Editorial

What's Your Evidence? Making Evidence-Based Claims and Why This Matters

Kristen Bieda and Sandra Crespo
Michigan State University

"The education system in the United States is broken."

"Reach all students with this new technology!"

"Get kids to pay attention and learn!"

These are a sampling of the kinds of headlines that reach our email in-boxes on a weekly basis. The widespread use of Twitter (#iteachmath) and blogs (#mtbos) brings prospective and in-service teachers unprecedented access to knowledge and guidance that can inform teaching, but the sheer volume of information available comes at a cost. The cost is that authors feel that they have to entice readers with catchier titles and bolder claims, a phenomenon that is referred to in the popular media as *clickbait*. As we are learning from our current political climate, our U.S. culture may be becoming increasingly entranced with compelling headlines and less engaged with the evidence provided to support those headlines.

In academic circles, we like to assert that we are immune to clickbait, and we certainly would not try to employ clickbait in our *own work*, but for many potential *Mathematics Teacher Educator* (*MTE*) authors, one of the most challenging aspects of writing a manuscript is framing a claim appropriate for the evidence they have. This issue of overstating claims has roots in some efforts that have emerged in the past few decades to question the value of academic research. These efforts have become manifest as initiatives to curate research that meets a particular set of standards for validity and rigor (e.g., What Works Clearinghouse), the difficulties scholars face in publishing replication or validity studies (**https://www.editage.com/insights/why-are-replication-studies-so-rarely-published**), and the greater emphasis being placed on bibliometrics (e.g., impact factors like the h-index) that encourage scholars to present bold, transformative claims.

Our responsibility as scholars is to avoid making sensational appeals and rather take great care to establish the trustworthiness of our scholarship, especially when our claims could directly inform the work of teachers and the learning of students. So how do we do this? We start by knowing about the range of different types of claims we can make as well as what evidence aligns best with the claims we seek to make.

Different Types of Claims

In her March 2013 *MTE* editorial, Peg Smith reminded the field of Lester and Williams's (2000) Commonsense Principle for Connecting Evidence and Claims: "Essentially, the principle says that for a body of Information (e.g., findings from an investigation) to be considered evidence for a claim, it must hold for a nonempty subset of the domain to which the claim applies" (p. 105). What goes unsaid here is how different types of claims require different kinds of evidence. In a 2016 report written for the Mathematica Center for Improving Educational Research, Chojnacki, Resch, Vigil, Martinez, and Bates outline four different types of evidence often presented in various educational media (from popular news to marketing to research articles and policy briefs): anecdotal, descriptive, correlational, and causal.

Anecdotal evidence consists of judgments from personal experience. Such judgments cannot be generalized to the experience of other individuals, but they can be used to make claims that an innovation or effect may be worth further study or investigation. In some ways, this is like the practice that mathematicians often follow of testing some example cases to determine whether a conjecture might be valid. Similarly, descriptive evidence can be obtained through a systematic process of gathering outcomes over time, but the evidence does not tell us more than what a particular group experienced or learned over time. There is no way to establish whether the program or innovation was the driver of the outcomes or whether some other factor or set of factors contributed to the change or improvement noted. However, descriptive evidence can, like anecdotal evidence, help determine what approaches warrant further research and investigation, as well as shed light on the features of the approach that drive the change or effect noted.

Mathematics Teacher Educator typically publishes work that draws primarily on correlational and causal evidence to support claims because, as Smith (2013) reminds us, "the extent to which the articles that appear in this journal influence the practice of teacher education depends in large measure on the extent to which the claims made are supported with compelling evidence and convincing

arguments" (p. 107). According to Chojnacki and colleagues (2016), correlational evidence can show an effect of an approach or innovation, but data collection did not systematically measure or account for effects of other variables on the outcome (e.g., prior knowledge, selection bias). Many, if not all, the research studies conducted in teacher education result in correlational evidence. Most research conducted within the context of a methods or content course in a teacher preparation program is inherently correlational because the teacher preparation program curriculum is not standardized in the United States, and one program likely does not have enough teacher candidates to create "control" group conditions during the data collection period. Moreover, even with a large teacher preparation program, the mentoring and experiences that teacher candidates receive in field placement components may be a likely and influential source of variation and one that is difficult to capture feasibly at scale.

