Teaching and Teacher Education 65 (2017) 48–60

Contents lists available at ScienceDirect

Teaching and Teacher Education

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

Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability

Kane Meissel*, Frauke Meyer, Esther S. Yao, Christine M. Rubie-Davies

University of Auckland, Faculty of Education and Social Work, Private Bag 92601, Symonds St, Auckland, New Zealand

HIGHLIGHTS

- Explored alignment of standardized achievement results with teacher judgments.
- Marginalized students received lower judgments after controlling for achievement.
- Classroom and school achievement composition inversely related to teacher judgments.
- Robust moderation of teacher judgments needed, both within and between schools.
- Professional development may assist teachers to make fair and consistent judgments.

ARTICLE INFO

Article history:
Received 22 June 2016
Received in revised form
16 January 2017
Accepted 25 February 2017
Available online 22 March 2017

Keywords:
Teacher judgments
Standardized testing
Student achievement
Social justice

ABSTRACT

Teacher judgments of student achievement are increasingly used for high-stakes decision-making, making it imperative that judgments be as fair and reliable as possible. Using a large national database from New Zealand, we explored the relation between psychometrically designed standardized achievement results and teacher judgments in reading (N = 4771 students) and writing (N = 11,765 students) using hierarchical linear modelling. Our findings indicated that judgments were systematically lower for marginalized learners after controlling for standardized achievement differences. Additionally, classroom and school achievement composition were inversely related to teacher judgments. These discrepancies are concerning, with important implications for equitable educational opportunities.

1. Introduction

The ability of teachers to accurately gauge student achievement is considered an important aspect of teachers' professional competence, as teacher judgments are often the primary source of information about student achievement (Ready & Wright, 2011; Südkamp, Kaiser, & Möller, 2012; Südkamp, Kaiser, & Möller, 2014). Teacher judgments are determinations made by teachers about students' current achievement (see Section 2 for more detail), and can impact teachers' ongoing instructional decision-making within the classroom, including instructional pace, level of support, and level of task difficulty (Alvidrez & Weinstein, 1999; Clark & Peterson, 1986; Hoge & Coladarci, 1989). For example, students judged to be more capable are more likely to receive higher quality learning opportunities than students judged less able (Clark & Peterson, 1986; Rubie-Davies, 2014; Rubie-Davies, Hattie, & Hamilton, 2006; Sharpley & Edgar, 1986). Furthermore, teacher judgments have implications for placement decisions in programs or ability groups, grade retention, and ultimately for students' future academic pathways (Begeny, Eckert, Montarello, & Storie, 2008; Begeny, Krouse, Brown, & Mann, 2011; Francis et al., 2016; Harlen, 2005; Parsons & Hallam, 2014; Wiliam & Bartholomew, 2004).

Internationally, much research has focused on teacher judgment alignment, mainly investigating the relations between teacher judgments and measured student performance. Reviews of this body of research have shown broad agreement between judgments and standardized assessments on average (r = 0.63, Südkamp et al., 2012), but the relations have been vastly inconsistent with a wide range of correlations reported (−0.03 to 0.92; Hoge & Coladarci, 1989; Südkamp et al., 2012). Südkamp et al. (2012) noted that teacher judgments showed higher correlations with measured

* Corresponding author. University of Auckland, Faculty of Education and Social Work, Private Bag 92601, Symonds St, Auckland 1150, New Zealand.
E-mail address: k.meissel@auckland.ac.nz (K. Meissel).

http://dx.doi.org/10.1016/j.tate.2017.02.021

achievement when teachers were informed about what measure their judgment was being compared with. Correlations were also higher when judgments and measures addressed the same domain or aspect of ability. Other test characteristics such as the number of points on the judgment scale did not impact the degree of alignment between judgments and measured student achievement.

Within the New Zealand context, overall teacher judgments (OTJs) were introduced as a specific achievement measure in 2010, and are assessed in relation to expected curriculum standards in reading, writing, and mathematics. These judgments are commonly referred to as National Standards (NS), and are intended to reflect a student's achievement in relation to the standard expected of students at the same year level nationally. Teachers are asked to consider a range of data, including observations of student learning, learning conversations, and formal assessments such as standardized achievement tests, to reach a decision on whether a student meets the demands of the New Zealand curriculum (Ministry of Education, 2011). Determination of whether a student meets the standard is up to the teacher, with no mandate with respect to which of these forms of evidence is utilized, nor the degree of weighting of specific types of data. The judgment, however, should focus solely on a student's achievement at that point in time and should not include construct irrelevant information such as a student's behavior or perceived potential ability.

