Your full name:

- Bas Senden

Affiliated authors with institutions:

- Trude Nilsen University of Oslo, Department of Teacher Education and School Research
- Nani Teig University of Oslo, Department of Teacher Education and School Research

Affiliation:

- University of Oslo, Department of Teacher Education and School Research

Current position

- Doctoral Research Fellow

Title of your paper

- Student ratings of instructional quality: validity, comparability, and the relationship with student achievement.

Abstract (224 words)

Researchers increasingly use student questionnaires to assess aspects of instructional quality in primary and secondary school. This necessitates in-depth knowledge of the properties and usefulness of such ratings. A novel opportunity arose when the Trends in Mathematics and Science Study (TIMSS) renewed their focus on instructional quality in 2019, contributing to new and updated items. In addition, Norway included several national items, thereby extending the measurement to assess three basic dimensions of instructional quality (classroom management, supportive climate and cognitive activation). The present study utilizes this opportunity by investigating the factorial structure, comparability, and contribution to student achievement of student ratings of instructional quality. To this end, TIMSS provided a large representative sample consisting of 3951 fifth-grade and 4575 ninth-grade students, who rated both their mathematics and science teachers' instructions. Results from multilevel confirmatory factor analysis support the three-dimensional factor structure across all investigated groups. Evidence from multilevel measurement invariance analysis supports the notion that classroom management and supportive climate are comparable across groups, whereas the comparability of cognitive activation is limited. Last, multilevel structural equation modelling was employed to investigate the contribution to student achievement. Results showed that classroom management correlates to student achievement in all groups. Other dimensions showed mixed results, especially across educational levels. The implications of these findings in improving the assessment of instructional quality are discussed.

Extended summary

1. Introduction

Teachers who provide high-quality instructions can foster students' cognitive and affective learning outcomes (Klieme, Pauli, & Reusser, 2009; Scheerens, Luyten, Steen, & de Thouars, 2007; Seidel & Shavelson, 2007).

One approach to assess instructional quality (INQUA) is by obtaining student ratings of teachers' behaviour through student questionnaires. A large body of studies has evaluated the psychometric properties and usefulness of this approach in the context of higher education. However, there is a remarkable lack of such research in primary and secondary education (Marsh, Dicke, & Pfeiffer, 2019). Bridging this gap is crucial as student ratings are an increasingly accepted and used measure for INQUA in primary and secondary school.

To date, several studies have investigated the properties and usefulness of student ratings of INQUA in primary and secondary school (e.g. Fauth, Decristan, Rieser, Klieme, & Büttner, 2014; Kyriakides et al., 2014; Rowley, Phillips, & Ferguson, 2019; van Der Scheer, Bijlsma, & Glas, 2019; Wagner, Göllner, Helmke, Trautwein, & Lüdtke, 2013; Wisniewski, Zierer, Dresel, & Daumiller, 2020). However, these studies seldom combine investigating the reliability and validity of student ratings of INQUA with examining their contribution to student outcomes. In addition, they often focus on a specific subject and educational level, leaving the comparability across educational levels and subjects understudied.

Furthermore, a majority of prior studies draw on data obtained in Germany or the USA. However, comparative studies using International Large-Scale Assessments (ILSAs) indicate that the role of INQUA on student outcomes often differs across countries. These results were even found between countries that are considered relatively similar, such as the Nordic or Confucian countries (Bellens, Damme, Noortgate, Wendt, & Nilsen, 2019; Nilsen, Scherer, & Blömeke, 2018; Yi & Lee, 2017). Thus, to effectively inform teacher education and professional development policies, there is a need for high-quality studies at the national level.

2. Theoretical framework

The construct of INQUA reflects instructional practices implemented in classrooms that are considered indicative of student learning (Blömeke, Kaiser, König, & Jentsch, 2020; Nilsen & Gustafsson, 2016).

One of the most prominent frameworks in Europe has conceptualized INQUA through three basic dimensions: Classroom Management, Cognitive Activation, and Supportive Climate (Klieme, Schümer, & Knoll, 2001). *Classroom Management* includes a teachers' ability to establish and maintain clear rules regarding content and social norms. To do so, a teacher needs to create stable routines, good planning and pacing, and keep the students engaged (Brophy, 1983; Klieme et al., 2009). S*upportive climate* features the social and emotional support that teachers provide to students, including a supportive teacher-student relationship, positive and constructive teacher feedback, a positive approach to student errors and misconceptions, individual learner support and caring teacher behaviour (Klieme et al., 2009, p. 141). Last, *cognitive activation* consists of several key instructional features that promote and encourage understanding, including challenging tasks, activating prior knowledge, content-related discourse and participation practices (Klieme et al., 2009). This

conceptualization is based on strong theoretical foundations and confirmed by a number of empirical studies that assessed its psychometric properties (see Praetorius, Klieme, Herbert, & Pinger, 2018 for a complete overview).

