BCES Conference Books, 2023, Volume 21. Sofia: Bulgarian Comparative Education Society
ISSN 2534-8426 (online), ISSN 1314-4693 (print)
ISBN 978-619-7326-13-0 (online), ISBN 978-619-7326-14-7 (print)

Martin Boško, Ondřej Papajoanu, Angie Moore & Hana Voňková

Examining Paradoxical Associations Between Students' Questionnaire Responses and Their Achievement Across PISA Cycles: The Case of Teacher Support

Abstract

In the context of Programme for International Student Assessment (PISA), paradoxical findings concerning the relationship between questionnaire scales and student achievement are often documented. These questionnaire scales are found to correlate positively with student achievement within countries at the individual level, but negatively at the between-country level (i.e., when correlating questionnaire scale and achievement values aggregated at the level of countries, the countries being the unit of analysis). These anomalous findings can be caused by the differences in reporting behavior of students in different countries and might lead educators and educational policy-makers to erroneous conclusions. In this paper, we examine the relationship between the teacher support scale and student achievement across three PISA cycles – 2012, 2015, and 2018. Our results show that there is a consistent negative between-country correlation between teacher support and student achievement in all three examined PISA cycles, which is in line with the previously documented paradoxes. We have also found that some countries, which participated in all three PISA cycles under study, consistently contribute to this paradox by having quite high levels of student-reported teacher support but rather low achievement scores. Future research should take into account the differences in reporting behavior between students when making cross-country analyses and consider the application of methodological approaches to identify and adjust for these differences such as the anchoring vignette method and the overclaiming technique.

Keywords: PISA, questionnaire, reporting behavior, teacher support

Introduction

The Programme for International Student Assessment (PISA) is a tri-yearly assessment administered to 15-year-old students across the globe, focusing on the three core school subjects of science, reading and mathematics (OECD, 2016a). This assessment provides data allowing researchers and educational policymakers to compare student performance across different countries. PISA data gives not only a snapshot of world-wide educational performance trends, but also has become highly influential in educational policy making (Rutkowski & Rutkowski, 2016). Although there is often a major focus on comparing the academic achievement of students from participating countries (Rutkowski & Rutkowski, 2016), PISA data also provides important information about a variety of non-cognitive factors such as students' attitudes, beliefs, interest, motivation, as well as their well-being (Bertling et al., 2016). These non-cognitive factors are growing in importance as they are seen as predictors of

See on pages 298-300 how to cite the papers published in this volume.

scholastic performance, educational attainment, as well as labor market success (Bertling et al., 2016).

However, the accurate measurement of these factors is a challenging and inappropriate measurement of non-cognitive factors that can limit the validity of the findings and usefulness of the data for policy decisions (Bertling et al., 2016; Kyllonen & Bertling, 2013).

Differences in reporting behavior and paradoxical results in the PISA study

PISA study questionnaires rely on four-point Likert scale items which can be prone to inaccuracies due to the differences in reporting behavior between students from different countries (Bertling et al., 2016; Kyllonen & Bertling, 2013). Such differences have been suggested as the cause of the "attitude-achievement paradox" when questionnaire scales are found to correlate with achievement positively within countries at the individual level (hereinafter "within-country correlation"), but negatively at the country level (i.e., when correlating questionnaire scale and achievement values aggregated at the level of countries, hereinafter "between-country correlation"; Bertling et al., 2016; Kyllonen & Bertling, 2013).

For example, in PISA 2003, the mean within-country correlation between the mathematics self-concept scale and mathematics achievement was r = .40, while the between-country correlation was r = -.20 (Kyllonen & Bertling, 2013). This means that countries with lower achievement tended to have, on average, higher self-concept values and vice versa. Similar paradoxical findings have been documented, for example, by He and van de Vijver (2016) who found, using PISA 2012 data, a median within-country correlation between mathematics achievement and intrinsic motivation r = .17 and extrinsic motivation r = .13, while the between-country correlations were r = -.53 and r = -.52 for intrinsic and extrinsic motivation, respectively. Further, Vonkova, Zamarro and Hitt (2018) found, using PISA 2012 data, a negative between-country correlation between math teacher's classroom management levels and math achievement r = -.33.

