EDITORIAL

Linking Claims and Evidence

Margaret S. Smith
Editor, Mathematics Teacher Educator

Building a trustworthy knowledge base for mathematics teacher education—the mission of *Mathematics Teacher Educator*—requires that manuscripts convey more than stories of practice, however compelling. Manuscripts must include evidence of the effectiveness of the intervention being described beyond anecdotal claims or personal intuitions. As the Editorial Panel articulated in the call for manuscripts, "the nature of evidence in a practitioner journal is different from that in a research journal, but evidence is still critically important to ensuring the scholarly nature of the journal. Thus, authors must go beyond simply describing innovations to *providing evidence of their effectiveness*. Note that *effectiveness* implies that something is *better* and not just *different* as a result of the innovation." Hence, claims must be supported by evidence. In this editorial, I discuss the nature of evidence appropriate for articles in *Mathematics Teacher Educator*.

In an essay that appeared in the *Journal of Research in Mathematics Education*, Lester and Wiliam (2000) proposed a Commonsense Principle for Connecting Evidence and Claims (CSP). Essentially the principle says that for a body of information (i.e., 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. To illustrate this principle, consider the following example drawn from the article "Mathematics Preservice Teachers Learning About English Language Learners Through Task-Based Interviews and Noticing," by Anthony Fernandes (2012), which appeared in the first issue of *Mathematics Teacher Educator*. Fernandes claimed that the interview experience in which preservice middle school teachers engaged had the potential for helping the PSTs (1) to notice the linguistic challenges faced by ELL students and the resources that ELL students used to communicate mathematically and (2) to develop concrete strategies that assisted ELL students and were aligned with best practices advocated in the research. Through an analysis of written reports in which PSTs responded to a set of questions after completing interviews with ELL students, Fernandes found that all 31 PSTs brought up the linguistic challenges that the ELL students faced during the interviews, most of the PSTs reported that the concrete materials were a resource that ELL students employed to understand and communicate their thinking, and PSTs had developed strategies for assisting ELL students. Excerpts from PSTs' reports were used to give the reader insight into what they said and did that supported this finding, and the few cases of disconfirming evidence were also presented and discussed. In addition, Fernandes triangulated these findings by examining portions of the videotape from the interviews, looking at responses to other questions, and checking with PSTs as needed to ensure that the PSTs' interpretations were grounded in their interactions.

Are the findings from his analysis evidence for the claim? In terms of the Commonsense Principle, we would consider preservice middle school teachers to be the domain, and the 31 PSTs enrolled in the content methods course for preservice middle school teachers who were subjects of the study to be a subset of domain. Since the researcher showed that the findings from his analysis hold for the participants in his study, we would say that the evidence supports the researcher's claim. It is worth noting that the excerpts from the reports used to illustrate what the PSTs did and said *alone* would not be sufficient evidence to support the claim, since pointing to specific cases that confirm the desired outcome (i.e., cherry picking) may ignore cases that contradict the position. In the Fernandes case, the excerpts are used to support and provide insight into the more general findings regarding the value of the interview experience (noted above), and cases of teachers who did not follow the more general pattern were discussed.

Lester and Wiliam (2000) argue that convincing others of knowledge claims is more than just presenting evidence that supports the claims; "it is also a matter of persuading them to accept the values the researcher holds about the objects and phenomena being studied as well as about the very purpose of research itself" (p. 136). In the case of the Fernandes example, a reader may be more persuaded by the findings if she identifies with the problem that lead to the innovation—PSTs' lack of preparation for working with diverse populations of students in general and ELL students in particular—and sees value in having PSTs conduct task-based interviews with K–12 students. Although not every reader will be compelled by the claims and related evidence provided in every manuscript that appears in the journal, authors need to consider whether the claims, evidence, and the argument (i.e., rationale for the overall importance and value of the research) are sufficiently compelling to cause readers (other mathematics teacher educators) to consider replicating or modifying the described innovation/intervention in their own contexts.

Data Sources for Claims

In the Fernandes example, data consisted primarily of written reports in which PSTs responded to a set of guiding questions (informed by Mason's framework of noticing) after conducting each of two interviews with ELL students. In this case, the data were a byproduct of the intervention itself and were carefully analyzed by the researcher to determine the extent to which PSTs were able to notice the challenges ELLs faced, identify resources on which ELL students drew in communicating their mathematical understandings, and employ appropriate strategies for supporting ELL students. As Fernandes indicated:

> **To document the impact of the intervention with respect to this goal, I initially focused on the PSTs' responses to questions 3, 4, and 9 (see Figure 3). I created a separate document that compiled each of the 31 PSTs' responses from both reports for these three questions and used this as the starting point for examining the impact of the intervention. I specifically looked at the linguistic challenges that the PSTs described and the resources the PSTs mentioned that the ELL students used in connection to these challenges. I triangulated these points with their responses to other questions, particularly the descriptions they provided in response to questions 1 and 2. I also examined portions of the videotape where they were interacting with the ELL students to ensure that their interpretation was grounded in their interactions. Further, I had close interactions with all the PSTs during the project, and during the feedback process I clarified my interpretation of their statements. (Fernandes, 2012, p. 16)**

PSTs' responses to particular questions served as the primary data source for the analysis. These responses were triangulated with their responses to other questions, the videotape record of the interaction, and direct interaction with the students. Through the analysis of the videotaped interactions between PSTs and ELL students, Fernandes was able to provide direct evidence that the PSTs actually developed new teaching strategies during the interview experience rather than relying solely on teacher's self-reports of what they did. In general, triangulation allows the researcher to have more confidence in his claims by having more than one data source that produced the same result.

