Teaching and Teacher Education 77 (2019) 170–182

Contents lists available at ScienceDirect

Teaching and Teacher Education

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

Consistency of results regarding teacher effects across subjects, school levels, outcomes and countries

Sigrid Blömeke*, Rolf Vegar Olsen

Centre for Educational Measurement (CEMO), University of Oslo, Norway

HIGHLIGHTS

- Consistency of TIMSS results across subjects, grades, outcomes, countries was examined.
- Instructional quality was consistently positively related to student outcomes.
- Other teacher characteristics were mostly inconsistently related to student outcomes.
- Instructional quality predicted student outcomes well and with large effect sizes.
- Effect sizes of other teacher characteristics were generally moderate only.

ARTICLE INFO

Article history:
Received 26 October 2017
Received in revised form
25 September 2018
Accepted 30 September 2018

Keywords:
mathematics
Science
Instructional quality
Student achievement
Motivation

ABSTRACT

How teacher characteristics are related to student outcomes may indicate ways to improve the effectiveness of schooling. This paper examines how consistent such relations are by utilizing TIMSS data on five teacher characteristics (teacher education major, teaching experience, professional development, preparedness and instructional quality) and two outcomes (achievement, motivation) in five countries (England, Norway, South Korea, Thailand, Tunisia) for two grades (4, 8) and two subjects (mathematics, science). Data revealed little consistency, if at all within countries only and regarding instructional quality as a predictor. Policy makers are advised not to make inferences across grades, subjects or outcomes.

The number of publications presenting results regarding teacher effects on student outcomes has substantially increased during the past decade. Researchers from many countries have provided evidence for effects on mathematics achievement for example (Baumert et al., 2010; Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Kersting, Givvin, Thompson, Santagata, & Stigler, 2012). Looking across the available evidence, the consistency of these results beyond one specific subject, school level, outcome or country examined is an open question though. Much of the research has focused on *mathematics*, often in *middle school*. Furthermore, outcome measures have often been limited to either cognitive *or* motivational outcomes of schooling. Finally, most of the research has relied on analysis of data from single countries.

If results were always discussed limited to the context examined in terms of subject, school level, outcome or country, reasons for concerns would be limited. However, even in peer-reviewed journals conclusions sometimes include rather general policy suggestions without supporting these with data from other subjects, school levels or countries. A particular problem emerges when policy makers want to use research to support a request for change, for example, of teacher education programs as Lingard and Lewis (2016) pointed out. Two reports by the Organisation for Economic Co-operation and Development [(OECD 2005]) and the European Commission (2015) may serve as examples. No matter how carefully these reports were based on data, they describe desirable teacher qualifications and teacher education programs on a general level across subjects and school grades although the underlying research was typically limited to a few grades and subjects – and this in a few countries only.

Against this background, the purpose of this paper is to examine how justified inferences from research results regarding teacher

* Corresponding author. Gaustadalleen 30, N – 0318, Oslo, Norway.
E-mail address: sigribl@cemo.uio.no (S. Blömeke).

https://doi.org/10.1016/j.tate.2018.09.018

effects for a selected set of predictors and outcomes are beyond one specific sample. If it turns out that relations of teacher effects to student outcomes are inconsistent across subjects, school levels, outcomes, and countries, it is difficult to use the available research to justify or argue for general reforms, for example of teacher education.

Hanushek and Luque (2003) did a pioneering study in this respect by comparing relations between teacher degree and teaching experience across about 15 countries with grade 4 mathematics data from the Trends in International Mathematics and Science Study (TIMSS) 1995 and across about 30 countries with grade 8 data. Studies by Akiba, LeTendre, and Scribner (2007) using TIMSS 2003 grade 8 mathematics data and teacher degree, teaching experience and teacher specialization as predictors, by Luschei and Chudgar (2011) using TIMSS 2003 grade 4 mathematics and science data and teacher degree, teaching experience and teachers' sense of preparedness as predictors as well as by Blömeke, Olsen and Suhl (2016) using TIMSS 2011 grade 4 mathematics data worked similarly with a few indicators and related these to one type of outcome.

All authors pointed out though that more teacher and teaching characteristics, subjects, grades and outcomes should be included in further studies and that a systematic estimation of the degree of consistency should take place. Such a systematic examination of the degree to which teacher and teaching characteristics consistently predict outcomes of schooling while transcending contexts is the first purpose of this paper. It adds, secondly, to the state of research by using TIMSS 2011 grade 4 *and* grade 8 data from mathematics *and* science with student achievement *and* student motivation as outcomes and including for the first time *instructional quality* as a predictor besides the common teacher characteristics. Although one can question how well the TIMSS variables characterize teacher and teaching quality and how suitable the cross-sectional design is for drawing conclusions about teacher effects, TIMSS provides still the richest dataset available when it comes to different countries, grades and subjects (for a discussion of limitations see the final section of this paper).