Teacher support in the PISA study

Teacher support behavior is one of the contextual factors commonly studied in the PISA assessment. Teachers' support of and care for students have been recognized as a relevant factor in a number of positive outcomes such as, for example, lower anxiety, higher academic motivation, higher engagement in learning, and better well-being (e.g., OECD, 2016b; OECD, 2019). In PISA study, a teacher support scale is commonly included in a student questionnaire, asking students to report whether their teacher shows interest in every student's learning, whether their teacher helps students with their learning, etc. (OECD, 2016b; OECD, 2019). Based on student's responses to these items, an index of teacher support is created and included in the student questionnaire data set (OECD, 2016b; OECD, 2019). In our study, we focus this data concerned with student-reported teacher support.

This study

In this study, we aim to investigate whether the paradoxical findings documented for a number of PISA questionnaire scales, which were found to correlate positively

with student achievement within countries but negatively at the between-country level, could also be found for teacher support scale and whether the pattern of paradoxical results would be the same across three PISA cycles.

Specifically, our research question is: What is the correlation between teacher support and student achievement in PISA 2012, 2015, and 2018?

Methodology

We use student questionnaire data from 3 cycles of PISA testing: 2012, 2015, and 2018. In each cycle, a teacher support scale was included in the questionnaire, asking about teacher behaviors in lessons of the subject which was the major tested subject in that cycle (note that in each cycle, one of the three subjects is tested in detail, taking almost half of the total testing time; OECD, 2016a).

The major subjects were mathematics, science, and reading in 2012, 2015, and 2018, respectively. For each cycle, we examine the relationship between the teacher support scale and student achievement in the corresponding (i.e., major) subject. Below we provide more information about the data for each cycle.

PISA 2012

We use data from 480,174 respondents from 65 countries and economies (hereinafter we use the term "countries"). In PISA 2012, teacher support was measured through question *ST77* where students reported the frequency of teacher supporting behaviors in mathematics classes on a scale *every lesson* (1), *most lessons* (2), *some lessons* (3), and *never or hardly ever* (4). The items for question *ST77* were as follows:

How often do these things happen in your mathematics lessons?
ST77Q01: The teacher shows an interest in every student's learning.
ST77Q02: The teacher gives extra help when students need it.
ST77Q04: The teacher helps students with their learning.
ST77Q05: The teacher continues teaching until the students understand.
ST77Q06: The teacher gives students an opportunity to express opinions.

In our analysis, we use the TEACHSUP index from the student questionnaire data set as an indicator of teacher support. The index was based on student responses to the *ST77* items above. The index was created so that the higher the index value, the higher the teacher support (for more details on the calculation of the index see OECD, 2014).

As the indicators of student achievement in mathematics, we use the first plausible value (PV1MATH) for within-country analyses and the country achievement scores in mathematics (*math score*; the values were taken from PISA Data Explorer; OECD, n.d.-a) for between-country analyses.

PISA 2015

We use data from 514,119 respondents from 72 countries. In PISA 2015, teacher support was measured through question *ST100* where students reported the frequency of teacher supporting behaviors in science classes on a scale *every lesson* (1), *most lessons* (2), *some lessons* (3), and *never or hardly ever* (4). The items for question *ST100* were as follows:

How often do these things happen in your <school science> lessons?
ST100Q01: The teacher shows an interest in every student's learning.
ST100Q02: The teacher gives extra help when students need it.
ST100Q03: The teacher helps students with their learning.
ST100Q04: The teacher continues teaching until the students understand.
ST100Q05: The teacher gives students an opportunity to express opinions.

In our analysis, we use the TEACHSUP index from the student questionnaire data set as an indicator of teacher support. The index was based on student responses to the *ST100* items above. The index was created so that the higher the index value, the higher the teacher support (for more details on the calculation of the index see OECD, 2017).

As the indicators of student achievement in science, we use the first plausible value (PV1SCIE) for within-country analyses and the country achievement scores in science (*science score*; OECD, n.d.-a) for between-country analyses.

