

The Next Generation of Quality Indicators for Group Design Research in Special Education

Exceptional Children
2023, Vol. 89(4) 359-378
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00144029221150801
journals.sagepub.com/home/ecx



Jessica R. Toste¹ , Jessica A. R. Logan²,
Karrie A. Shogren³ , and Brian A. Boyd⁴

Abstract

Group design research studies can provide evidence to draw conclusions about *what works*, *for whom*, and *under what conditions* in special education. The quality indicators introduced by Gersten and colleagues (2005) have contributed to increased rigor in group design research, which has provided substantial evidence about what works across various content areas. However, there is a need to more fully understand the range of effects within studies and individual differences in treatment response. In this article, we identify contemporary considerations for group design research in special education. First, we propose an expanded set of quality indicators that broaden our lens of inquiry to promote understanding of for whom and under what conditions interventions, programs, and practices are more or less effective. Next, we introduce new quality indicators to further methodological rigor in using open science practices. We recommend that researchers consider the full set of quality indicators and document their decision-making related to the design, implementation, and analysis of group design research. These efforts can generate new scientific knowledge with the potential to advance equity and inclusion of students with disabilities.

Advancing and implementing best practices in the field of special education should be based on an accumulation of evidence derived from rigorous, high-quality research. Group design experimental and quasiexperimental research studies can provide evidence to draw conclusions about *what works*, *for whom*, and *under what conditions*. In 2005, Gersten and colleagues proposed a set of quality indicators for group design research. The purpose of these indicators was to inform and enhance rigorous, scientific educational research and served to expand on the Study Design and Implementation Assessment Device (Valentine & Cooper, 2003) published by the What Works Clearinghouse (WWC). Whereas the WWC guidelines were meant to support evaluation of a completed research report, the quality indicators Gersten et al. (2005) proposed were meant to be used “not only to evaluate the merits of a completed

research report or article, but also to evaluate research proposals, dissertation proposals, and grant applications submitted to funding agencies” (p. 150).

In 2019, the Institute of Education Sciences (IES) introduced the Standards for Excellence in Education Research (SEER) to emphasize “additional factors that can make research transformational” (para. 1). Taken together, researchers have used the Gersten et al. (2005) quality

¹The University of Texas at Austin

²Vanderbilt University

³University of Kansas

⁴University of North Carolina at Chapel Hill

Corresponding Author:

Jessica R. Toste, 1 University Station, D5300, Austin, TX 78712, USA.

Email: jrtoste@austin.utexas.edu

indicators as explicit guidance for the application of SEER, particularly within the unique context of special education research. The purpose of this article is to respond to advancements in the field and introduce a set of guiding questions to increase the purposeful design, implementation, analysis, and reporting of group design research in special education with the ultimate goal of improving equitable educational outcomes for all students with disabilities. Building from Gersten et al. (2005) and SEER, we propose an expanded set of quality indicators to inform group design research.

Contributions of Group Design Research

Group design research is one of many methods that can be employed to address research questions in special education. Although it is essential that multiple methods are considered to address the complex research questions in special education, the focus of the present article is to propose an expanded set of quality indicators specific to group design research. Since the first set of quality indicators for group design research was published in 2005, special education researchers have increased methodological rigor in conceptualizing, designing, conducting, and reporting special education intervention studies.

The first set of quality indicators (Gersten et al., 2005) included, for research proposals, 14 essential indicators—summarized as conceptualization underlying the study, participants and sampling, intervention implementation and nature of control condition, outcome measures, and data analysis—and an additional seven desirable indicators. For research reports, the authors proposed a similar set of 10 essential and eight desirable quality indicators. These group design indicators were developed to ensure a high-quality inference about the main effect of an intervention: what works.

As special education science advances, it is necessary to reflect upon the questions we ask and the way we conduct our research, then adjust accordingly. In recent years, the focus of much work in the intervention space has moved toward testing for whom and under what conditions particular interventions or programs work,

considerations that were not fully addressed in the quality indicators proposed by Gersten et al. (2005). For example, there has been an increased focus on centering equity in the application of group design research and we may learn from similar advancements in related fields. The Consolidated Reporting Standards for Randomized Trials (CONSORT) developed a health equity extension to provide reporting standards for group design research focused on historically marginalized or minoritized populations (Petkovic et al., 2020). Implementation research, “the scientific study of methods to promote the uptake of research findings” into routine practice (Eccles et al., 2005, p. 107), may also advance approaches to study design through systematic conceptualization and testing of both implementation and effectiveness (Curran et al., 2012). Ongoing advancements in analytic approaches and open science practices can also advance how data are used to address the complexity of special education research.

In this article, we aim to identify contemporary considerations for researchers designing, conducting, reporting, and evaluating group design research. We have not proposed a rewriting of the 2005 quality indicators; instead, we have expanded and introduced new considerations for a next generation of quality indicators that advance our understandings of what works, for whom, and under what conditions through group design research. The article is organized into two sections. In the first section, we argue that the initial set of indicators is necessary but perhaps insufficient in addressing the current and future directions of the special education research field. To date, much of the focus in group design research has been placed on identifying effective interventions. Addressing this question is critical to special education practice, and the evidence for what works has accumulated over several decades. However, most typical educational intervention studies report modest effects at best (e.g., Cheung & Slavin, 2016; Chow et al., 2022; Kraft, 2020; Wanzek et al., 2018) and it is common for these effects to diminish over time (Bailey et al., 2020). There is a need to better understand the range of effects within these studies and individual differences in treatment

response to avoid making false generalizations about the effectiveness of interventions.

Stated more directly, although we have much evidence about what works, there remains much to be understood about for whom and under what conditions interventions are most effective.

With this in mind, we review the 2005 quality indicators and describe updates that broaden the focus to include considerations of for whom and under what conditions interventions are more or less effective (see Table 1).

Second, we acknowledge that it is impossible for every group design research study to fully address what works, for whom, and under what conditions, particularly given the complexity of schools, research implementation, and limited funding for large-scale group design studies. This is one place where open science practices can advance the integration of knowledge gained across research teams and research studies. For these reasons, we have introduced a new set of quality indicators related to open science practices (see Table 2) and discuss this unique set of indicators in the second half of this article. Although open science practices should be woven through all aspects of study conceptualization, design, and implementation, these new quality indicators can stand on their own. Each practice has its own distinct quality indicators that must be considered in group design research.

In applying these new quality indicators, special education researchers can increase rigor, accountability, and reproducibility. Open science practices serve not only to make study information more transparent and readily available but also to increase equity in access to this information. For example, open data extend the use of previous funding investments by propelling innovative research questions and accelerating the pace of discovery (Adolph et al., 2012). This practice also serves to democratize access to high-quality and high-powered experimental data, supporting more equitable access among trainees or early-career scholars, scholars of

color, female-identified scholars, and researchers at underresourced institutions (e.g., Ginther et al., 2011; Magua et al., 2017; Murray et al., 2016; Pohlhaus et al., 2011). Therefore, we introduce new quality indicators to advance the field through use of open science practices.

Expanded Quality Indicators to Broaden Lens of for Whom and Under What Conditions

Although special education has made progress in advancing understandings of what works, there has been less focus on concurrently seeking to understand the heterogeneity in such effects. At its foundation, a central goal of special education research has always been to advance equity and inclusion of students with disabilities. High-quality, rigorous research must be designed with intentionality. Considering access and differential response to interventions, practices, and programs can help to inform for whom and under what conditions these interventions work, recognizing that each student holds multiple social identities and other markers of difference that shape their experiences within the education system (S. Garcia & Ortiz, 2013; Hernández-Saca et al., 2018).¹

For these reasons, we must continue to advance rigor in special education research and at the same time also broaden our lens to address questions about the conditions under which effective outcomes occur. The quality indicators introduced by Gersten et al. (2005) can and should continue to play a role in advancing these understandings, but additional focus must be placed on the nuance of their application given the complexity of researching effective practices in school and community settings.

