Your full name: *Bas Senden*

Affiliated authors with institutions: *Bas Senden, Nani Teig, Trude Nilsen*

Affiliation: *All authors are affiliated with the* F*aculty of Educational Sciences, Department of Teacher Education and School Research, University of Oslo, Norway*

Current position: *PhD Candidate*

Student perceptions of teaching quality across 39 countries: Evaluating the comparability and impact of teachers' gender

Abstract (300 words)

Teaching quality is well known to be related to student learning. Therefore, it has frequently been assessed through student questionnaires in international large-scale assessments (ILSAs), such as the Trends in International Mathematics and Science Study (TIMSS). Although cross-national comparisons are at the heart of ILSAs, student perceptions are rarely used to compare teaching quality across countries due to the unattainable requirement of exact measurement invariance. Hence, the current study aims to bridge this gap by applying a novel method, namely Alignment, to examine student ratings of disorderly behaviour and instructional clarity using data from TIMSS 2019 in Grade 8 across 39 countries. It aims to (1) evaluate the comparability of disorderly behaviour and instructional clarity across countries, (2) examine the cross-national comparison of their mean scores, and (3) investigate the impact of teachers' gender on student ratings of the two aspects of teaching quality. Preliminary results revealed that comparability across countries varied between disorderly behaviour and instructional clarity. Both scales also showed significant differences in mean scores across countries. Finally, the impact of teachers' gender on student ratings of disorderly behaviour and instructional clarity seemed to be negligible across countries.

Extended summary (1000 words, excluding reference list) Include introduction, theoretical background, methods, aims, preliminary findings/findings, results, theoretical and education significance, relevance to the QUINT ambition and the reference list.

1. Introduction

Student questionnaires are commonly used to acquire student perceptions of teaching quality and, subsequently, examine its association with student learning outcomes (Seidel & Shavelson, 2007). The majority of studies investigating student perceptions of teaching quality have been performed in one specific, mostly western, country (e.g., Baumert et al., 2010; Wisniewski et al., 2020). Using findings from single-country studies to make inferences about cross-national differences and similarities can lead to erroneous conclusions (Avvisati et al., 2019; van de Vijver, 2018). Unfortunately, few studies have examined student perceptions of teaching quality cross-nationally and have done so primarily across a small number of countries (e.g., Scherer et al., 2016; van der Lans et al., 2021). Additional research is needed to identify cross-national differences and similarities in students perceived teaching quality and, consequently, compare - and learn from - findings across countries. In addition, cross-national investigations might inform the extent to which measures work universally or provide a foundation for including more local or regional measures (Rutkowski & Rutkowski, 2018). Against this

background, there is a need for novel investigations of student perceptions of teaching quality across a variety of culturally diverse countries.

2. Theoretical background

2.1 Teaching quality

Teaching quality can be defined as those interactions between students and teachers that are hypothesized to encourage learning processes that lead to desired student learning outcomes (Klieme et al., 2009; Nilsen et al., 2016). Scholars agree that teaching quality is a multidimensional and complex construct (Creemers & Kyriakides, 2008; Teddlie et al., 2006) which consists of at least three core overarching dimensions: classroom management, cognitive activation, and supportive climate (Klieme et al., 2009; Pianta et al., 2008). These dimensions are referred to as being generic and are thus considered relevant across a variety of subjects, grades, and possibly countries and cultures (Praetorius et al., 2018).

In the current study, we focus on classroom management and supportive climate. More specifically, we focus on two aspects of these dimensions that were, for the first time, included in the international student questionnaire of TIMSS 2019: *Disorderly behaviour* and *instructional clarity* (Mullis & Martin, 2017).

3. Methods

For meaningful cross-national comparisons, the underlying measurement structure of the constructs should be stable, that is, measurement invariant (Davidov et al., 2018). A common approach to assess measurement invariance is by conducting a multi-group confirmatory factor analysis (MGCFA), followed by systematically restricting parameters in the model to be invariant. However, the requirement of exact invariance is an almost unattainable ideal (Asparouhov & Muthén, 2014; Marsh et al., 2018). Applied research has consistently run into this problem and often refrained from mean-score comparisons, used single-item measures (and are thus unable to assess measurement invariance), or simply ignored possible non-invariance (He et al., 2019; Nilsen et al., 2016).

With this severe limitation in mind, Asparouhov and Muthén (2014) have recommended another approach to measurement invariance for many groups, namely alignment optimization. The alignment optimization method is an approach of approximate invariance that extracts trustworthy means and variance from the data without constricting the parameters to be exactly equal across groups (Asparouhov & Muthén, 2014). The approach also provides an overview of the extent of measurement invariance for each model parameter in every country (Muthén & Asparouhov, 2018). The latter could help for further scale development by indicating which items are more useful for cross-country comparison (Marsh et al., 2018). Recently the alignment method has been extended to include general structural equation models (Asparouhov & Muthén, 2022).

4. Aims

This study investigates the comparability of student perceptions of teaching quality across 39 countries who participated with eighth-grade students in TIMSS 2019. In addition, the study aims to make a subsequent cross-national mean-score comparison and investigates the impact of teachers'

gender on student ratings of teaching quality across countries. The following research questions address the above aims:

1. To what extent are student perceptions of teaching quality measurement invariant across a diverse set of countries?
2. To what extent are there significant mean-score differences of student perceived teaching quality across countries?
3. To what extent does teachers' gender influence student ratings of teaching quality?