PISA 2018

We use data from 566,793 respondents from 75 countries. In PISA 2018, teacher support was measured through question *ST100* where students reported the frequency of teacher supporting behaviors in test language classes on a scale *every lesson* (1), *most lessons* (2), *some lessons* (3), and *never or hardly ever* (4). The items for question *ST100* were as follows:

How often do these things happen in your <test language lessons>?
ST100Q01: The teacher shows an interest in every student's learning.
ST100Q02: The teacher gives extra help when students need it.
ST100Q03: The teacher helps students with their learning.
ST100Q04: The teacher continues teaching until the students understand.

In our analysis, we use the TEACHSUP index from the student questionnaire data set as an indicator of teacher support. The index was based on student responses to the *ST100* items above. The index was created so that the higher the index value, the higher the teacher support (for more details on the calculation of the index see OECD, n.d.-b).

As the indicators of student achievement in reading, we use the first plausible value (PV1READ) for within-country analyses and the country achievement scores in reading (*reading score*; OECD, n.d.-a) for between-country analyses.

Analysis

As for within-country analyses, teacher support (TEACHSUP) and student achievement (PV1MATH, PV1SCIE, and PV1READ for PISA 2012, 2015, and 2018, respectively) were correlated at the individual level in each country using the *final student weight* (W_FSTUWT). As for between-country analyses, we first calculated the weighted mean of the TEACHSUP index using the *final student weight* (W_FSTUWT) for each country and then correlated these weighted means with the country achievement scores (*math score*, *science score*, and *reading score* for PISA 2012, 2015, and 2018, respectively). Note that participating countries differ across the PISA cycles, so the between-country correlations from different cycles are not directly comparable.

Results

For 2012, the correlation of TEACHSUP and PV1MATH within countries ranges from -0.164 to 0.219 with a median correlation of 0.006. In total, 30 countries had a negative within-country correlation between TEACHSUP and PV1MATH, while 35 countries had a positive within-country correlation between TEACHSUP and PV1MATH. The lowest within-country correlation was found, for example, in Serbia, Montenegro, Tunisia, Slovakia, and Austria. Contrarily, the highest within-country correlation was found, for example, in Norway, Denmark, Australia, Jordan, and Korea. The country level correlation for TEACHSUP and country *math score* is -0.479.

For 2015, the correlation of TEACHSUP and PV1SCIE within countries ranges from -0.124 to 0.161 with a median correlation of -0.010. In total, 43 countries had a negative within-country correlation between TEACHSUP and PV1SCIE, while 29 countries had a positive within-country correlation between TEACHSUP and PV1SCIE. The lowest within-country correlation was found, for example, in Uruguay, Slovakia, Romania, Tunisia, and Peru. Contrarily, the highest within-country correlation was found, for example, in Malta, Vietnam, the Chinese provinces of Beijing, Shanghai, Jiangsu and Guangdong, Norway, and Finland. The country level correlation for TEACHSUP and country *science score* is -0.461.

For 2018, the correlation of TEACHSUP and PV1READ within countries ranged from -0.106 to 0.186 with a median correlation of 0.020. In total, 24 countries had a negative within-country correlation between TEACHSUP and PV1READ, while 51 countries had a positive within-country correlation between TEACHSUP and PV1READ. The lowest within-country correlation was found, for example, in Austria, Germany, Israel, Slovakia, and Panama. Contrarily, the highest within-country correlation was found, for example, in Malaysia, Sweden, Norway, Jordan, and Korea. The country level correlation for TEACHSUP and country *reading score* is -0.528.

A closer analysis of the countries that participated in all three PISA cycles (54 countries) has revealed that if we rank their mean TEACHSUP and achievement score for each cycle, some countries consistently rank among the top 10 countries when it comes to teacher support, yet they tend to be among bottom 10 countries when it comes to achievement score. Such countries are for example Brazil, Costa Rica, Jordan, Mexico, and Peru. Curiously, the opposite paradox (i.e., ranking bottom 10 in teacher support and top 10 in achievement score) is not present in the data.