Central to such efforts must be renewing our focus on equity not only as an end goal but also as the starting place for the design and implementation of group design research.

Recently, SEER added an equity standard that calls attention to equity considerations throughout the research process—from conceptualization through dissemination. The current set of expanded quality indicators

integrates these considerations and their application to special education group design research.

Table 1 outlines expanded quality indicators based on ongoing advancements in education science. The items in italics are the original quality indicators carried forward from Gersten et al. (2005), which remain largely unchanged. The additional items are the new or substantively revised quality indicators we propose. We believe the consideration of each of these quality indicators in group design research can advance the rigor of our work, particularly informing for whom and under what conditions interventions, practices, and programs are effective. For these reasons, we do not present different quality indicators for research proposals and reports; this comprehensive set of expanded indicators can be used to guide all phases of research, from development to implementation to review. Further, we elected not to distinguish between desirable and essential indicators. As previously stated, it is not realistic to expect every study to address all quality indicators. Instead, we recommend that all researchers consider and document their decision-making steps related to the full set of quality indicators when designing, conducting, and reporting on group design studies. In the sections that follow, we provide additional context for each of the new or revised indicators presented in Table 1.

Conceptualization and Rationale for the Study

Essential to group design research is a clear rationale, rooted in past research, driven by a clear theory of change, with compelling research questions. This is essential to the scientific process. To understand the evidence for any intervention, practice, and program, we must consider for whom and under what conditions the evidence was collected. We propose four additional quality indicators to guide future group design research in these efforts.

First, in establishing the rationale for a study, a compelling justification should be made for the target population, setting, and conditions being included. This also includes a justification if certain populations, settings,

or conditions are being excluded. Further, if a study includes questions of for whom and under what conditions, then the researcher must provide a compelling case not only for the main effect of the intervention or program but also for these types of questions. For example, why might the intervention be particularly effective for a certain population of learners or for educators in particular school settings? Establishing these questions in advance provides a firmer foundation for scientific questioning, which makes the answers to the questions compelling regardless of the findings. Equity-centered work requires interrogation of the advantages or disadvantages that differentiate groups, as well as the sources of inequities (Brownson et al., 2021). Some direction for conducting equity-centered group design research has been offered by the field of implementation science, which includes producing a compelling rationale for systematic adaptations to an intervention that identifies both underlying assumptions about the intervention or program and potential sources of disparities that may exist (Baumann & Cabassa, 2020; Gaias et al., 2021).

Next, when examining prior literature, it is important to consider what populations, settings, and conditions have been represented. The special education field, although focused on equity for students with disabilities, has not always systematically considered forms of historic marginalization in the development of theory, intervention, and practice recommendations (see N. García et al., 2018). Researchers should ask themselves what other theories, perspectives, or voices should be brought into future work to ensure that heterogeneity of treatment effects can be unpacked. For example, as a methodological subfield of critical race theory that aims to advance quantitative methods related to the study of race and racism, QuantCrit asserts that quantitative research has the potential to uncover the ways in which practices differentially affect groups—presenting counternarratives to race neutrality in our research (Sablan, 2019). To advance our work, we must consider what additional elements may need to be part of the theory of change and what research questions need to be addressed to

Table 1. Expanded Quality Indicators to Broaden Lens of What Works, for Whom, and Under What Conditions.

Conceptualization and Rationale for the Study

1. *Is a compelling case for the importance of the research made? Is the conceptualization based on well-designed studies and does it reflect the scope of extant knowledge?*
2. *If an innovative approach is used, is it based on a sound conceptualization formed from sound research?*
3. *Are the research questions appropriate and stated clearly for the purposes of this study?*
4. *Are valid arguments supporting the nature of intervention in the comparison group(s) presented?*
5. *Does the rationale for the study provide a compelling justification for the targeted population and conditions being studied (including why certain groups or conditions may be excluded)?*
6. *Does the rationale for the study consider issues of equity and make a case for how the research advances our understanding of what works, for whom, and under what conditions?*
7. *Is previous research used to build study rationale generalizable to the target population and conditions?*
8. *Are mediators and moderators purposefully included in the conceptualization of the study?*

Participants and Sampling

1. *Are appropriate procedures used to ensure that participants are comparable across intervention conditions on relevant characteristics? If random assignment is used, is information about participants prior to the intervention made available to ensure that samples are comparable on salient characteristics?*
2. *Is sufficient information be provided to determine whether the participants demonstrated the disability (disabilities) presented?*
3. *Are appropriate procedures used to increase the probability that teachers or interventionists are comparable across conditions?*
4. *Was overall attrition and attrition rates among intervention samples documented and reported? Is any attrition comparable across samples? Is overall attrition less than 30%^a*
5. *Is there a robust description of the sample, including both student and implementers, provided that includes a range of sociodemographic characteristics (e.g., race-ethnicity)?*
6. *Is adequate information provided about participants' disability-related support needs and experiences?*
7. *Is the sample selection at the school, teacher, and student levels transparently described with attention to how the sample does or does not generalize to the range of students, educators, and schools for whom the intervention might be implemented?*

Implementation and Context

1. *Is the intervention clearly described?*
2. *Are procedures for ensuring and assessing fidelity of implementation described?*
3. *Is the nature of instruction or services provided in comparison conditions described?*
4. *Is there documentation of the nature of instruction or services provided in comparison conditions?^a*
5. *Are study conditions masked from data collectors and scorers and equally (un)familiar to examinees across study conditions?^a*
6. *Are multiple indicators of fidelity of implementation measured and reported? Are these indicators considered in interpretation of study findings?^b*
7. *Are intervention materials available that allow others to understand implementation?^b*
8. *Is there transparency and robustness in the description of the classroom, school, and community and is this appropriately considered in the analysis plan?*
9. *Is school context considered in the generalizations made about study findings or intervention effects?*

Outcome Measures

1. *Are multiple measures used to provide an appropriate balance between measures closely aligned with the intervention and measures of generalized performance?*
 2. *Is evidence of the validity of the measures discussed?^a*
 3. *Are outcomes for capturing the intervention's effect measured at the appropriate times?*
 4. *Are outcomes for capturing the intervention's effect measured beyond an immediate posttest?^a*
 5. *Is evidence of reliability for the outcome measures provided (e.g., internal consistency, test-retest, interrater reliability), including sample-specific reliabilities?^b*
 6. *Are any researcher-designed measures made available?*
 7. *Do test batteries include measures that are commonly and widely used in similar intervention*
-

(continue)

Table I. (Continued)

research? If not, is attention paid to describing the similarities and differences between the measures used and those that are more common?
Research Design and Data Analysis
1. <i>Is the variability within each sample accounted for either by sampling techniques (e.g., restricting range) or appropriate statistical techniques?</i>
2. <i>Does the research report include not only inferential statistics but also effect size calculations?</i>
3. <i>Were the results presented in a clear, coherent fashion?</i>
4. Are descriptive statistics (i.e., means, standard deviations, correlations) supplied for all key measures?
5. Are the research design and the data analysis techniques appropriate and linked to key research questions and hypotheses? ^b
6. Are a priori power analyses conducted for key parameters involved in each inferential research question? Is the power analysis aligned with the research questions, levels of nesting in the data, and proposed analyses? ^b
7. Is the effect size used in power analyses reasonable to expect and well justified based on prior literature?