1. Conceptual framework

1.1. Teacher characteristics, instructional quality and student achievement

TIMSS collects data from representative 4th and 8th grade student samples of intact classrooms, including their mathematics and science teachers (Joncas & Foy, 2012). The data set provides thus a unique opportunity to link responses from students with those from their teachers. Teacher characteristics have, to varying degrees, been shown to have effects on student outcomes (Wayne & Youngs, 2003). These teacher characteristics cover a range of indicators.

Teacher education specialization in terms of major academic disciplines studied can be interpreted as a rough approximation of opportunities to learn during teachers' programs. This indicator has been identified as a strong predictor of learning outcomes by educational effectiveness research (Berliner, 1985; Carroll, 1963; Scheerens, 2016). We have general evidence from U.S. studies that a major in a subject and a major in the pedagogy of this subject are positively related to student achievement (Clotfelter, Ladd, & Vigdor, 2007; Cochran-Smith & Zeichner, 2005). Subject-specific evidence exists with respect to mathematics teachers from studies that used knowledge tests instead of majors as a proxy (Baumert et al., 2010; Kersting et al., 2012).

Teaching experience is another teacher characteristic that has proven to be relevant in a range of studies. Evidence exists in particular with respect to mathematics teachers (Clotfelter et al., 2007; Kersting et al., 2012). However, it seems as if this indicator only played out fully if the experience was gained in the same grade that was examined or if teachers were at the beginning of their career (Huang & Moon, 2009).

Teachers feel to varying degrees *prepared* for their task (Kee, 2012) which may reasonably be assumed to reflect teachers' self-efficacy (Bandura, 1986). Self-efficacy beliefs influence thought patterns and emotions, which in turn enable or inhibit actions. Teachers with strong self-efficacy are typically more persistent and make stronger efforts to overcome classroom challenges than others (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998). A recent research synthesis revealed that a significant relation to student achievement and student motivation exists (Zee & Koomen, 2016).

Professional development (PD) activities taken by teachers may vary from very short courses to comprehensive and continuous programs (Goldsmith, Doerr, & Lewis, 2014). Meta-analyses of single-subject studies supported the hypothesis that PD is positively related to student achievement if the activities meet certain quality characteristics (Timperley, Wilson, Barrar, & Fung, 2007). Desimone, Smith, and Frisvold (2010) classified these quality features into a focus on content, active learning, coherence, a certain minimum length and collaborative activities.

Besides teacher characteristics, teaching characteristics in terms of *instructional quality* (INQUA) matter for student outcomes (Seidel & Shavelson, 2007). Educational effectiveness studies and qualitatively oriented classroom observational studies converge on key INQUA features and their relation to student achievement and motivation. Classroom management, cognitive activation, clarity of instruction and a supportive climate are regarded as essential features (Decristan et al., 2015).

1.2. Dimensions of consistency

The question how consistent relations between teacher characteristics, INQUA and student outcomes are may be raised in several respects. Similarities and differences across different *subjects* are one important issue in this context. Expertise research has from early on pointed out that learning and achievement are domain-specific (Berliner, 2001) because, for example, experts are not able to transfer their speed and accuracy from one domain to another (Glaser & Chi, 1988; De Groot, 1946/1978). If a situation presented to an expert was completely different compared to those where he/she had gained the expertise, performance was not better than that of novices.

TIMSS allows examining potential differences between subjects with respect to predictors and outcomes of mathematics and science education. These two subjects are relatively similar though, for example in both relying on hypotheses and providing evidence for these. At the same time, differences exist. Whereas experiments and observations as well as hypothesis testing against natural phenomena are crucial in science, axiomatic structures, logical deduction and modelling patterns are crucial in mathematics. Correspondingly, the TIMSS assessment framework stresses number sense, operations and algebraic thinking (collapsed in grade 4), patterns and relationships, geometric shapes and data organisation in mathematics (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). The science assessment framework distinguishes between biology, chemistry (collapsed as life science in grade 4), earth science and physics. If it turns out that relations of teacher characteristics and INQUA to student outcomes are little consistent already across these two subjects that are relatively similar, it is difficult to implement general reforms across even broader ranges of subjects.

In addition to consistency across subjects, consistency across different *grades* is an issue – in particular if they belong to different school levels. TIMSS allows examining potential differences in relations between predictors and outcomes between grade 4 and 8. A