So do we throw in the towel? Do we give up because of the impossibility of generating "gold standard" causal evidence? Thankfully, *MTE* authors have been brave enough to say no! The heart of the *MTE* mission is to be a repository where knowledge about the practice can be made public, shared, and stored, as well as verified and improved over time (Hiebert, Gallimore, & Stigler, 2002). This means that, as teacher educators, we should not hold onto information about some potentially promising approaches until we have the "gold standard," causal evidence to support robust claims. Instead, we temper our claims and share the evidence we have that suggests that our approaches warrant other teacher educators considering their use, verifying their value, and (most important) improving them over time.

The Bieda, Cavanna, and Ji 2015 *MTE* manuscript titled "Enhancing Learning in Field Experiences Through Mentor-Guided Lesson Study" was initially submitted with the following research questions presented as the focus of the study: "To what extent does mentor-guided lesson study [MSGL] support PST-Ms [prospective secondary teachers-mentors'] reflections about teaching that promote development of MKT [mathematical knowledge for teaching]? What factors of mentor-guided lesson study influence the frequency and quality of reflections about practice that promote development of MKT?" The initial manuscript made claims such as the following:

> **These results suggest that the MGLS experience does help focus PST-Ms' attention upon student thinking during study lessons and, thus, may support the development of knowledge of content and students (Ball, Thames, & Phelps, 2008). The results also suggest that PST-Ms' capacity for noticing aspects of instruction related to students' thinking and key teaching moves increases in Cycle 2 with more experience in doing MGLS.**

However, the reviewer feedback questioned the trustworthiness of these claims because the authors did not have baseline evidence for PST-Ms' noticing prior to the Lesson Study Cycle 1.

The reviewers' feedback prompted a lot of introspection about the nature of the evidence Bieda and colleagues had collected to measure the effectiveness of the mentor-guided lesson study approach, and it informed changes to both the research questions and the claims asserted in a revised (and eventually published) manuscript. Specifically, the second question was dropped, and the first question was reworded to read, "To what extent does mentor-guided lesson study support PSTs in noticing features of instruction relevant to developing knowledge for mathematics teaching?" (p. 20). In the previous version of the manuscript, the question of identifying factors in the mentor-guided lesson study approach demanded that the authors present baseline evidence to better identify what changed in PST noticing. In the published version, a focus on the extent to which a particular kind of noticing (related to development of knowledge for mathematics teaching) emerged in the reflections during mentor-guided lesson study could be addressed by having only descriptive evidence of PSTs' reflections during the lesson study cycles.

As discussed above, one way to address *MTE*'s manuscript criteria for presenting evidence that warrants the claims being made is to carefully word claims to fit the evidence the authors have available. Another, complementary, rhetorical move is to make clear the limitations of the inquiry you have carried out. Smith's 2013 *MTE* editorial highlights the work of Groth (2012) as a model for this kind of move in an *MTE* article.

In this current *MTE* issue, we see the authors make different kinds of claims and use complementary kinds of evidence aligned with those claims. These articles also lay bare the weaknesses of their research, putting into perspective the generalizability of their claims. Although counterintuitive to novice researchers, this kind of rhetorical move enhances the trustworthiness of the research. Both Spitzer and Phelps-Gregory (2018, this issue) and Fukawa-Connelly, Klein, Silverman, and Shumar (2018, this issue) feature research into ways that online spaces for teacher education promote teacher learning, with the former in the context of supporting preservice teachers' abilities to analyze teaching, and the latter in the context of supporting teachers in an online professional development format to notice and wonder about student thinking.

Spitzer and Phelps-Gregory, in particular, use both quantitative and qualitative analytic tools to assess the effectiveness of the online format; however, the evidence presented is descriptive of PSTs' work throughout the course. These authors model effective ways to temper their claims about the likelihood that the intervention would be effective in other contexts ("This shift toward a critical view of the lesson provides some tentative evidence for the success of the intervention" [p. 34]) as well as the importance of sharing both significant and nonsignificant results. Although there appeared to be significant effects when responses from all participants were included, further analyses showed that these effects were attributed predominantly to the subgroup of PSTs who had a particular initial response to the intervention. This particular case shows that knowing when a result is nonsignificant has important applications for the practice of mathematics teacher education.