Note. Quality indicators carried forward from Gersten et al. (2005) are in italics. We integrated indicators for research proposals and reports; as such, they were edited for clarity and are not presented verbatim. For description of these quality indicators, see original article.

^aIndicator that was originally listed as desirable rather than essential.

^bIndicator related to this general issue was present in Gersten et al. (2005) but has been revised to reflect advancements in the field.

determine if an intervention or program does (or does not) promote equitable outcomes for students with disabilities. This is particularly important for students with disabilities who may experience multiple marginalized identities (e.g., students of color, LGBTQ+ students). Researchers make innumerable decisions when conducting research—and, collectively, we must attend to transparency of decision-making to further advance rigor and replication in the field. Issues related to equity and how they are or are not addressed need to be explicitly integrated into the conceptualization that justifies group design research studies.

Finally, as research questions become more focused on for whom and under what conditions interventions work, researchers should plan, in advance, for the investigation of potential mediators and moderators of intervention or program effects. Researchers will need to determine the groups, settings, or conditions under which there might be expected variability. The selection of mediators and moderators have specific and multifaceted implications on study design. Because they lie in similar orthographic and phonological neighborhoods, the terms “mediation” and “moderation” often get

confused. We define each of these approaches briefly here.

A mediation analysis attempts to understand the mechanism through which an intervention is effective. As such, candidate mediator variables must be measured after the assignment to treatment has occurred but before the outcome is assessed (i.e., temporal precedence; see Kendall et al., 2017). In addition, mediator variables should be hypothesized to differ primarily because of treatment assignment and to be directly related to the outcome. One way of conceptualizing mediation is that we expect the difference between the treatment and control groups on the outcome to be minimized (or zero) once the mediating variable is accounted for (Baron & Kenny, 1986). Mediation can also be conceptualized as the indirect effect of the treatment on the outcome through the mediator (Agler & de Boeck, 2017; Hayes & Preacher, 2014; Kendall et al., 2017). For example, in a teacher-implemented intervention to improve students’ reading skills, the researcher might ask a mediation question such as, Is there a difference between reading skills of students in my treatment and control groups once I have controlled for

the extent to which educators implement core intervention components with fidelity? Or, Is there a direct relation between the assignment to condition and students reading outcomes or is that relation indirect, mediated by the extent to which educators implement core intervention components with fidelity?

Conversely, moderation analysis tests whether the relation between treatment and outcome varies as a function of an additional variable (the moderator). Moderator variables have fewer restrictions for the timing of their measurement than mediators. Moderators can be measured at any point during the study, but their inclusion should similarly be rooted in theory. For a dichotomous moderator, a moderation analysis would hypothesize that the expected difference between treatment and control groups on the outcome is different for each group. For example, a researcher might ask, Is this intervention differentially effective for students with and without disabilities? For a continuous moderator, a moderation hypothesis would posit that the treatment would be differentially effective for students with high versus low values on the moderator variable. For example, one might consider individual differences based on a foundational skill as a potential moderator of treatment effects on a more complex skill. For example, does students' working memory moderate the effectiveness of an intervention designed to improve understanding of fractions?

Participants and Sampling

We propose three expanded quality indicators focused on providing adequate information about participants and sampling procedures. There has been a push in special education, even before Gersten et al. (2005), to describe participants thoroughly in an effort to promote generalizability of study results. However, this quality indicator has not been consistently implemented in the special education intervention literature. At a basic level, researchers frequently do not report student sociodemographic characteristics (e.g., race and ethnicity, gender identity), nor are these characteristics regularly considered in establishing what works (Graves et al., 2021; Steinbrenner et al., 2022). This is also true for

implementer characteristics, despite research suggesting that there can be an interaction of student and implementer characteristics in study outcomes (see Redding, 2019, for review). This not only limits generalizability but also reflects a lack of focus on establishing for whom and under what conditions interventions work.

Beyond just describing participants, quality indicators have focused on ensuring comparability across conditions and confirming disability status of study participants. With these expanded indicators, we focus additional attention on how disability is defined in studies and how disability-related support needs are measured and reported. Diagnostic information may not be the only or most important factor to understand in determining why there is heterogeneity of treatment effects; additional information needs to be provided on disability-related support needs and students' educational experiences, including educational placement and previous access to instruction and intervention.

Finally, more transparency is also needed in decision-making about sample selection, as the participants in any given study are rarely representative or generalizable to the population as a whole (Tipton & Olsen, 2022). For example, if only non-Title I schools are included in a study that uses school-based random assignment, this needs to be communicated, aligned with the theory of change, and justified. Relatedly, if schools included in the study have high rates of restrictive placements (e.g., self-contained classrooms) for students with certain disability labels, then this needs to be described, as it may limit who participates in intervention studies, particularly if students are being recruited from general education classrooms. We recommend that decision-making be guided by a focus on understanding heterogeneity of treatment effects, at both the student and the school level, and that this decision-making about sample selection is documented. Defaulting to samples that are "accessible" or "convenient" cannot be a justification, particularly when attempting to advance equity by understanding for whom interventions work. This will necessitate new and different steps to engage with marginalized communities and enable their participation in group design research.

Implementation and Context

Past indicators have led to improved documentation of and reporting on the intervention, study conditions, and research procedures across these conditions. We propose additional indicators to guide researchers in these efforts. For example, it is necessary and generally feasible to measure multiple indicators of fidelity and consider these findings in the interpretation of study outcomes. The core components of an intervention, particularly given the complexity of school-based implementation, must be clearly defined as well as measured in multiple ways (e.g., adherence, quality, responsiveness) to clarify what works as well as inform for whom and under what conditions (Ferber et al., 2019).

Next, researchers should ensure that intervention materials are available to allow others to understand the study implementation. Journal guidelines may not always provide enough space in published articles for robust descriptions, but materials can be made available through other avenues. This access supports replication efforts in the field and promotes a better understanding of the contextual factors that support (or impede) implementation. Additional attention should be directed to transparency and documentation of decision-making in key aspects of the implementation to inform the conditions under which interventions, practices, or programs are effective. For example, describing community, school, and educational program factors can allow for the examination of these issues as mediators and moderators, examining the influence of context, consistent with research questions and aims.

The notion that context matters is widely accepted; however, there is a need to better define context in efforts to explain sources of influence on intervention implementation and outcomes in group design research studies. Although describing the intervention and comparison conditions remains essential, researchers should also attend to operationalizing and documenting contextual factors that shape implementation (Shogren et al., 2014). "Education settings are not laboratories. The contexts and conditions under which an intervention is implemented are part of the intervention" (Conaway et al.,

2022). Although random assignment is still recognized as a means to control for contextual factors, more fully addressing, documenting, and examining factors related to community-level socioeconomic status, school climate, school inclusivity, and collaboration between general and special education are needed to understand the conditions under which interventions are being tested, as these factors are likely to influence the generalizability of findings in group design research.

In researchers' efforts to address these issues more fully, they might look to implementation science research for guidance. Group design researchers focus on intervention or program effectiveness and, in general, are deeply interested in monitoring or controlling implementation (e.g., fidelity) or enhancing the *practice* of implementation. Understanding implementation science provides an opportunity for special education research to systematically study implementation through, for example, use of effectiveness-implementation hybrid trial designs (Curran et al., 2012). These designs are used to concurrently examine both the effectiveness of the intervention within real-world settings, such as schools, and implementation strategies or outcomes (e.g., acceptability of the intervention) related to uptake. There are three types of hybrid designs distinguished based on whether the primary research question is related to effectiveness or testing an implementation strategy (for review, see Bernet et al., 2013; Nilsen, 2015).