As teacher educators who engage PSTs and practicing teachers in a range of activities that are intended to develop some aspect of their knowledge base for teaching, the careful analysis of the work produced around engagement in these activities (e.g., videotapes of discussions, written work, reflections on the experiences) can be a valid source of data for supporting the claims we make. The key, as exemplified in the Fernandes article, is to carefully analyze the data, and, when possible, use more than one source of data to triangulate the findings. In addition, it is critical not to claim more than what can be supported by the evidence. For example, Fernandes did not claim that the PSTs in his course were better prepared to teach ELL students. Although we may hope that this is indeed the case, no evidence was presented that indicated how the interview experience influenced PSTs' teaching directly. Also, we need to be clear about whether we have direct evidence of subjects' knowledge and/or teaching skills/practice, rather than self-reports or statements of their beliefs about their knowledge or practice. Beliefs about an important topic in mathematics education may serve as valuable evidence for certain claims; beliefs about one's own knowledge or practice are not strong evidence.

Although the work produced by teachers during an intervention is one source of data that can be used to produce evidence to support claims, it is not the only possibility. Consider for example the article titled, "The Role of Writing Prompts in a Statistical Knowledge for Teaching Course" by Randall Groth (2012) that appeared in the first issue of *Mathematics Teacher Educator*. In the article, Groth describes an intervention (course) that was intended to develop teachers' knowledge base (broadly defined) for teaching statistics that included writing prompts as a key feature. He administered two measures—a statistics test developed by the Learning Mathematics for Teaching (LMT) project and the Comprehensive Assessment of Outcomes in a First Statistics Course (CAOS)—at the beginning and end of the course in order to determine what PSTs learned. Groth noted that results from the LMT and CAOS tests indicated that prospective teachers made notable progress toward learning goals for the course, and he provided appropriate statistics to show that the pre- to post-test differences were significant.

The strength of using instruments such as the LMT and CAOS is they have been previously used, and in some cases validated by others, and therefore it may be possible to compare the results with other samples, as was the case with the CAOS. Such evidence may be particularly compelling, but it is not always possible to find an existing instrument that is well aligned with what you are trying to measure.

The Groth article raises another issue that is important to consider. Although LMT and CAOS showed gains in teachers' knowledge of statistics, Groth acknowledges that these gains cannot be attributed directly to the

writing prompts that were at the core of the invention. He indicates ". . . the scores provide evidence that the writing prompts can play a prominent role in courses that build both subject matter knowledge and pedagogical content knowledge." He also makes clear that the analysis he performed on PST responses to the writing prompts using the Structure of the Observed Learning Outcome (SOLO) taxonomy provided insight into prospective teachers' thinking at various time points, and that there were limitations in the claims that could be made based on these data alone. He explains:

> **… it was not feasible to track changes in SOLO levels across tasks because the sets of tasks all dealt with different statistical content. Hence, any changes in the level of response seem just as easily attributable to the difficulty of the content as they would be to general cognitive gains in SKT. SOLO analyses could, however, be used to track learning gains if similar sets of tasks were administered periodically throughout a course. (Groth, 2012, p. 33)**

Hence, Groth is careful not to link improvements in scores directly to the success of the intervention and to articulate what claims could and could not be made based on the nature of the writing responses.

Conclusion

In making claims and supporting them with evidence, it is important to consider the audience. For the purposes of this journal, the evidence needs to be strong enough to convince a colleague that a particular intervention or innovation is worth considering. Such evidence might not be compelling enough to influence policy or large-scale reforms, and may not satisfy the readers of research journals who are looking for larger sample sizes and more methodological rigor, but our purpose here is different and so, too, is our standard for evidence.

My intent here is to begin a conversation regarding what counts as reasonable evidence for articles that appear in *Mathematics Teacher Educator* and the importance of limiting the claims that are made to those that can be supported with the evidence available. 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.

References

Fernandes, A. (2012). Mathematics preservice teachers learning about English language learners through task-based interviews and noticing. *Mathematics Teacher Educator, 1,* 10–22. http://www.nctm.org/publications/article.aspx?id=33983

Groth, R. E. (2012). The role of writing prompts in a statistical knowledge for teaching course. *Mathematics Teacher Educator, 1,* 23–40. http://www.nctm.org/publications/article.aspx?id=33985

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.

Author

Margaret S. Smith, Department of Instruction and Learning, University of Pittsburgh, Pittsburgh, PA 15260; pegs@pitt.edu