2.1 The present study

The current study aims to shed light on the properties and usefulness of student ratings of INQUA. To this end, we wanted to test the assumption that students can provide reliable and valid assessments of three dimensions of INQUA. We outlined three research questions (RQs) to guide this study:

RQ1: To what extent does the factorial structure of students' ratings of INQUA reflect the proposed three-dimensional conceptualization of classroom management, cognitive activation, and supportive climate?

RQ2: Do students perceive INQUA similarly between educational levels and subjects?

RQ3: To what extent do student ratings of INQUA in primary and secondary education relate to student achievement in mathematics and science?

3. Methodology

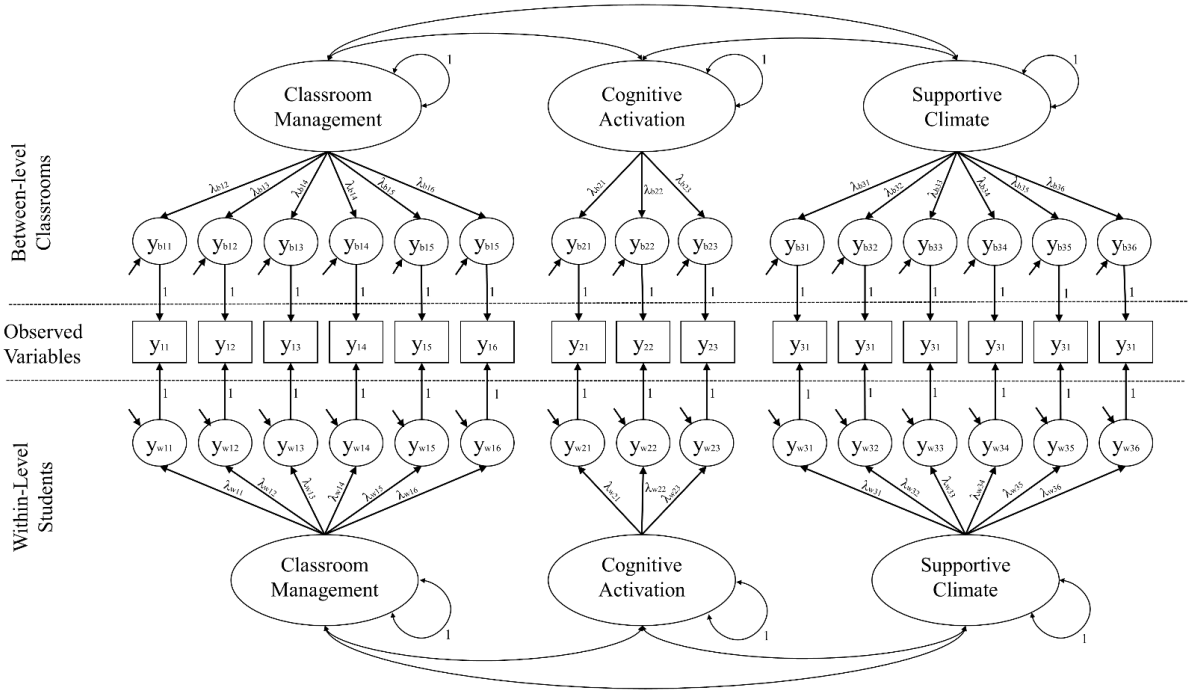
Data from the Trends in International Mathematics and Science Study (TIMSS) provided a large representative sample consisting of 3951 fifth-grade and 4575 ninth-grade students, both divided over 231 classes. All students rated both their mathematics and science teachers' instructions and completed a mathematics and science assessment.

The data were analyzed using a multilevel approach with individual students on the first level and classrooms on the second. Three analytical steps, each pertaining to a separate research question, were employed. To this end, we investigated (1) the factorial structure of student ratings of INQUA with multilevel confirmatory factor analysis; (2) the comparability across educational levels and subject with multilevel measurement invariance analysis; (3) the relationship between student ratings of INQUA and student achievement with multilevel structural equation modelling.