Outcome Measures

We propose three additional quality indicators to supplement those originally proposed in 2005. Gersten et al. provide clear guidance on appropriate selection and reporting of a study's outcome measures, including use of multiple outcome measures and evidence of the reliability and validity of these measures (see Table 1). Although it has been long accepted that researchers should report evidence of reliability for the outcome measures being used, we further suggest reporting of reliability that informs our understanding of for whom and under what conditions programs work. Specifically, researchers should report the sample-specific reliabilities for any

measure, even standardized measures, as they may or may not hold within a given sample. This means that data should be collected, entered, and stored at the item level. We believe that this are essential to the advancement of the field and accurate interpretation of the body of evidence drawn from group design studies.

Next, researcher-designed measures are often used to investigate intervention outcomes alongside other standardized tests or assessment tools. Because these measures are closely aligned with the intervention, researchers can better understand the specific intervention content or components that appeared to benefit students. In recent commentary, Clemens and Fuchs (2022) further argued for the use of multiple measures—both commercial, norm-referenced tests and researcher-designed tests—when seeking to understand effects on complex outcomes, such as reading comprehension. Researchers should provide a clear rationale for the selection and use of all study measures, including measures that assess skills closely aligned with the content of the intervention and those that assess transfer to more global skills or performance indicators. We further recommend, with these expanded indicators, that researcher-designed measures used in group design studies be made available for other researchers to review items and use for purposes of replication.

Finally, we suggest that researchers aim to include measures that are commonly and widely used in similar intervention research. These need not be commercial, norm-referenced measures, but they must be valid and reliable for the target population of learners and the key outcomes under investigation. We echo the recommendation made in the recent National Academies of Sciences, Engineering, and Medicine (NASEM) report on the future of education research, which stated that “encouraging, but not requiring, common measures is ideal and allows investigators to pursue innovative measures as called for by theory and the needs of particular studies” (NASEM, 2022, pp. 145–146). Federal funding agencies have made policy recommendations related to the use of

common measures, with the stated goal of maximizing researchers’ ability to address a greater range of significant questions across more diverse samples (e.g., Hodes et al., 2013). For example, the National Institutes of Health has been encouraging the use of common data elements, standardized questions, and instruments used systematically across studies to increase the rigor of research and sharing of data across projects for more than a decade (e.g., www.nihtoolbox.org). Similarly, IES has described the importance of “making common measures more common” (Schneider, 2020) so that study outcomes more clearly inform evidence-based decision-making and allow for comparisons of relative impact and cost of various interventions. There is a growing number of online repositories (e.g., EdInstruments.com) aimed at increasing access to educational measures. Increased engagement in open science and the practice of sharing materials will further enhance ease of access to a range of measures.

If no common measures are used, researchers should make an effort to describe the similarities and differences between the measures they opted to use and those that have been used to test the efficacy of similar interventions. This quality indicator speaks to the ultimate goal of conducting group design research—contributing to the knowledge base from which conclusions can be drawn about what works, for whom, and under what conditions. No single group design study, regardless of the quality or rigor of the study, should be relied on for evidence that directs wide-ranging recommendations for special education policy and practice. For example, the findings from a large-scale experimental study may offer compelling evidence for the efficacy of a new intervention; however, these findings must be replicated with attention to testing and explaining potential sources of heterogeneity in these treatment effects (Bryan et al., 2021). To interpret the range of evidence contributed by various studies, researchers must be able to report effects on similar constructs across these studies (e.g., measures aligned with the key outcomes that the intervention claims to target). Furthermore, various methodological

advances, such as integrative data analysis, allow for investigation of effects across multiple studies—and facilitate researchers' ability to address questions of critical importance to special education that are not often possible to address within a single study (Bauer & Hussong, 2009). This technique, which allows researchers to better understand the extent to which findings generalize to different groups of learners, requires that each study include common or very similar measures. This is further discussed in the section on open materials.

Research Design and Data Analysis

We propose four additional quality indicators to guide researchers based on the advancements related to design and analysis of group design research studies. First, reported study findings should include thorough documentation of descriptive statistics (i.e., means, standard deviations, correlations) for all key measures, either in the manuscript itself or as supplemental material. Thorough documentation of such descriptive statistics will allow for clearer interpretation of the results and facilitates future meta-analytic work to better understand evidence in the field.

Second, the data analysis proposed (in a grant) or conducted (in a paper) should be clearly aligned and matched to the research questions posed. For example, a question about the efficacy of a mathematics program in elementary school would not be addressed by an analysis aimed at identifying subgroups within a sample based on shared characteristics (i.e., latent class analysis). The concept of alignment is particularly important to consider as new research methods and designs are being developed. Leveraging any type of research design has ripple effects on all aspects of the research process, from database design to participant recruitment, from how and when data should be collected to data analysis and many stages between. The field continues to make advances in research design and analytic methods. As researchers learn about new methods, it may be especially important to consult or collaborate with experts who can guide a match between design and analyses.

Inferential group design research should include a priori power analyses (e.g., IES CFDA 84.305 and 84.324). Although a priori consideration of statistical power is a quality indicator, this does not mean that studies should conduct post hoc power analyses; such analyses are meaningless at best and misleading at worst (Zhang et al., 2019). Second, although power analyses are often used to determine the sample size during study planning, statistical power is not a property of a sample size. Rather, power is a property of a single statistical test within a given statistical model (e.g., a *t* test of the parameter representing the difference in the expected means for the treatment and control groups within a multiple regression model). Statistical power varies as a function of many different features, including the expected effect for the particular outcome measure, planned analytic technique, and inclusion of covariates. Thus, any change to the included variables, statistical model, or focus of the research question will result in a different estimate of statistical power (Pek et al., 2022). As the focus of much of group design research is moving toward questions of for whom and under what conditions particular interventions or programs work, researchers should be sure to include these questions in their a priori power analyses. Conducting a power analysis for an interaction (or moderation) effect requires additional parameters, particularly when the design necessitates a multilevel structure, but should still be conducted as part of the study design process (see Dong et al., 2021). Power analyses should be aligned directly with each research question in terms of the sample size, the inclusion of nesting structure, the proposed analytic technique, and the specific parameter being evaluated.

Finally, each research question must be powered for a specific effect. In group design research, effects typically are the difference between the treatment (or treatments) and control groups on some key outcomes of interest. It is important that this effect both be meaningful and potentially achievable based on prior literature and address the justification for the targeted population made in the

study conceptualization. The effect should be meaningful in that it should be based on an understanding of the outcome rather than arbitrary (e.g., based on Cohen's benchmarks; Kraft, 2020). For students' academic outcomes, consider basing an effect on empirical benchmarks, such as those set for reading and mathematics (Hill et al., 2008) or language (Schmitt et al., 2017). An effect should be achievable, meaning needs to be justified based on prior literature that has used a similar sample, context, intervention, dose, delivery system, or outcome. It is important to consider each of these factors when justifying and estimating how large of an effect may be detectable under the specified sample and conditions.

Summary of Expanded Quality Indicators

Through these expanded quality indicators, our hope is that special education researchers will more purposefully design studies that broaden our lens to help us better understand for whom and under what conditions interventions, programs, or practices work. In doing so, we have the opportunity to address issues of equity from the conceptualization of a study and through all phases of our research. These expanded quality indicators direct attention toward the need for more robust critique of how past research has (or has not) addressed equity; to expand participants that are included in research studies, planning from the start to advance equitable access and participation in knowledge generation; to document and analyze contextual factors at the student, family, school, and community levels that impact intervention implementation and outcomes; and to be planful in using outcome measures and study design in a way that advances understanding across our field.

