Learning and Instruction 45 (2016) 49–60

Contents lists available at ScienceDirect

Learning and Instruction

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

Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes?

Anthony J. Gabriele*, Elana Joram, Ki H. Park

University of Northern Iowa, United States

ARTICLE INFO

Article history:
Received 26 June 2015
Received in revised form
23 June 2016
Accepted 30 June 2016

Keywords:
Teacher knowledge
Judgment accuracy
Instructional decision-making
Metacognitive monitoring
Student achievement

ABSTRACT

In this study we investigated whether elementary mathematics teachers' knowledge of their students, as reflected in both the accuracy and confidence with which they are able to estimate their students' task-specific performance on sets of mathematics problems, predicted students' overall mathematics achievement. Thirty-nine teachers made predictions about the performance of a random sample of target students (n = 150) in their classrooms on sets of "easy" and "difficult" multiplication and division problems. Teachers also provided confidence ratings for those judgments. From these data, indicators of teachers' judgment accuracy, judgment confidence and calibration accuracy (a measure of metacognitive monitoring) were then related to all of their students' (n = 834) performance on year-end standardized mathematics achievement tests. Multilevel analyses indicate that teachers' calibration accuracy, but not their task-specific judgment accuracy, significantly predicted students' mathematics achievement. Implications for future research on teacher knowledge as well as professional development programs are discussed.

1. Introduction

1.1. Teacher judgment accuracy and teaching effectiveness

The accuracy with which teachers can gauge their students' understanding and skill is likely to play a significant role in the quality of their teaching (e.g., Baumert & Kunter, 2013). Teachers who know their students well are more likely to make wise instructional decisions that allow them to effectively adapt and adjust their teaching to the needs of their students in ways that ultimately favorably impact student learning (e.g., Vogt & Rogalla, 2009). For example, a teacher who can accurately rank order the reading skill level of his students is more likely to be able to make an instructionally effective decision such as placement in an appropriate reading group (Shavelson & Borko, 1979). Similarly, a teacher who can accurately estimate how well her students understand a mathematical concept will more likely provide an effective explanation that is neither overly complex nor too simplistic but instead builds on students' existing understanding, thereby increasing student learning (Wittwer & Renkl, 2008). As such, measures of teachers' knowledge of their students' competencies and characteristics – sometimes referred to as teachers' *judgment accuracy* - may provide a useful indicator of a teacher's level of professional development and effectiveness in particular instructional contexts.

Surprisingly, teachers' knowledge of their students, although generally recognized as an important characteristic of effective teachers, has received little empirical attention in the research literature linking teacher characteristics to student outcomes. The few studies that have looked at the relationship between teacher judgment accuracy and teaching effectiveness report mixed results. For example, in the area of mathematics, some researchers report a direct relationship between a teacher's judgment accuracy and the mathematics achievement of students in their classroom (Carpenter, Fennema, Peterson, & Carey, 1988; Thiede et al., 2015), whereas other studies report that the relationship between teacher judgment accuracy and student mathematics achievement outcomes is moderated by instructional variables (Helmke & Schrader, 1987), suggesting that certain instructional actions are effective

* Corresponding author. Department of Educational Psychology & Foundations, University of Northern Iowa, Schindler Education Center, Cedar Falls, IA 50614-0607, United States. fax: +3192737732.
E-mail address: gabriele@uni.edu (A.J. Gabriele).

http://dx.doi.org/10.1016/j.learninstruc.2016.06.008
0959-4752/

only when teachers are sufficiently accurate in their judgments about students' competencies. Finally, some researchers have failed to find any relationship – direct or moderated – between indicators of teacher judgment accuracy and student achievement in mathematics (Karing, Pfost, & Artelt, 2011). To complicate matters more, some researchers report findings that suggest that the relationship between teacher judgment accuracy and student mathematics achievement outcomes (i.e., direct or moderated by instructional variables) may depend upon the way teacher judgment accuracy is measured (Anders, Kunter, Brunner, Krauss, & Baumert, 2010; Artelt & Rausch, 2014). Given the hypothesized importance placed on teacher judgment accuracy in models of effective teaching (Baumert & Kunter, 2013), firmly establishing and understanding the relationship of indicators of teachers' knowledge of their students and student learning outcomes is essential.

In addition, other aspects of teacher judgment, such as teachers' confidence in the accuracy of their judgments about their students (i.e., judgment confidence) and the degree to which teachers' confidence in their judgments corresponds with the accuracy of their judgments (i.e., calibration of confidence judgments) have rarely been addressed in studies of teacher judgment accuracy. Yet, there are theoretical reasons to believe that teachers' confidence in their judgments about students may play a critical role in teachers' effective use of this knowledge that ultimately affects their ability to influence student learning outcomes (Praetorius, Berner, Zeinz, Scheunpflug, & Dresel, 2013). In this study, we empirically explore this possibility by examining the role that teacher judgment confidence may play in teachers' instructional effectiveness.

