Journal of Mathematical Behavior 45 (2017) 111–120

Contents lists available at ScienceDirect

The Journal of Mathematical Behavior

journal homepage: www.elsevier.com/locate/jmathb

Inspecting the foundations of claims about cognitive demand and student learning: A citation analysis of Stein and Lane (1996)

Samuel Otten[a,*], Corey Webel[b], Zandra de Araujo[b]

[a] University of Missouri, 121 B Townsend Hall, Columbia, MO 65211, USA
[b] University of Missouri, USA

ARTICLE INFO

Article history:
Received 25 August 2016
Received in revised form
17 December 2016
Accepted 23 December 2016
Available online 8 January 2017

Keywords:
Mathematical tasks
Cognitive demand
Literature review
Citation analysis
Student learning
Curriculum
Replication studies

ABSTRACT

Educational research communities bear responsibility for establishing a substantial body of evidence to support claims that drive the field. For example, one commonly accepted claim is that there is a relationship between the cognitive demand of mathematical task enactments and students' learning. One study that is often cited in association with this claim is Stein and Lane (1996), and in 44% of those citations, Stein and Lane (1996) is the sole reference. Citation analysis reveals that many of these claims go beyond the warrants provided by the Stein and Lane study, either by granting more confidence in the relationship than the study design allows or by phrasing the claim as a causal relationship between cognitive demand and student learning. A few other studies are occasionally cited in conjunction with Stein and Lane (1996) and are summarized in this article, but there remains a need for replication studies to provide better empirical support for claims about cognitive demand and student learning and to refine our shared understanding.

Educational research faces profound challenges, according to the National Research Council (2002). One challenge is that educational practice is tied to strong societal and personal values, and "social ideals inevitably influence the research that is done, the way it is framed and conducted, and the policies and practices that are based on research findings" (p. 17). These values are not only held by policymakers and members of communities, but also by the educational researchers themselves. In mathematics education, specifically, there are those who value skill repetition and automaticity, those who value authentic problem solving and reasoning, and those who value both in equal measure. In some sense, the importance of these mathematical experiences is a matter of opinion, but there are also empirically-testable claims that arise. For example, one may claim that higher levels of cognitive demand of mathematical tasks (and task enactments) (Henningsen & Stein, 1997; Stein, Grover, & Henningsen, 1996) is predictive of (or perhaps causes) positive student learning outcomes, including higher performance on achievement tests.

We suspect that this claim about cognitive demand and student learning is held by a large number of researchers in mathematics education, and indeed the claim may very well be true, but as the National Research Council (2002) pointed out, another challenge for education research is to establish robust empirical warrants for our central claims, relying on a variety of perspectives and methodologies, and to actively seek out discrediting evidence. We, as a field, are not always diligent in pursuing discrediting evidence or buttressing our warrants; for example, replication studies are rarer in education research than in other fields (Makel & Plucker, 2014). With regard to the warrants for cognitive demand and student learning, in our

* Corresponding author.
E-mail addresses: ottensa@missouri.edu (S. Otten), webelcm@missouri.edu (C. Webel), dearaujoz@missouri.edu (Z. de Araujo).

http://dx.doi.org/10.1016/j.jmathb.2016.12.008
0732-3123/

own past work (e.g., Otten, 2012; de Araujo, 2012), we noticed extensive literature on the nature of cognitive demand (Doyle, 1983; Stein et al., 1996; Stein, Smith, Henningsen, & Silver, 2009) and factors influencing cognitive demand throughout mathematical task implementations (Boston & Smith, 2009; Jackson, Garrison, Wilson, Gibbons, & Shahan, 2013; Wilhelm, 2014), but a weaker empirical foundation for the direct link between cognitive demand and student learning. Nevertheless, the link is extremely important for practitioners and indeed provides much of the justification for studying the construct of cognitive demand itself.