There remains an essential question of how to bring these quality indicators together to conduct studies that explore what works, for whom, and under what conditions. Efforts to center equity within the context of group design research have often been more aspirational and resulted in too few tangible examples. As previously noted, current reviews suggest that even basic reporting of participants' race and ethnicity

remains lacking across special education research (e.g., Steinbrenner et al., 2022). We can look to other fields for examples of how to advance our work while maintaining rigorous standards; for instance, Mbuagbaw and colleagues (2017) developed the PROGRESS Plus framework (place of residence, race/ethnicity/culture/language, occupation, gender, religion, education, socioeconomic status, social capital, "plus" other context-specific factors) to support the integration of health equity and social determinants of health into randomized controlled trials. They propose that the framework could be used to formulate research questions focused on one or more of these characteristics if the research is being conducted with a marginalized or vulnerable group. Further, the framework could be used to guide decision-making related to the study design or disaggregation of the data based on specific characteristics. The PROGRESS framework also acknowledges intersecting identities, suggesting that more than one characteristic may be important to consider. Of note, to further guide reporting of clinical trials, a health equity extension was added to the CONSORT reporting standards that infuses the PROGRESS framework (Petkovic et al., 2020; Welch et al., 2017).

These ongoing efforts to advance rigor in special education research are further supported through systematic documentation and sharing of information to increase transparency in our work. The open science practices described in the following section serve to enhance methodological rigor in special education research and further our understanding of what works, for whom, and under what conditions.

New Quality Indicators to Further Methodological Rigor Through Open Science

A relatively recent advancement in education research is the concept of open science. Recent work in the metascientific literature suggests that education researchers are becoming more aware of these practices, even though many are not yet engaging in open science practices themselves (Makel et al., 2021).

Table 2. New Quality Indicators to Further Rigor Through Open Science Practices.**Preregistration**

1. Is the preregistration posted online in a study registry?
2. Does the preregistration contain all planned confirmatory research questions?
3. Does the preregistration include detailed information about how each construct will be measured, including how variables will be calculated or represented?
4. Does the preregistration include data processing and cleaning plans?
5. Does the preregistration contain detailed plans for how data analyses will be approached for each confirmatory research question?
6. Does the preregistration contain a detailed analysis plan for each confirmatory research question?

Open Results

1. Is the manuscript, including all supplementary materials, freely available online?
2. Is the manuscript shared as a preprint in an online repository or in an open access journal?

Open Materials

1. Are the materials used to conduct the study available openly available?
2. Do the materials include detailed documentation of the intervention and control group?
3. Does documentation include detailed information about the sample and context of the work?
4. Is the code used to run statistical analyses openly available?

Data Sharing

1. Are the data shared in a data repository, either with fully open access or through a clearance process?
2. Are the data shared at the item level, as possible?
3. Are the data shared with thorough documentation of study design and study purpose?
4. Are the data shared with thorough documentation of all variable names and values?

Although open science was not addressed by Gersten et al. (2005), guidelines for open science in education do exist in some form. In October 2021, IES published an update to SEER. These standards now include two specific open science practices: preregistration and making findings, methods, and data open. Although inclusion of these standards is an important step forward, they are not comprehensive of the current state of open science, nor are they written specific to group design research in special education. We expand on these standards and propose new quality indicators specific to open science practices to guide group design research (see Table 2). In the sections that follow, we describe indicators related to preregistration, open results, open materials, and data sharing.

Preregistration

As the field advances, preregistration of group design studies can support increased transparency. The first SEER standard in this area states, "Causal impact studies must be pre-registered in a recognized study registry, documenting their confirmatory research questions and planned

analytic activities" (IES, 2019). Preregistration is the process of documenting or mapping out the study (see Cook et al., 2022; Munafò et al., 2017). We propose specific quality indicators to guide development of a quality preregistration specific to group design research.

Researchers should make preregistrations publicly available by publishing them in an online repository. For example, readers may be familiar with the Society for Research in Educational Effectiveness and its Registry for Efficacy and Effectiveness trials (<https://sreereg.icpsr.umich.edu>). This is a highly structured registry designed to be aligned with IES's SEER principles. There are several other options for preregistration repositories, and the choice may be influenced by the particular study being conducted or requirements of the funding agency.

Much of the information included in a preregistration will align with the researchers' decision-making related to the previously introduced set of quality indicators (see Table 1). Specifically, in developing a preregistration, researchers should address five key components. First, the preregistration should include all planned confirmatory research questions. In

group design research, these are typically questions about the effectiveness of a treatment (i.e., the difference between the treatment and control groups). This will also include any planned contrasts to examine for whom or under what conditions a particular intervention works. Next, a preregistration should include detailed information about how each construct will be measured—not only which measures will be used but details for how study variables will be calculated or represented. For example, a preregistration might indicate that reading achievement will be measured with the Word Identification subtest of the Woodcock-Johnson III. However, readers familiar with this subtest are aware that this can be expressed as a raw score, total score, norm-referenced standard score, W score, or percentile rank. It is important that the preregistration describe the measures and relevant variables to be identified. Third, researchers should detail the data-cleaning steps that will be undertaken to preprocess the data prior to analysis. How will outliers be identified, and what is the plan if they are identified? How will missing data be handled? The preregistration should document plans for any preliminary analyses required before the inferential questions can be tested (e.g., if a confirmatory factor analysis model is required to be fit to the data prior to a structural path testing the research question, explain how the final factor model will be selected). This should also include plans to address or account for missing data. Fourth, the preregistration should include detailed plans for how data analyses will be approached for each confirmatory research question. What analysis will be used? Which covariates will be included (and how will they be measured)? Finally, power analyses should also be included in the preregistration of study materials. As previously described, this should include power analysis reported for each a priori inferential research question.

Resources. For more general information about the importance of preregistration, researchers can be directed to the article by Munafò and colleagues (2017). For researchers wanting to learn more about deciding how and where to preregister a study, guidelines exist to compare the different available

preregistration templates (Fleming, 2021) and platforms (Haroz, 2022) based on study features. Finally, researchers may consult a practical how-to guide, such as the one written by Cook and colleagues (2022) specific to special education research.

Open Results

The next open science practice is open results, or open access principles, which refers to making results or findings of a given study publicly available free of charge (Klein et al., 2018). In this instance, the term “results” refers not specifically to the statistical analyses conducted but to the final products that have been created as part of the research process. We propose two connected quality indicators: that the manuscript and all supplemental materials be made freely available online and that the manuscript also be shared as a preprint in an online repository or in an open access journal.

Although the principle of open results is important to all of science, it can be particularly important in group design research for several reasons. First, intervention science is frequently evolving and changing as researchers begin to discover teaching and implementation techniques that work or do not work. Advances in delivery, implementation, and adaptation are key to the success of group design work and need to be available to all researchers in order to design effective interventions moving forward. Further, open access allows for findings to be included in future meta-analytic work. Meta-analyses are key tools that help education science advance our understanding of what works, for whom, and under what conditions. For example, although any one study might be able to demonstrate that a particular practice or intervention is effective for the group of students involved, it can generalize only to one population of students. Through meta-analysis of multiple studies of similar intervention targets, conducted at many different sites and with different samples, researchers can ask for whom and under what conditions particular approaches work to improve student outcomes.

