review. *Review of Educational Research, 86*(2), 602-640.

*Macaro, E., Handley, Z., & Walter, C. (2012). A systematic review of CALL in English as a second language: focus on primary and secondary education. *Language Teaching*, 45(1), 1-43.

*Major, L., Warwick, P., Rasmussen, I., Ludvigsen, S., & Cook, V. (2018). Classroom dialogue and digital Technologies: a scoping review. *Education and Information Technologies*, 23, 19995-2028.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, *151*(4), 264-269.

Polanin, J. R., & Joshi, M. (2022). Evidence Gap Maps in Education Research [Data set]. Open Science Framework. https://doi.org/10.17605/OSF.IO/WP5RF

PubMed. (2020). *Meta-Analysis*. National Library of Medicine Bethesda (MD). https://www.ncbi.nlm.nih.gov/pubmed/

R Core Team. (2020). *R: a language and environment for statistical computing*. Vienna, Austria. https://www.R-project.org/

*Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: reading interventions for students in grades 4–12, 1914–2014. *Review of Educational Research, 86*(3), 756–800. https://doi.org/10.3102/0034654316652942

*Schneider, R. M., & Plasman, K. (2011). Science teacher learning progressions: a review of science teachers' pedagogical content knowledge development. *Review of Educational Research, 81*(4), 530–565. https://doi.org/10.3102/0034654311423382

*Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: the role of theory and research design in disentangling meta-analysis results. *Review of Educational Research, 77*(4), 454–499. https://doi.org/10.3102/0034654307310317

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

*Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: an analysis of the interaction hypothesis. *Review of Educational Research*, *75*(4), 457–490. https://doi.org/10.3102/00346543075004457

Snilstveit, B., Vojtkova, M., Bhavsar, A., Stevenson, J., & Gaarder, M. (2016). Evidence & gap maps: A tool for promoting evidence informed policy and strategic research agendas. *Journal of clinical epidemiology, 79*, 120-129.

Tipton, E. (2015). Small-sample adjustment for robust variance estimation with meta-regression. *Research Psychological Methods, 20*(3), 375-393.

Tipton, E., Pustejovsky, J. E., & Ahmadi, H. (2019). Current practices in meta-regression in psychology, education, and medicine. *Research synthesis methods, 10*(2), 180-194.

U.S. Department of Education. (2020). *Using the WWC to find ESSA tiers of evidence.* Washington, DC: Institute of Education Sciences, What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/essa

*Webster-Wright, A. (2009). Reframing professional development through understanding authentic professional learning. *Review of Educational Research, 79*(2), 702–739. https://doi.org/10.3102/0034654308330970

Weir, M. C., Grimshaw, J. M., Mayhew, A., & Fergusson, D. (2012). Decisions about lumping vs. splitting of the scope of systematic reviews of complex interventions are not well

justified: a case study in systematic reviews of health care professional reminders. *Journal of Clinical Epidemiology*, *65*(7), 756-763.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag.

*Winkle-Wagner, R. (2015). Having their lives narrowed down? The state of Black women's college success. *Review of Educational Research, 85*(2), 171–204. https://doi.org/10.3102/0034654314551065

*Wozney, Lori Mae. (2011). *A systematic review of instructional interventions to improve school completion: Mapping the evidence (Doctoral dissertation)*. Available from Dissertation Abstracts International Section A: Humanities and Social Sciences.

Yau, N. (2011). *Visualize this: the flowing Data guide to design, visualization, and statistics*. Hoboken, NJ: John Wiley & Sons.

*Zepeda, S. J., & Mayers, R. S. (2006). An analysis of research on block scheduling. *Review of Educational Research, 76*(1), 137–170. https://doi.org/10.3102/00346543076001137