Conclusion

In PISA, paradoxical findings are commonly found when questionnaire scales and student achievement correlate positively within countries at the individual level, but negatively at the country level. These findings have been attributed to the differences in reporting behavior between students from different countries. In our study, we documented consistent moderate negative between-country correlations between the teacher support index (TEACHSUP) and the countries' achievement scores in PISA 2012, 2015, and 2018. This means the higher the country's score on an achievement test, the lower the student-reported level of teacher support and vice versa. The within-country correlations are both positive and negative and weaker in strength compared to the between-country correlations. Thus, our study suggests that there is an unexpected paradox also in the case of teacher support scale and shows its consistency across different PISA cycles. Interestingly, some countries, which participated in all three

PISA cycles examined in our study, have been found to consistently contribute to this paradox by having quite high levels of student-reported teacher support but rather low achievement scores.

Our findings indicate the need to take into account the differences in reporting behavior between students when making cross-country analyses. Future research could thus consider the application of methodological approaches, such as the anchoring vignette method (King et al., 2004) and the overclaiming technique (Vonkova, Papajoanu & Stipek, 2018), to identify differences in student reporting behavior and adjust student-reported data for these differences. The previous analyses of PISA data have shown that these approaches have the potential to explain some of the paradoxical findings and improve the cross-country comparability of student-reported data (e.g., Kyllonen & Bertling, 2013; Vonkova, Zamarro & Hitt, 2018).

References

Bertling, J. P., Marksteiner, T., Kyllonen, P. C. (2016): General noncognitive outcomes. In S. Kuger, E. Klieme, N. Jude & D. Kaplan (Eds.) *Assessing contexts of learning. Methodology of educational measurement and assessment* (pp. 251-281). Cham: Springer.

He, J. & Van de Vijver, F. (2016): The motivation-achievement paradox in international educational achievement tests: Toward a better understanding. In R. B. King & A. B. I. Bernardo (Eds.) *The psychology of Asian learners: A festschrift in honor of David Watkins* (pp. 253-268). Singapore: Springer.

King, G., Murray, C., Salomon, J. & Tandon, A. (2004): Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98(1), 191-207.

Kyllonen, P. C. & Bertling, J. (2013): Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier & D. Rutkowski (Eds.) *A handbook of international large-scale assessment data analysis: Background, technical issues, and methods of data analysis* (pp. 277-285). New York: Chapman and Hall/CRC Press.

OECD (2014): *PISA 2012 Technical Report*. Paris: OECD.

OECD (2016a): *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD.

OECD (2016b): *PISA 2015 Results (Volume II): Policies and Practices for Successful Schools*. Paris: OECD.

OECD (2017): *PISA 2015 Technical Report*. Paris: OECD.

OECD (2019): *PISA 2018 Results (Volume III): What School Life Means for Students' Lives*. Paris: OECD.

OECD (n.d.-a): *PISA data explorer*. https://pisadataexplorer.oecd.org/ide/idepisa/ (Accessed 9 February 2023).

OECD (n.d.-b): *PISA 2018 Technical Report – Chapter 16: Scaling procedures and construct validation of context questionnaire data*. Paris: OECD.

Rutkowski, L. & Rutkowski, D. (2016): A call for a more measured approach to reporting and interpreting PISA results. *Educational Researcher*, 45(4), 252-257.

Vonkova, H., Papajoanu, O. & Stipek, J. (2018): Enhancing the cross-cultural comparability of self-reports using the overclaiming technique: An analysis of accuracy and exaggeration in 64 cultures. *Journal of Cross-Cultural Psychology*, 49(8), 1247-1268.

Vonkova, H., Zamarro, G. & Hitt, C. (2018): Cross-country heterogeneity in students' reporting behavior: The use of the anchoring vignette method. *Journal of Educational Measurement*, 55(1), 3-31.

Acknowledgements

This conference contribution was made possible with the support of Czech Science Foundation's grant GA CR 21-09064S project "Heterogeneity in reporting behavior on surveys among different countries, schools and groups of students".

Mgr. Martin Boško, Charles University, Czech Republic

Dr. Ondřej Papajoanu, Charles University, Czech Republic

Angie Moore, M.A., Charles University, Czech Republic

Assoc. Prof. Dr. Hana Voňková, Charles University, Czech Republic