Resources. One way to post findings openly is to publish in a journal that allows for open

access. These often are costly to the submitter, depending on the journal (see Cook et al., 2018; van Dijk et al., 2021). However, openly sharing findings can be as simple as posting a version of the finished paper online as soon as it is completed and sent in for peer review. This practice is called posting a preprint (called a “preprint” because it is available to read before it has been printed or published online; Bourne et al., 2017). One simple option is to post a preprint on an individual or project website. However, there are several advantages to posting a preprint on a dedicated preprint server or repository. Preprint repositories are searchable, and with several stored in one place, the preprint will be more discoverable. One resource for special education researchers to consider for hosting their preprints is EdArVix (<https://edarxiv.org>), a domain-specific repository for education science. In either instance, preprints are both free to access and free to post as a depositor. Finally, for readers concerned about this new practice, note that many journals now have policies that explicitly allow for preprints. For example, the *Journal of Learning Disabilities* provides instructions for how to reference both preregistrations and preprints in its author guidelines. To learn more, Cook and colleagues (2018) provide an excellent discussion of open access options specific to special education research.

Open Materials

Open materials, or materials sharing, involves the research team providing freely available access to the materials used in conducting the research study. This generally includes all of the materials necessary to successfully run a replication study (Grahe, 2018). Sharing such materials is aligned with IES’s policy regarding public access to research (IES, n.d.), where it calls for the “use of transparent research methods.” We propose four quality indicators related to open materials for group design researchers. First, the materials used to conduct the study should be openly available. For group design research in special education, these should include recruitment protocols, detailed testing procedures and schedules for students, and copies of any

locally developed measures. For the second quality indicator, we specifically highlight the need for information about the intervention and control conditions, which should include examples of intervention materials, details of the intervention scope and duration, the theoretical active ingredients of the intervention, details of implementation, copies of locally developed fidelity checklists, and information about the activities provided to the control group or the activities and practices of the business-as-usual condition. Third, researchers should make openly available thorough documentation of the sample characteristics and the context in which the work was conducted, such as was discussed in detail in the previous quality indicators of implementation and context. Finally, open materials should also include sharing the code used to conduct the analyses in pursuing the research aims (Cook et al., 2018).

These indicators are particularly important because group design research is often focused on evaluating an entire package that may have many active ingredients, with a primary focus on the sample as a whole. Page length constraints may force some key details to be left out of standard journal article reporting. Providing access to key materials allows researchers to learn from others’ experiences. Researchers can build from these materials, adapt them for their own studies, or replicate the work in a different target population. Further, meta-analytic work or integrative data analysis can allow researchers to look at generalizability of an effect or practice across multiple group design studies based on their features or for tests of whether particular practices work across populations. In this way, providing access to study materials supports science to determine what works, for whom, and under what conditions (van Dijk et al., 2021).

Resources. There are many services that can host the materials used to conduct a group design study (e.g., figshare, <https://figshare.com>; Inter-University Consortium of Political and Social Research, www.icpsr.org; Open Science Framework maintained by the Center for Open Science, <https://osf.io>); however, they can also be stored as supplementary material to a published article or can also be stored

wherever data are stored. It is also possible to publish a paper specifically about the methods used to conduct a study that thoroughly defines the sample, recruitment strategies, and the context of the study. This becomes a direct citation for any future publications using those data. For a recent example of publishing about the methods of a study in learning disabilities research, see van Dijk et al. (2022). In regard to code sharing, code can be uploaded in the same manner as other supplemental materials or can be hosted on code-specific websites (e.g., GitHub). For a discussion on the benefits and challenges of sharing code, including several resources, see Allen and Mehler (2019).

Data Sharing

Data sharing has been a requirement for any data collected with federal grant funds since a directive from the Obama administration's Office for Science and Technology Policy (2013). Yet recent survey work finds that education researchers report rarely or never engaging in data sharing practices (Makel et al., 2021). Even though it is far from a customary practice, providing access to data is a strong quality indicator for several reasons. First, shared data makes it possible for reviewers to conduct checks of the data analyses for accuracy. They may be uncommon, but mistakes can and do happen. Variables can accidentally be interpreted incorrectly (e.g., Aboumatar & Wise, 2019) or merged incorrectly (e.g., Goldberg et al., 2008), or missing data coded incorrectly (e.g., Beheim et al., 2021). In each case, the errors were found by a second research team when the data were shared, and the errata provided an opportunity to correct the scientific record. These corrections should be recognized as contributions to rigor in our work.

Assuming data and analyses are correct, there are still several additional reasons that data sharing is important for group design research. For one, though a group design research project is typically set up to examine one specific question, there are always several additional questions that can be asked from the same data set (Logan et al., 2021), perhaps focusing on a specific

subgroup (for whom), the role of a particular contextual factor, or features of a specific measure that were not explored in the initial analysis (under what conditions). Second, sharing data removes some of the barriers to science for researchers who have fewer resources. Third, new methodological and statistical techniques are regularly being developed. Through data sharing, any released data can be analyzed using newly developed techniques, and such reevaluation can advance the field's understanding of for whom and under what conditions the main effects of the study hold. Finally, some research questions cannot be addressed at the scale of individual data sets. Researchers studying smaller populations, low base rates, or low frequency of the targeted behavior or skill rely on integrating multiple data sets (e.g., Bainter & Curran, 2015).

Resources. The first quality indicator in this section is that the data be stored in a data repository, either with fully open access or through a clearance process. Although research teams have for many years relied on the disclaimer that "data will be made available on request," this is a historically unreliable method of data sharing (Wicherts et al., 2006). It is time-consuming for both the requester and the data holder, and availability of the data declines over time (Vines et al., 2014). Therefore, we encourage researchers to instead share their data in a domain-specific repository (e.g., LDbase, www.ldbase.org). Further, research teams should include item-level data whenever possible, as item-specific information can support error identification (Logan et al., 2021) and increased usability through integrative data analysis (IDA; Bauer & Hussong, 2009). IDA involves combining multiple data sets to create a new analytic data set that maintains variability at the item or individual level, which allows for future analyses that may advance knowledge of what works, for whom, and under what conditions (e.g., van Dijk et al., 2022).

The last two quality indicators refer to the documentation that accompanies a shared database: documentation of study design and

purpose, and documentation of variable names and values. These materials are necessary components to understanding and interpreting the data. Logan and colleagues (2021) present an in-depth discussion of data sharing for education research, including common concerns, refutations, and resources. Further, Meyer (2018) provides specific recommendations related to the ethical sharing of data.

Conclusion: Implementing the Next Generation of Quality Indicators

We have built on earlier recommendations for high-quality, rigorous group design studies introduced by Gersten et al. (2005). We propose an expanded set of quality indicators for group design research in special education; these indicators broaden the lens used to conceptualize and evaluate group design research to better understand for whom and under what conditions our interventions are most effective. We have also integrated new quality indicators related to advancing methodological rigor through open science practices. We present these quality indicators with the hope that special education researchers will take them up not as a “checklist” to be addressed when designing or evaluating a study but rather a set of guiding questions to increase the purposeful design, implementation, and analysis of group design research in special education to advance the ultimate goal of equitable educational outcomes for all students with disabilities.

These quality indicators should be seen as a guide to conducting the most thoughtful, context-relevant, rigorous research that we are able.

This is not to say that there are not real constraints on study design and implementation. Science is incremental, and we will not be able to address every important issue in one study alone; however, it is essential to document constraints on our research so that they are identified. This identification informs future equity-driven research. These quality

indicators should serve to guide us in reflection upon the limitations of the study, interpretation of findings, and considerations as we plan future work. Further, as previously described, the indicators presented in Tables 1 and 2 complement each other. The purposeful engagement in open science practices not only advances the methodological rigor of special education research but also allows for researchers to investigate questions that may not have been answerable within the constraints of a single study. Further, increasing open access allows for other researchers to explore new ways of testing the same research questions even with the same data set. Addition of new covariates, a slight change in the conceptualization of the outcome variable, or a focus on a particular subset of the sample may be important conceptual questions that are key to building a robust understanding of what works, for whom, and under what conditions.