The articles by Tan and Thorius (2018, this issue) and Rubel and Stachelek (2018, this issue) focus on professional development experiences for practicing teachers aimed at supporting equity goals in the mathematics classroom. Tan and Thorius focus their article on inclusive practices for special education students in the mathematics classroom, whereas Rubel and Stachelek focus their professional development project on helping teachers diversify student participation opportunities. Both sets of authors designed professional learning tools that they shared with the teacher participants and that documented the effects of their intervention. Neither of these author teams goes on to make wild claims about the kinds of effects the professional development program had on the teachers in the program. Tan and Thorius make it clear that their results should be interpreted with caution as their results are localized in a particular place, time, and context. However, by describing portions of conversations that were made possible by their tool and how it supported teachers to shift from using deficit to strength language, Tan and Thorius show restraint when making claims about the success of their professional development program. Rubel and Stachelek write about two contrasting case studies and explain that they "selected Lucy and Teresa for case studies because Lucy had the largest change in mean DPP [Difference in Participation Proportion] and Teresa's DPP means remained consistent and unchanged" (p. 13). By exploring what happened with a teacher whose teaching practice (as represented by DPP means) remained unchanged, these authors are able to not only add credibility and trustworthiness to their analysis but also help the field learn to expect these kinds of results and to understand the importance of exploring and reporting them as well.

Although the focus of this editorial has been on the authors' practice of aligning claims and evidence, the recent attack on Rochelle Gutiérrez's work illustrates that, when making claims that critique scholarly work, we also have a responsibility to ensure that those claims are based on careful reading of, and evidence from, the work. We are fortunate to hear directly from Gutiérrez (2018, this issue) about the implications of the degradation of scholarly critique and how we can respond to unfounded attacks on our research.

We hope that this editorial promotes more conversation among teacher educators about how and when we share what we are learning from systematic inquiry into our practice. Given the typical publication cycle, waiting to publish what was once "exploratory" and "ground-breaking" new approaches until we have the kind of evidence that can support causal claims realistically means that we end up learning about this "new" work when it is at least several years old. Taking a stance that we are willing to present work with limited evidence (and making limited claims) may expose our scholarship to attacks from those outside our field about the validity and relevancy of the scholarship in the field of mathematics teacher education. It is therefore important to be clear about who our audience is and what our goals are for publishing our research. The *MTE* audience includes those who have "boots on the ground" doing the work of mathematics teacher education. As important as it is for scholars to speak directly to the public at large, *MTE* is not designed to do so. This publication is not meant to be a compendium of "best practices" to be used for public consumption. As we continue to grow and expand our knowledge base, it will be important for us individually and collectively to learn how to reach and speak to nonacademics about the work we do and the nature of the claims we make. As Gutiérrez's commentary proposes, we hope this will be the start of many more conversations.

References

Bieda, K. N., Cavanna, J., & Ji, X. (2015). Mentor-guided lesson study as a tool to support learning in field experiences. *Mathematics Teacher Educator, 4*(1), 20–31.

Chojnacki, G., Resch, A., Vigil, A., Martinez, I., & Bates, S. (2016). Understanding types of evidence: A guide for educators. Mathematica Policy Research Reports, Mathematica Policy Research.

Fukawa-Connelly, T., Klein, V., Silverman, J., & Shumar, W. (2018). An online professional development model to support teachers' abilities to examine student work and thinking. *Mathematics Teacher Educator, 6*(2), 39–51.

Groth, R. E. (2012). The role of writing prompts in a statistical knowledge for teaching course. *Mathematics Teacher Educator, 1*(1), 23–40.

Gutiérrez, R. (2018). When mathematics teacher educators come under attack. *Mathematics Teacher Educator, 6*(2), 68–74.

Hiebert, J., Gallimore, R., & Stigler, J. W. (2002). A knowledge base for the teaching profession: What would it look like and how can we get one? *Educational Researcher, 31*(5), 3–15.

Lester, F. K., & Wiliam, D. (2000). The evidential basis for knowledge claims in mathematics education research. *Journal for Research in Mathematics Education, 31*, 132–137.

Rubel, L., & Stachelek, A. (2018). Tools for rethinking classroom participation in secondary mathematics. *Mathematics Teacher Educator, 6*(2), 8–25.

Smith, M. S. (2013). Linking claims and evidence. *Mathematics Teacher Educator, 1*(2), 105–107.

Spitzer, S., & Phelps-Gregory, C. (2018). Using online collaboration to improve prospective teachers' analysis of teaching. *Mathematics Teacher Educator, 6*(2), 26–38.

Tan, P., & Thorius, K. K. (2018). En/countering inclusive mathematics education: A case of professional learning. *Mathematics Teacher Educator, 6*(2), 52–67.

Authors

Sandra Crespo, Teacher Education and Mathematics Education, Department of Teacher Education, College of Education, 620 Farm Lane, Room 362, Erickson Hall, Michigan State University, East Lansing, MI 48824-1034

Kristen Bieda, Teacher Education Associate Director of Mathematics, CREATE for STEM Institute, Michigan State University, 317 Erickson Hall, East Lansing, MI 48824