5. Preliminary findings

The traditional MGCFA approach indicated that students' perceptions of disorderly behaviour and instructional clarity were metric but not scalar invariant. As these results did not allow mean-score comparisons, we continued with the alignment optimization method. Findings of the alignment method indicated 50.2% non-invariant parameters for disorderly behaviour and 33.8% for instructional clarity. Disorderly behaviour was well above the guidelines of a maximum of 25% non-invariance proposed by Muthén and Asparouhov (2014), whereas instructional clarity was only slightly above this value. Therefore, and in line with Muthén and Asparouhov (2014) recommendations, we decided to run a Monte Carlo simulation study to test further the extent to which the obtained mean scores are trustworthy. Simulated correlations between the estimated and generated factor means and factor variances were sufficiently high and above the recommended cut-off of 0.98 proposed by Muthén and Asparouhov (2018). The obtained mean scores showed significant differences in how students rate teaching quality between countries. In addition, the impact of teachers' gender on the ratings was negligible and only significant a small number of countries.

6. Theoretical and educational significance

This study provides further understanding of how we can compare student perceptions of teaching quality across countries. In addition, the alignment method will provide insights into which scales and items measuring teaching quality in TIMSS 2019 are most (in)comparable across countries, thus creating room for scale development. The study also implies that, in contrast to findings from higher education, teachers' gender hardly influences student ratings of teaching quality.

7. Relevance to the QUINT ambition

Especially the latter part of the study - the mean score comparisons - will give more insight into how students in the Nordic countries perceive teaching quality in relation to other countries in the world. In addition, the findings show that there is no significant impact of teachers' gender on how students rate teaching quality in either Norway or Sweden.

References

Asparouhov, T., & Muthén, B. (2014). Multiple-Group Factor Analysis Alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495-508. https://doi.org/10.1080/10705511.2014.919210

Avvisati, F., Le Donné, N., & Paccagnella, M. (2019). A meeting report: cross-cultural comparability of questionnaire measures in large-scale international surveys. *Measurement Instruments for the Social Sciences*, *1*(8), 1-10. https://doi.org/10.1186/s42409-019-0010-z

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., & Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133-180. https://doi.org/10.3102/0002831209345157

Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness : a contribution to policy, practice and theory in contemporary schools*. Routledge.

Davidov, E., Muthen, B., & Schmidt, P. (2018). Measurement Invariance in Cross-National Studies:Challenging Traditional Approaches and Evaluating New Ones. *Sociological Methods & Research*, *47*(4), 631-636. https://doi.org/10.1177/0049124118789708

He, J., Barrera-Pedemonte, F., & Buchholz, J. (2019). Cross-cultural comparability of noncognitive constructs in TIMSS and PISA. *Assessment in Education: Principles, Policy & Practice*, *26*(4), 369-385. https://doi.org/10.1080/0969594X.2018.1469467

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). Waxmann. http://www.ciando.com/ebook/bid-40151

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524-545. https://doi.org/10.1037/met0000113

Mullis, I. V. S., & Martin, M. O. (2017). TIMSS 2019 Assessment Frameworks. *Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2019/frameworks/*

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, *5*, 978. https://doi.org/10.3389/fpsyg.2014.00978

Muthén, B., & Asparouhov, T. (2018). Recent Methods for the Study of Measurement Invariance With Many Groups:Alignment and Random Effects. *Sociological Methods & Research*, *47*(4), 637-664. https://doi.org/10.1177/0049124117701488

Nilsen, T., Gustafsson, J.-E., & Blömeke, S. (2016). Conceptual Framework and Methodology of This Report. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher Quality, Instructional Quality and Student Outcomes: Relationships Across Countries, Cohorts and Time* (pp. 1-19). Springer Publishing & IEA. https://doi.org/10.1007/978-3-319-41252-8

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H Brookes Publishing.

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

Rutkowski, L., & Rutkowski, D. (2018). Improving the Comparability and Local Usefulness of International Assessments: A Look Back and A Way Forward. *Scandinavian Journal of Educational Research*, *62*(3), 354-367. https://doi.org/10.1080/00313831.2016.1261044

Scherer, R., Nilsen, T., & Jansen, M. (2016). Evaluating Individual Students' Perceptions of Instructional Quality: An Investigation of their Factor Structure, Measurement Invariance, and Relations to Educational Outcomes. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00110

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. https://doi.org/10.3102/0034654307310317

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for Teacher Observation and Feedback: Evolution of an international study of teacher effectiveness constructs. *Educational Research and Evaluation*, *12*(6), 561-582. https://doi.org/10.1080/13803610600874067

van de Vijver, F. J. R. (2018). Towards an Integrated Framework of Bias in Noncognitive Assessment in International Large-Scale Studies: Challenges and Prospects. *Educational Measurement: Issues and Practice*, *37*(4), 49-56. https://doi.org/https://doi.org/10.1111/emip.12227

van der Lans, R. M., Maulana, R., Helms-Lorenz, M., Fernández-García, C.-M., Chun, S., de Jager, T., Irnidayanti, Y., Inda-Caro, M., Lee, O., Coetzee, T., Fadhilah, N., Jeon, M., & Moorer, P. (2021). Student Perceptions of Teaching Quality in Five Countries: A Partial Credit Model Approach to Assess Measurement Invariance. *SAGE Open*, *11*(3), 21582440211040121. https://doi.org/10.1177/21582440211040121

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*, Article 101303. https://doi.org/10.1016/j.learninstruc.2020.101303