A perfect correlation is unlikely and arguably undesirable since the two measures should be used for different purposes – standardized assessments often focus on a specific aspect of a student's learning whereas teacher judgments should take into account a number of aspects of a student's achievement within a whole subject area. Nonetheless, while previous research has investigated the relation between judgments and standardized achievement, the properties of teacher judgments and what informs these decisions remain relatively unexplored. The question remains whether lower correlations simply reflect differences in the nature of the assessments, or whether there are construct irrelevant factors which influence teachers when they make judgments about student performance. For example, although neither ethnicity nor special needs status *should* affect a judgment about a student's achievement, previous research has indicated that such factors might indeed influence teachers' judgments (see e.g., Glock, Krolak-Schwerdt, & Pit-ten Cate, 2015; Martínez, Stecher, & Borko, 2009; Ready & Wright, 2011). Although discrepancies in teacher judgments are to be expected given that there is random error in all assessments of student performance, systematic differences relating to specific subgroups would suggest a degree of bias.

Alignment between standardized tests and teacher judgments may also be affected by the inherently different interpretive approaches (Hattie & Brown, 2003; Hattie et al., 2003). Standardized tests are specifically designed to maximize reliability and consistency across students, classrooms, schools, and regions. In contrast, individual teachers typically make evaluations of student performance in relation to local (class or school) level evidence. That is, although teacher judgments may be defined as criterion-referenced, judgments are likely to be influenced by normative evaluations, such as how well each student is performing in relation to other students within the teacher's class (Angoff, 1974).

Despite considerable work investigating the properties of teacher judgments, the majority of these studies have been conducted in a North American context. Notable exceptions include the earlier studies by Doherty and Conolly (1985) and Sharpley and Edgar (1986), which were undertaken in Australia and the UK respectively, as well as the more recent research undertaken in Germany by Kaiser, Retelsdorf, Südkamp, and Möller (2013).

The current study extends previous work investigating the nature of teacher judgments in several ways. It most closely aligns with the work of Ready and Wright (2011), but draws on a sample of older students (approximately 9–13 years old) in both reading and writing. Ready and Wright's (2011) study focused on kindergarten students and research has shown that the alignment between teacher judgments and measured achievement can differ across grade levels, highlighting the need for further study with respect to older students (Südkamp et al., 2014). In addition, teachers in Ready and Wright's (2011) study had no access to standardized assessment results of students, whereas teachers in the current study had access to each students' standardized achievement results, and were advised by the New Zealand Ministry of Education that this was a source of evidence that could be drawn on when making judgments about students' achievement. The meta-analysis by Südkamp et al. (2012) indicated that research has yet to examine the way in which teacher judgments are affected by knowledge of standardized assessment results *prior* to making a holistic judgment about students' achievement within a learning domain. Furthermore, relatively few studies have utilized data collected as part of regular school routine. The current study uses teacher judgments and standardized achievement results collected in actual classroom contexts where data collection was not an imposed measure for schools.

Previous research has frequently focused on relatively small, localized samples of students; the average sample size of the 75 studies reported in a recent meta-analysis by Südkamp et al. (2012) was 518 students. The current study drew on data from a large-scale teacher professional development project with almost 5000 students represented in reading, and around 12,000 additional students for writing. Since teacher judgments are inherently likely to violate statistical assumptions of independence because one teacher determines the judgments for all students in his/her class, we employed three-level hierarchical linear modelling in the analyses with students nested within classrooms and nested within schools. This enabled the inherent clustering of the data to be accounted for.

Furthermore, the majority of teacher judgment studies have not attended to between-group differences with respect to student characteristics. The extent to which students' characteristics influence teachers' overall judgments of achievement remains largely inconclusive. Due to the importance of equitable educational opportunities, this is a key focus of the current study.

2. Review of teacher judgment literature

The following sections provide a review of the extant literature on teacher judgments – their alignment with standardized achievement results and the impact of student characteristics and school composition on these judgments. Due to the overlap of teacher expectations and teacher judgments, the review begins with a brief discussion of this issue.

2.1. Teacher expectations and teacher judgments

Teacher expectations and teacher judgments are similar in that both represent subjective teacher estimates about student achievement. They mainly differ in that expectations are typically predictions about *future* achievement while judgments are a *current* estimate of a student's performance. The latter are mostly made in circumstances in which the teacher has taught the student for some time and therefore can take into account a range of information. In contrast, teacher expectations focus on expected improvement or performance over a future time period, and are predictions of the possible academic progression of a student rather than an assessment of their existing skills and knowledge (see for example Rubie-Davies, Peterson, Sibley, & Rosenthal, 2015).