4. Results

RQ1. The empirical data was in line with the proposed three-dimensional factor structure, reflecting the three Basic Dimensions of INQUA based on Klieme et al. (2001): Classroom Management, Supportive Climate and Cognitive Activation (figure 1). This model showed an acceptable model fit (RMSEA=0.016, CFI=0.984, TLI=0.981, $SRMR_{within}$=0.023, $SRMR_{between}$=0.085). In addition, factor correlations were below .85 in all models, providing evidence that students can adequately distinguish between three dimensions of instructional quality.

Figure 1. Factorial structure of the proposed three dimensional model of INQUA



RQ2. The comparability of each of the three dimensions was tested between educational levels (grade 5 vs grade 9) and between subjects (mathematics vs science). Cognitive activation was excluded from the analysis between subjects because different items were used to measure the construct in science and mathematics. Findings from the analysis (table 2) indicate that the empirical data is in line with a three-dimensional factor structure across all the investigated groups. Thus, students in both primary and secondary school students can adequately distinguish between the different dimensions regardless of whether they rate their mathematics or science teachers' instructions.

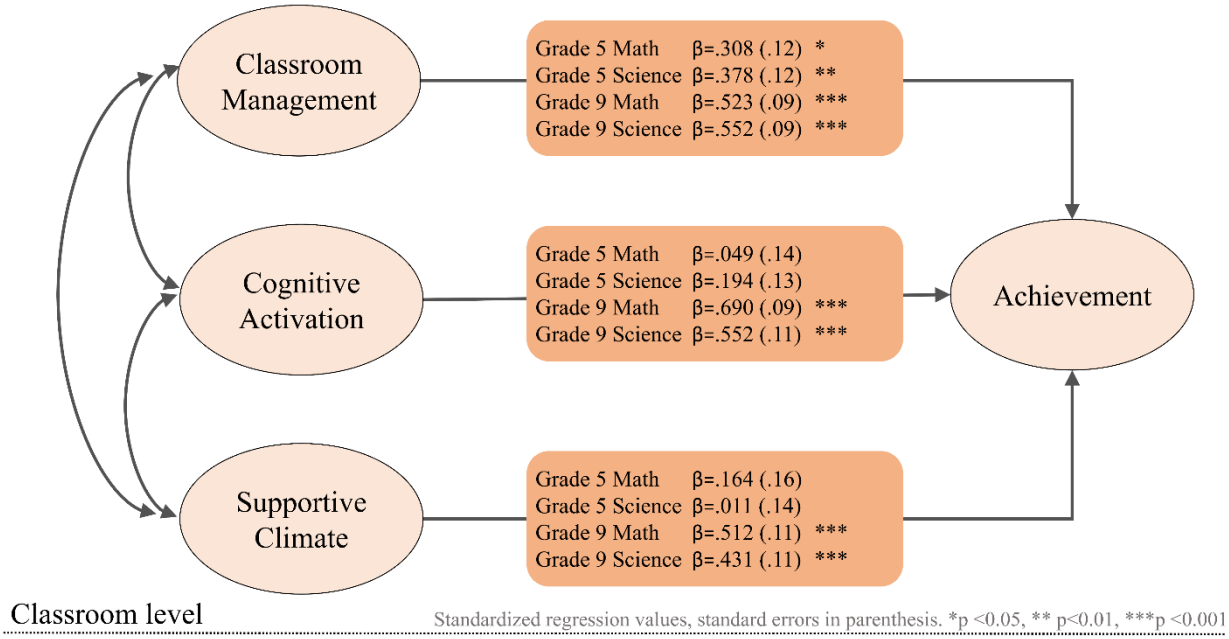
However, the dimensions need to attest at least metric invariance to allow a meaningful comparison of correlations across groups (RQ3). Therefore, in the results of RQ3, we can compare the relationships of student ratings of INQUA with achievement across both educational levels and subjects for classroom management and a supportive climate, but not for cognitive activation.

Table 2. level of measurement invariance across grades and subjects

Groups	Grade 5 vs 9		Mathematics vs Science	
Measures	Mathematics	Science	Grade 5	Grade 9
Classroom Management	Scalar	Scalar	Scalar	Scalar
Supportive Climate	Scalar	Scalar	Scalar	Metric
Cognitive Activation	Configural	Configural	N.A.	N.A.