Note

1. The term “identities” refers to socially constructed categories used to explain individual or group characteristics. These categories are often influenced by how individuals are perceived and treated, at both interpersonal and institutional levels, and the simultaneous interactions of these identities shape experiences of marginalization. For example, S. Garcia and Ortiz (2013) explain, “While a disability label may assign students to a subordinate status in a general education classroom on the basis of their perceived disabilities, their gender, social class and/or racial identities may mitigate this status in different ways, creating privilege for some but disadvantage for others” (p. 34).

References

- Aboumatar, H., & Wise, R. A. (2019). Notice of retraction. Aboumatar et al. Effect of a program combining transitional care and long-term self-management support on outcomes of hospitalized patients with chronic obstructive pulmonary disease: A randomized clinical trial. *Journal of the American Medical Association*, 322(14), 1417–1418. <https://doi.org/10.1001/jama.2019.11954>
- Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological*


- Inquiry*, 23(3), 244–247. <https://doi.org/10.1080/1047840X.2012.705133>
- Agler, R., & DeBoeck, P. (2017). On the interpretation and use of mediation: Multiple perspectives on mediation analysis. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01984>
- Allen, C., & Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. *PLOS Biology*, 17(5), e3000246. <https://doi.org/10.1371/journal.pbio.3000246>
- Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest*, 21(2), 55–97. <https://doi.org/10.1177/1529100620915848>
- Bainter, S. A., & Curran, P. J. (2015). Advantages of integrative data analysis for developmental research. *Journal of Cognition and Development*, 16(1), 1–10. <https://doi.org/10.1080/15248372.2013.871721>
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>
- Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*, 14(2), 101. <https://doi.org/10.1037/a0015583>
- Baumann, A. A., & Cabassa, L. J. (2020). Reframing implementation science to address inequities in healthcare delivery. *BioMed Central Health Services Research*, 20, 190. <https://doi.org/10.1186/s12913-020-4975-3>
- Beheim, B., Atkinson, Q. D., Bulbulia, J., Gervais, W., Gray, R. D., Henrich, J., Lang, M., Monroe, M. W., Muthukrishna, M., Norenzayan, A., Purzycki, B. G., Shariff, A., Slingerland, E., Spicer, R., & Willard, A. K. (2021). Treatment of missing data determined conclusions regarding moralizing gods. *Nature*, 595(7866), E29–E34. <https://doi.org/10.1038/s41586-019-1043-4>
- Bernet, A. C., Willens, D. E., & Bauer, M. S. (2013). Effectiveness-implementation hybrid designs: Implications for quality improvement science. *Implementation Science*, 8(1), 1–2. <https://doi.org/10.1186/1748-5908-8-S1-S2>
- Bourne, P. E., Polka, J. K., Vale, R. D., & Kiley, R. (2017). Ten simple rules to consider regarding preprint submission. *PLOS Computational Biology*, 13(5), e1005473. <https://doi.org/10.1371/journal.pcbi.1005473>
- Brownson, R. C., Kumanyika, S. K., Kreuter, M. W., & Haire-Joshu, D. (2021). Implementation science should give higher priority to health equity. *Implementation Science*, 16(1), 1–16. <https://doi.org/10.1111/1468-0009.12112>
- Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, 5(8), 980–989. <https://doi.org/10.1038/s41562-021-01143-3>
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292. <https://doi.org/10.3102/0013189X16656615>
- Chow, J., Zhao, H., Sandbank, M., Bottema-Beutel, K., & Woynaroski, T. (2022). Empirically-derived effect size distributions of interventions for young children on the autism spectrum. *Journal of Clinical Child & Adolescent Psychology*. Advance online publication. <https://doi.org/10.1080/15374416.2021.2007485>
- Clemens, N. H., & Fuchs, D. (2022). Commercially developed tests of reading comprehension: Gold standard or fool's gold? *Reading Research Quarterly*, 57, 385–397. <https://doi.org/10.1002/rrq.415>
- Conaway, C., Tipton, E., & Artiles, A. J. (2022). *The value of variation: Why we need to attend to heterogeneity in intervention research*. Stanford GSE Public Scholarship Collaborative. <https://publicscholarship.stanford.edu/psc-blogs-0>
- Cook, B. G., Fleming, J. I., Hart, S. A., Lane, K. L., Therrien, W. J., van Dijk, W., & Wilson, S. E. (2022). A how-to guide for open-science practices in special education research. *Remedial and Special Education*, 43(4), 270–280. <https://doi.org/10.1177/07419325211019100>
- Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, 85(1), 104–118. <https://doi.org/10.1177/0014402918793138>
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*, 50(3), 217. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3731143>
- Dong, N., Spybrook, J., Kelcey, B., & Bulus, M. (2021). Power analyses for moderator effects with (non) randomly varying slopes in cluster randomized trials. *Methodology*, 17(2), 92–110. <https://doi.org/10.5964/meth.4003>

- Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). Changing the behavior of health-care professionals: The use of theory in promoting the uptake of research findings. *Journal of Clinical Epidemiology*, 58(2), 107–112. <https://doi.org/10.1016/j.jclinepi.2004.09.002>
- Ferber, T., Wiggins, M. E., & Sileo, A. (2019). *Advancing the use of core components of effective programs*. Forum for Youth Investment. <https://forumfyi.org/knowledge-center/advancing-core-components>
- Fleming, J. (2021). *Preregistration registry figure*. Figshare. <https://doi.org/10.6084/m9.figshare.16746112.v4>
- Gaias, L. M., Arnold, K. T., Liu, F. F., Pullmann, M. D., Duong, M. T., & Lyon, A. R. (2021). Adapting strategies to promote implementation reach and equity (ASPIRE) in school mental health services. *Psychology in the Schools*. Advance online publication. <https://doi.org/10.1002/pits.22515>
- Garcia, N. M., López, N., & Vélez, V. N. (2018). Quantcrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education*, 21(2), 149–157. <https://doi.org/10.1080/13613324.2017.1377675>
- Garcia, S. B., & Ortiz, A. A. (2013). Intersectionality as a framework for transformative research in special education. *Multiple Voices for Ethnically Diverse Exceptional Learners*, 13(2), 32–47. <https://doi.org/10.5555/muvo.13.2.yv7822w58116kw42>
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71(2), 149–164. <https://doi.org/10.1177/001440290507100202>
- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333(6045), 1015–1019. <https://doi.org/10.1126/science.1196783>
- Goldberg, L. R., Lee, K., & Ashton, M. C. (2008). Comment on Anderson and Ones (in press). *European Journal of Personality*, 22(2), 151–156. <https://doi.org/10.1002/per.663>
- Grahe, J. (2018). Another step towards scientific transparency: Requiring research materials for publication. *Journal of Social Psychology*, 158(1), 1–6. <https://doi.org/10.1080/00224545.2018.1416272>
- Graves, S. L., Phillips, S., Jones, M., & Johnson, K. (2021). A systematic review of the What Works Clearinghouse's behavioral intervention evidence: Does it relate to Black children? *Psychology in the Schools*, 58(6), 1026–1040. <https://doi.org/10.1002/pits.22485>
- Haroz, S. (2022). Comparison of preregistration platforms. MetaArXiv. <https://doi.org/10.31222/osf.io/zry2u>
- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451–470. <https://doi.org/10.1111/bmsp.12028>
- Hernández-Saca, D. I., Gutmann Kahn, L., & Cannon, M. A. (2018). Intersectionality disability research: How disability research in education engages intersectionality to uncover the multidimensional construction of disabled experiences. *Review of Research in Education*, 42(1), 286–311. <https://doi.org/10.3102/0091732X18762439>
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177. <https://doi.org/10.1111/j.1750-8606.2008.00061.x>
- Hodes, R. J., Insel, T. R., & Landis, S. C. (2013). The NIH toolbox: Setting a standard for biomedical research. *Neurology*, 80(11 Suppl. 3), S1–S1. <https://doi.org/10.1212/WNL.0b013e3182872e90>
- Institute of Education Sciences. (2019). *Standards for excellence in education research*. <https://ies.ed.gov/seer>
- Institute of Education Sciences. (n.d.). *IES policy regarding public access to research*. <https://ies.ed.gov/funding/researchaccess.asp>
- Kendall, P. C., Olino, T. M., Carper, M., & Makover, H. (2017). On the importance of temporal precedence in mediational analyses. *Journal of Consulting and Clinical Psychology*, 85(1), 80–82. <https://doi.org/10.1037/ccp0000152>
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., Ijzerman, H., Nilsson, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4(1), 20. <https://doi.org/10.1525/collabra.158>
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- Logan, J. A., Hart, S. A., & Schatschneider, C. (2021). Data sharing in education science. *AERA Open*, 7, 1–11. <https://doi.org/10.1177/23328584211006475>
- Magua, W., Zhu, X., Bhattacharya, A., Filut, A., Potvien, A., Leatherberry, R., Lee, Y.-G.,