Before describing in more detail our specific research questions, we first briefly describe some of the differences in the methods that have been used to measure teacher judgment accuracy, and some of the arguments made for their validity. We then lay out in more detail two alternative perspectives on how teacher judgment confidence may influence teachers' instructional effectiveness.

1.2. Teacher judgment accuracy: how it has been studied and measured

In studies of teacher judgment accuracy teachers are asked to estimate, rank, or rate their students' performance on an assessment of that performance; these judgments are then related to students' actual performance on that assessment, collected within close temporal proximity to the teachers' judgments. To date, most studies of teacher judgment accuracy have focused on establishing *whether* teachers are accurate at judging their students' achievement or affective characteristics and exploring the factors which influence the accuracy of these judgments (e.g., Begeny, Krouse, Brown, & Mann, 2011; Feinberg & Shapiro, 2009; Mulholland & Berliner, 1992). In general, studies of teacher judgment accuracy report moderate to high correspondence between teachers' judgments of student performance and their students' actual performance, suggesting that teachers are by and large accurate (Südkamp, Kaiser, & Möller, 2012). However, the degree of accuracy can vary considerably across teachers (e.g., Helmke & Schrader, 1987), types of students judged (e.g., Feinberg & Shapiro, 2009), and measures used (e.g., Artelt & Rausch, 2014) demonstrating substantial individual differences in teacher judgment accuracy and the presence of a number of influential moderating variables (Südkamp et al., 2012).

Measures of teacher judgment accuracy have focused either on the class level (e.g., Anders, et al., 2010) or the individual level (e.g., Helmke & Schrader, 1987) and include judgments either about all students (e.g., Thiede et al., 2015) or a random sample of students from a teacher's classroom (e.g., Karing et al., 2011). Measures of teacher judgment accuracy have also varied with respect to the specificity of the object of judgment, with some studies focused on general aspects of student performance, such as student mathematics achievement on standardized tests (e.g., Coladarci, 1986) and others focused on more task-specific aspects of student performance such as solutions to a few carefully selected mathematical tasks that differ in their difficulty level (e.g., Carpenter, et al., 1988).

Of the different methods that have been used to compute teacher judgment accuracy, the most common is to compute the correlation between a measure of teachers' judgments of their students' performance and a measure of their students' actual performance, which reflects a teacher's knowledge of the *relative* standing of their students' performance on some criterion (Südkamp et al., 2012). For example, Behrmann and Souvignier (2013) asked teachers to rate the reading comprehension ability of every student from their respective class on a five-point Likert-type scale. For each teacher, a Spearman's rank-order coefficient was computed between a teacher's rating of each student's reading comprehension ability and each of their corresponding student's actual reading comprehension score on an 18-item standardized reading comprehension test. Each correlation coefficient served as that teacher's indicator of judgment accuracy.

Other indicators of teacher judgment accuracy, used less often by researchers, capture how close a teacher's estimate of their students' performance is to their students' actual performance (i.e., teacher judgment error) or how frequently a teacher accurately predicts their students performance (e.g., teacher prediction accuracy) in *absolute* terms. For example, to create a task-specific measure of teacher judgment error at the class level Brunner, Anders, Hachfeld, and Krauss (2013) asked teachers to estimate the percentage of students in their class who would correctly answer each of four mathematical tasks; they then computed the absolute difference between a teacher's estimate and the actual percentage of correct answers in the class on each of the four tasks. The mean task-specific judgment error across the four tasks was calculated, with a judgment error of zero indicating that a teacher estimated without error the percentage of correct solutions in their class on all four tasks. To create a task-specific measure of prediction accuracy at the individual student level, Carpenter et al. (1988) asked teachers to predict how each of six randomly selected "target" students from the teacher's class would perform on six different addition and subtraction word problems. For each teacher, the total number of correct predictions across target students and tasks was summed to produce a prediction accuracy score, with a score of 36 indicating that the teacher was 100% accurate in his or her predictions.

To date, no consensus has emerged in the literature on how best to measure teacher judgment accuracy. Some empirical evidence in research on mathematics teaching suggests that different indicators of teacher judgment accuracy may be unrelated to each other, suggesting that teacher judgment accuracy is better conceptualized as a multidimensional – rather than a one-dimensional – construct, with different indicators providing information on different aspects of teacher judgment accuracy (Brunner et al., 2013). The few studies that have included multiple indicators of teacher judgment accuracy in the same study and examined their relationship to student outcomes, report an inconsistent pattern of results (Anders et al., 2010; Karing et al., 2011). This suggests that the choice of method used to construct teacher judgment accuracy measures should be carefully considered with respect to its purpose (Artelt & Rausch, 2014).

Some researchers have argued on theoretical grounds that more global indicators of teacher judgment accuracy based on ranking or rating students' ability disclose little about a teacher's specific knowledge of what a student knows and understands in some