RQ3. Findings from multilevel structural equation modelling (figure 2) indicate that classroom management significantly relates to student achievement in mathematics and science regardless of the grade and subject. However, the results indicated significant differences between educational levels.

Figure 2. Correlations of each of the three dimensions with achievement



References

Bellens, K., Damme, J., Noortgate, W., Wendt, H., & Nilsen, T. (2019). Instructional quality: catalyst or pitfall in educational systems' aim for high achievement and equity? An answer based on multilevel SEM analyses of TIMSS 2015 data in Flanders (Belgium), Germany, and Norway. *An IEA-ETS Research Institute Journal, 7*(1), 1-27. doi:10.1186/s40536-019-0069-2

Blömeke, S., Kaiser, G., König, J., & Jentsch, A. (2020). Profiles of mathematics teachers' competence and their relation to instructional quality. *ZDM, 52*(2), 329-342. doi:10.1007/s11858-020-01128-y

Brophy, J. (1983). Classroom Organization and Management. *The Elementary School Journal, 83*(4), 265-285. doi:10.1086/461318

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9. doi:10.1016/j.learninstruc.2013.07.001

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras Study. Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom.* (pp. 137-160). Münster u.a.: Waxmann.

Klieme, E., Schümer, G., & Knoll, S. (2001). Mathematikunterricht in der Sekundarstufe I. "Aufgabenkultur" und Unterrichtsgestaltung.(Nebst) CD-ROM. In *TIMSS - Impulse für Schule und Unterricht.* (pp. 43-57). Bonn: Bundesministerium für Bildung u. Forschung.

Kyriakides, L., Creemers, B., Panayiotou, A., Vanlaar, G., Pfeifer, M., Cankar, G., & McMahon, L. (2014). Using student ratings to measure quality of teaching in six European countries. *European Journal of Teacher Education, 37*. doi:10.1080/02619768.2014.882311

Marsh, H. W., Dicke, T., & Pfeiffer, M. (2019). A tale of two quests: The (almost) non-overlapping research literatures on students' evaluations of secondary-school and university teachers. *Contemporary Educational Psychology, 58*, 1-18. doi:10.1016/j.cedpsych.2019.01.011

Nilsen, T., & Gustafsson, J.-E. (2016). *Teacher Quality, Instructional Quality and Student Outcomes : Relationships Across Countries, Cohorts and Time*(1st ed. 2016. ed., Vol. 2).

Nilsen, T., Scherer, R., & Blömeke, S. (2018). The relation of science teachers' quality and instruction to student motivation and achievement in the 4th and 8th grade: A Nordic perspective. In *Northern Lights on TIMSS ansd PISA 2018* (pp. 61-94). Denmark: Nordic Council of Ministers.

Praetorius, A.-K., Klieme, E., Herbert, B., & Pinger, P. (2018). Generic dimensions of teaching quality: the German framework of Three Basic Dimensions. *Mathematics Education, 50*(3), 407-426. doi:10.1007/s11858-018-0918-4

Rowley, J. F. S., Phillips, S. F., & Ferguson, R. F. (2019). The stability of student ratings of teacher instructional practice: examining the one-year stability of the 7Cs composite.

School Effectiveness and School Improvement, 30(4), 549-562. doi:10.1080/09243453.2019.1620293

Scheerens, J., Luyten, J. W., Steen, R., & de Thouars, Y. C. H. (2007). *Review and meta-analyses of school and teaching effectiveness.* Enschede: Universiteit Twente, Afdeling Onderwijsorganisatie en management.

Seidel, T., & Shavelson, R. J. (2007). Teaching Effectiveness Research in the Past Decade: The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research, 77*(4), 454-499. doi:10.3102/0034654307310317

van Der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement, 30*(1), 30-50. doi:10.1080/09243453.2018.1539015

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction, 28*, 1-11. doi:10.1016/j.learninstruc.2013.03.003

Wisniewski, B., Zierer, K., Dresel, M., & Daumiller, M. (2020). Obtaining secondary students' perceptions of instructional quality: Two-level structure and measurement invariance. *Learning and Instruction, 66*. doi:10.1016/j.learninstruc.2020.101303

Yi, H., & Lee, Y. (2017). A latent profile analysis and structural equation modeling of the instructional quality of mathematics classrooms based on the PISA 2012 results of Korea and Singapore. *Asia Pacific Education Review, 18*(1), 23-39. doi:10.1007/s12564-016-9455-4