- Jens, M., Malikireddy, D., Carnes, M., & Kaatz, A. (2017). Are female applicants disadvantaged in National Institutes of Health peer review? Combining algorithmic text mining and qualitative methods to detect evaluative differences in R01 reviewers' critiques. *Journal of Women's Health, 26*(5), 560–570. <https://doi.org/10.1089/jwh.2016.6021>
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. *Educational Researcher, 50*(8), 493–504. <https://doi.org/10.3102/0013189X211001356>
- Mbuagbaw, L., Aves, T., Shea, B., Jull, J., Welch, V., Taljaard, M., Yoganathan, M., Greer-Smith, R., Wells, G., & Tugwell, P. (2017). Considerations and guidance in designing equity-relevant clinical trials. *International Journal for Equity in Health, 16*(1), 1–9. <https://doi.org/10.1186/s12939-017-0591-1>
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science, 1*(1), 131–144. <https://doi.org/10.1177/2515245917747656>
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*(1), 1–9. <https://doi.org/10.1038/s41562-016-0021>
- Murray, D. L., Morris, D., Lavoie, C., Leavitt, P. R., MacIsaac, H., Masson, M. E., & Villard, M. A. (2016). Bias in research grant evaluation has dire consequences for small universities. *PLOS ONE, 11*(6), e0155876. <https://doi.org/10.1371/journal.pone.0155876>
- National Academies of Sciences, Engineering, and Medicine. (2022). *The future of education research at IES: Advancing an equity-oriented science*. National Academies Press. <https://doi.org/10.17226/26428>
- Nilsen, P. (2015). Making sense of implementation theories, models, and frameworks. *Implementation Science, 10*(1), 53. https://doi.org/10.1007/978-3-030-03874-8_3
- Pek, J., Hoisington-Shaw, K. J., & Wegener, D. T. (2022). Avoiding questionable research practices surrounding statistical power analysis. In W. O'Donohue, A. Masuda & S. Lilienfeld (Eds.), *Questionable research practices in applied psychology* (pp. 243–267). Springer.
- Petkovic, J., Jull, J., Yoganathan, M., Dewidar, O., Baird, S., Grimshaw, J. M., & Welch, V. (2020). Reporting of health equity considerations in cluster and individually randomized trials. *Trials, 21*(1), 1–12. <https://doi.org/10.1186/s13063-020-4223-5>
- Pohlhaus, J. R., Jiang, H., Wagner, R. M., Schaffer, W. T., & Pinn, V. W. (2011). Sex differences in application, success, and funding rates for NIH extramural programs. *Academic Medicine, 86*(6), 759–767. <https://doi.org/10.1097/ACM.0b013e31821836ff>
- Redding, C. (2019). A teacher like me: A review of the effect of student-teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of Educational Research, 89*(4), 499–535. <https://doi.org/10.3102/0034654319853545>
- Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American Educational Research Journal, 56*(1), 178–203. <https://doi.org/10.3102/0002831218798325>
- Schmitt, M. B., Logan, J. A. R., Tambyraja, S. R., Farquharson, K., & Justice, L. M. (2017). Establishing language benchmarks for children with typically developing language and children with language impairment. *Journal of Speech, Language, and Hearing Research, 60*(2), 364–378. https://doi.org/10.1044/2016_JSLHR-L-15-0273
- Schneider, M. (2020). *Making common measures more common*. Institute of Education Sciences. <https://ies.ed.gov/director/remarks/5-05-2020.asp>
- Shogren, K. A., Luckasson, R., & Schalock, R. L. (2014). The definition of “context” and its application in the field of intellectual disability. *Journal of Policy and Practice in Intellectual Disabilities, 11*(2), 109–116. <https://doi.org/10.1111/jppi.12077>
- Steinbrenner, J. R., McIntyre, N., Rentschler, L. F., Pearson, J. N., Luelmo, P., Jaramillo, M. E., Boyd, B. A., Wong, C., Nowell, S. W., Odom, S. L., & Hume, K. A. (2022). Patterns in reporting and participant inclusion related to race and ethnicity in autism intervention literature: Data from a large-scale systematic review of evidence-based practices. *Autism*. Advance online publication. <https://doi.org/10.1177/13623613211072593>
- Tipton, E., & Olsen, R. B. (2022). *Enhancing the generalizability of impact studies in education* (NCEE 2022-003). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <http://ies.ed.gov/ncee>

- Valentine, J. C., & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. What Works Clearinghouse.
- van Dijk, W., Norris, C. U., Al Otaiba, S., Schatschneider, C., & Hart, S. A. (2022). Exploring individual differences in response to reading intervention: Data from Project KIDS (Kids and Individual Differences in Schools). *Journal of Open Psychology Data*, 10(1), 2. <http://doi.org/10.5334/jopd.58>
- van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. *Journal of Learning Disabilities*, 54(2), 139–152. <https://doi.org/10.1177/0022219420945267>
- Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, 24(1), 94–97. <https://doi.org/10.1016/j.cub.2013.11.014>
- Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities*, 51(6), 612–624. <https://doi.org/10.1177/0022219418775110>
- Welch, V. A., Norheim, O. F., Jull, J., Cookson, R., Sommerfelt, H., & Tugwell, P. (2017). CONSORT-Equity 2017 extension and elaboration for better reporting of health equity in randomised trials. *BMJ*, 359, <https://doi.org/10.1136/bmj.j5085>
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726. <https://doi.org/10.1037/0003-066X.61.7.726>
- Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., & Tu, X. M. (2019). Post hoc power analysis: Is it an informative and meaningful analysis? *General Psychiatry*, 32(4). <https://doi.org/10.1136/gpsych-2019-100069>

ORCID iDs

Jessica R. Toste  <https://orcid.org/0000-0002-6327-0054>

Karrie A. Shogren  <https://orcid.org/0000-0001-7925-1299>

Manuscript received March 2022; accepted December 2022.