The hypothesis motivating this study is that a large portion of the warrants for claims in the mathematics education literature about the link between cognitive demand and learning depend on a single reference—Stein and Lane (1996). If this hypothesis were true, then it would become imperative to critically analyze the research design, evidence, and claims made in Stein and Lane (1996) and to consider possibilities of replication, especially given that a modified replication (Otten, 2012) failed to verify the findings. Using citation analysis techniques similar to Leatham and Winiecke (2014), we identified peer-reviewed articles and examined the claims for which Stein and Lane (1996) was included as a citation, and we identified other references, if they existed, that were also cited for those same claims. One might note that this study takes as its starting point the citations to Stein and Lane (1996) rather than all claims about cognitive demand and learning because it is unfeasible to systematically capture the latter and because, as discussed below, mainstream sources in mathematics education tend to at least cite Stein and Lane (1996) in relation to cognitive demand and learning.

In the following sections, we briefly summarize the Stein and Lane (1996) study, describe our method for compiling and analyzing the citations to Stein and Lane (1996), and then share our key results. We also discuss what this analysis reveals with regard to the foundation for claims about cognitive demand and student learning and we share implications of this work.

1. Stein and Lane (1996)

1.1. Summary

Stein and Lane (1996) was published in *Educational Research and Evaluation*, which is not a journal specific to mathematics education, but it stems from a project well known in mathematics education—Quantitative Understanding: Amplifying Student Achievement and Reasoning (QUASAR). QUASAR was funded by the Ford Foundation in the mid-1990s and was based at the University of Pittsburgh (see Silver & Stein, 1996; for an overview). Six urban middle schools participated in QUASAR with the overall goal of promoting mathematics instruction aligned with recommendations from the National Council of Teachers of Mathematics (NCTM) (1989, 1991) and to investigate whether such ambitious instruction was feasible and responsible in schools with a history of poor mathematics performance.

Other works from QUASAR presented the construct of cognitive demand and the Mathematical Tasks Framework (Henningsen & Stein, 1997; Stein et al., 1996), but Stein and Lane (1996), specifically, had the following purpose:

> The purpose of this paper is to present a conceptual framework for linking teaching and learning within the classrooms of teachers who are attempting to reform their instruction. . . and to present empirical evidence regarding the degree to which the presence of reform features of instruction are linked to increases in student understanding of mathematics. (p. 51, emphasis added)

Thus, the study was centrally concerned with warrants for the link between instructional features and student learning. The study focused on 4 of the 6 middle schools from the QUASAR project over a three-year period. Data consisted of narrative summary field notes and video recordings of three-day observation cycles in three teachers' classrooms in each school each year, a classroom observation instrument completed based on the field notes and video recordings, and Fall and Spring administrations of a project-developed assessment instrument (Lane, 1993). Mathematical tasks were identified in the observation data and the two tasks in each observation that received the largest amount of time were identified for further analysis. Of the 620 main tasks, a stratified random sample of 144 was drawn and the task set-up and task implementations of these 144 tasks were coded for cognitive demand based on "what *the majority of* students appeared to be doing" ((Stein & Lane, 1996; p. 65), emphasis in original). Levels of cognitive demand were collapsed from 6 (doing mathematics, procedures with connections, procedures without connections, memorization, unsystematic exploration, nonmathematical activity) to 2 (high and low). A 25% sample of the 144 tasks was double coded with 79% agreement.

The assessment instrument consisted of 36 open-ended tasks distributed into four forms (9 questions per form) and a 5-point scoring rubric (0–4) for each task. Analysis focused on 11 of the tasks and used not the scores themselves but "the average percentage of student responses across tasks that were scored at the two most proficient score levels (3 or 4)" (p. 68) and how this average percentage shifted between different time points (e.g., between Fall Year 1 and Spring Year 3). Stein and Lane (1996) hypothesized that sites where tasks were set-up and implemented at high levels would have relatively high learning gains compared to sites where tasks were set-up and implemented at low levels.

To generate their findings, the researchers rank ordered the four schools based on their gains in percentage of students at the top two levels of proficiency and then focused on Site A, which had gained the most (36%), and Site D, which had gained the least (17%). (The other two schools had gained 27% and 22% in the top two levels of proficiency.) They compared these learning gain rankings with the school profile for task enactments and no ted the following: