**Who overclaims? An analysis of PISA 2012 data**

# **Objectives**

Comparing the United States with 63 other education systems that participated in the 2012 Program for International Student Assessment (PISA), this study examines the background questionnaire items designed to measure students’ tendency to overstate what they know. In particular, this study (a) explores students’ tendency to overclaim their familiarity to mathematics concepts across participating education systems, (b) attempts to adjust such over-claiming response pattern, (c) identifies student subgroups based on their response patterns, and (d) examines the relationship between students’ overclaiming patterns and their mathematics performance.

# **Perspectives and Theoretical Framework**

The self-reported Likert scale is commonly used in social science research to understand the participants’ attitude or belief in the topic of interest (Beaton, et al. 2000). It is of no exception in large-scale international studies such as PISA, which provide not only students’ assessment data on various subjects, but also responses to a series of self-reported Likert scale background questionnaires from students, parents, teachers, and school principals. The relationship between the assessment data and the background questionnaire data are explored in a wide range of cross-cultural research (Loveless 2017, Thomson, Bortoli and Underwood 2017, Haahr, et al. 2005).

However, literature points out many potential challenges facing self-reported Likert scale data (Ludeke and Makransky 2015, Buckley 2009). For instance, social desirability bias is participants’ tendency to choose answers that are viewed as desirable by society at large (Paulhus 2002). There are also evidence of individuals overstating the levels of traits that they themselves see as desirable (Ludeke, Weisberg and DeYoung 2013). In making cross-cultural comparisons based on self-reported data, such problems become more pronounced as differences in responses to the background questionnaire Likert scales may be in part due to differences in response styles across gender (Watkins and Cheung 1995), race (Marin, Gamba and Marin 1992), and cultures in general (Javaras and Ripley 2007, King, et al. 2004). For example, the research found that Chinese and Japanese secondary students are more likely to use the midpoint of a seven-point Likert-type item, while U.S. students exhibit a greater tendency toward extreme response style than the Asian students or their Canadian counterparts (Chen, Lee and Stevenson 1995). Therefore, interpreting responses to background questionnaires at face value would cause validity threats to the research results, which would then incorrectly inform education policy.

Efforts to address these problems include methods such as anchoring vignettes, forced-choice comparisons, situational-judgement-test, behavioral items, bi-factor models, and the overclaiming technique (He, Buchholz and Klieme 2017, Kyllonen and Bertling 2014, Brown and Maydeu-Olivares 2011, Cheung and Rensvold 2000, Rossi 2001).They all help alleviate the comparability issue, although there is no consensus among the limited literature on which method is the best (He and Van de Vijver 2016). Background questionnaire items to detect students’ tendency of overclaiming were designed in PISA 2012 as one way to enable adjustments for cross-cultural differences in response tendencies (OECD 2014). However, it has not received much research attention. No previous studies have examined the overclaiming questions at the item-level, nor have they identified student subgroups who tend to overclaim. This analysis aims at filling these gaps by comparing – in a global context – students’ responses to the overclaiming questions and by exploring characteristics of students who tended to overclaim.

# **Methods and Data Sources**

PISA is an assessment of 15-year-old students coordinated by the Organization for Economic Cooperation and Development (OECD). This study uses PISA 2012 mathematics assessment data and student background questionnaire data, which asked students about their familiarity with 16 mathematical concepts (see Appendix A). There are 13 actual mathematical concepts such as “exponential function”, “complex number”, and “polygon”. In addition, there are 3 pseudo concepts designed to detect overclaiming: “proper number”, “subjunctive scaling”, and “declarative fraction”. These non-existing concepts were created by combining a term from grammar with a mathematical term (OECD 2014).

The response options for both the real and the pseudo concepts are the same: (a) never heard of it, (b) heard of it once or twice, (c) heard of it a few times, (d) heard of it often, and (e) know it well, understand the concept. These options are assigned values from 1 to 5, respectively. Three indices were calculated from students’ responses. The first index (REAL\_MEAN) is a simple mean score of students’ responses to the 13 real mathematical concepts. The second index (PSEUDO\_MEAN) is the mean score of students’ responses to the 3 non-existing concepts. The last index (REAL\_MEAN\_ADJUSTED) is the difference between the first index and the second index, which is a conventional way of correcting over-claiming (Paulhus, Harms, et al. 2003, Zimmerman, et al. 1977).

Based on the quartile distribution of the scores of the first two indices, students in each education system were identified in one of the four groups[[1]](#footnote-1), which can be characterized as following:.

* **Irrational respondents** are students who tend to choose “never heard of it” to real concepts but “know it well, understand the concept” to pseudo concepts.
* **Over claimers** are students who tend to choose “know it well, understand the concept” to both real and pseudo concepts.
* **Lower claimers** are students who tend to choose “never heard of it” to both real the pseudo concepts.
* **Ideal respondents** are students who tend to choose “know it well, understand the concept” to real concepts but “never heard of it” to pseudo concepts.

Results are presented at the education system level for the United States and 63 other education systems around the world[[2]](#footnote-2). Two-tailed *t*-tests were performed for comparisons between student subgroups. Test results with *p*-values under 0.05 are considered statistically significant.

# **Results and Conclusions**

**Q1: How do students respond to real and pseudo mathematical concepts?**

* Although students’ response varies to each of the 16 mathematics familiarity questions, there are general trends about how students responded to real versus pseudo mathematical concepts. Figure 1 illustrates students’ reported familiarity level with an example real concept, “exponential function” to the left, and with an example pseudo concept, “proper number” to the right.
* Across the 64 PISA participating education systems, the percentage of students who reported that they had never heard of the real concept “exponential function” ranged from 8% in Chinese Taipei to 80% in Tunisia. In the United States, 14% of the students fall into this category.
* When presented with the made-up item, “proper number”, it would be expected that the majority of the students would report “never heard of it”. This is the case in 5 education systems where more than half of the students reported this way: Iceland (73%), Spain (65%), Korea (63%), and Chinese Taipei (61%), and Hong Kong – China (55%). However, in the remaining education systems, the percentage of students who reported “never heard of it” ranged from 48% in Sweden to 3% in Albania. In the United States, 15% of the students reported that they have never heard of the pseudo concept. In other words, 85% of them overstated what they know. These patterns are similar for other pseudo items not displayed.
* This is a clear indication of the existence of the over-claiming response style among PISA participating students. Given that students tend to over claim with varying degrees across education systems, interpreting students’ familiarity with real mathematical concepts without any adjustment would be misleading.

**Q2: Could over-claiming be adjusted?**

* Across the 64 education systems, on average, students indicated a 3.4 unadjusted familiarity rating for the real mathematics concepts (REAL\_MEAN) and a 2.3 familiarly rating for the pseudo concepts (PSEUDO\_MEAN). The average mathematics performance for these 64 education systems are 474 out of 1000.
* Table 1 reveals that, on average, the within-education-system correlations between the unadjusted familiarity ratings and math achievement are r = 0.46, which range from -0.01 in Albania (not statistically different from 0) to 0.64 in Korea. The average within-education-system correlation between the adjusted familiarity ratings and the math achieving are r = 0.43, which ranges from 0.01 in Albania (still not statistically different from 0) to 0.69 in Liechtenstein. The change in the correlation brought by the adjustment varies by the education system, ranging from decreasing 0.27 in Macao – China to increasing 0.26 in Sweden.
* At the cross-education-system level, before adjustment, there is a very weak relationship (r = 0.16) between students’ familiarity and math performances. Although there may be a substantive explanation for having such a weak relationship at the across-education-system level but not at the within-education-system level, it is more reasonable to believe that the across-education-system level correlation of r = 0.16 does not tease out the bias brought by different response style across education systems. After adjustment, the correlation increases to r = 0.68, which is in line with what the findings from the field test of PISA 2012 (Kyllonen and Bertling 2014).

**Q3: What percentage of students are identified as over claimers in each education system?**

* Based on students’ average response to the real and the pseudo concepts, this analysis identifies 4 types of students in each education system: low claimers, irrational respondents, ideal respondents, and over claimers. As seen in Figure 2, across the 64 education systems, the percentage of over claimers ranged from 8% in Korea and Spain to 18% in Sweden. In the United States, 12% of the students are identified as over claimers.
* Most of the education systems with high percentages of over claimers have low performance in mathematics, defined as having an average mathematics performance below the OECD average. A few exceptions include New Zealand and Denmark, where there are high percentages of over claimers but also higher-than-OECD-average mathematic performance.

**Q4: Are there gender gaps in the percentage of over claimers?**

* Depending on the education system, there may be a gender gap in the percentage of over claimers, as shown in Figure 3. In the United States, for example, 11% of girls are over claimers, while 12% of boys are the same. The gender gap is not statistically significant in this case. In 31 education systems, however, there are higher percentages of boys than girls who are over claimers. In 3 education systems, which all have below OECD average mathematics performance, higher percentages of girls than boys are identified as over claimers.

**Q5: What are the mathematics scores for students who are over claimers and ideal respondents?**

* Figure 4 illustrates the average mathematics scores by student group in each education system. The average mathematics scores for over claimers range from 386 in Indonesia to 639 in Shanghai – China, which are generally higher than the scores of low claimers and the irrational respondents. In the United States, students who are identified as over claimers scored 516 on average, while its low claimers and irrational respondents scored 428 and 388, respectively.
* However, compared with ideal respondents, over claimers tend to have lower average mathematics scores across the 64 education systems. Except for Macao – China, Albania, and Tunisia where the gap in the average mathematics scores between over claimers and ideal respondents are not measurably different, in all education systems, over claimers score on average, from 26 points (Hong Kong – China) to 100 points (Qatar) lower than ideal respondents.
* The score of ideal respondents provides adjustment to the score of all students in measuring mathematics performance. The former indicator excludes students who are over claimers, lower claimers, and irrational respondents, and therefore provides new information about the mathematics performance of an education system. The mathematics score of ideal respondents is the lowest in Albania (388 points) and is the highest in Chinese Taipei (672 points).
* When ranking the 64 education systems based on mathematics scores of ideal respondents rather than of all students, education systems move up as many as 19 places (from the 23rd to the 4th in New Zealand), and moves down as many as 23 places (from the 6th to the 29th in Macao – China)

# **Significance**

This analysis provides a cross-national picture of students’ tendency to overclaim based on the PISA 2012 background questionnaire item. The findings contribute to the existing literature by focusing on students’ response to non-existing concepts, identifying over claimers, and comparing their academic outcomes. The results help to understand students’ response style across education systems, and they highlight the importance of accounting for students’ different response patterns when interpreting the results, especially in cross-cultural comparisons. Further research is invited to explore ways to improve validity and reliability of participants’ self-reported data on the background questionnaire items, as more valid and reliable background information data can help us better understand how contextual information is related to achievement and allows researchers to have potentially less biased comparisons across education systems.

# **References**

Beaton, Dorcas E., Claire Bombardier, Francis Guillemin, and Marcos Bosi Ferraz. 2000. "Guidelines for the process of cross-cultural adaptation of self-report measures." *SPINE* 3186-3191.

Brown, A, and A Maydeu-Olivares. 2011. "Item response modeling of forced-choice questionnaires." *Educational and Psychological Measurement* 460-502.

Buckley, Jack. 2009. "Cross-national response styles in international educational assessments: Evidence from pisa 2006."

Chen, C., S. Lee, and H. W. Stevenson. 1995. "Response style and cross-cultural comparisons of rating scales among east asian and north american students." *Psychological Science* 170-175.

Cheung, G. W., and R. B. Rensvold. 2000. "Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling." *Journal of Cross-Cultural Psychology* 187-212.

Haahr, J. H., T. K. Nielsen, M. E. Hansen, and S. T. Nielsen. 2005. "Explaining student performance: Evidence from the international PISA, TIMSS and PIRLS surveys." Danish Technological Institute.

He, Jia, and Fons Van de Vijver. 2016. "Correcting for scale usage differences among Latin American countries, Portugal, and Spain in PISA." *RELIEVE.*

He, Jia, Janine Buchholz, and Eckhard Klieme. 2017. "Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures." *Journal of Cross-Cultural Psychology* 319-334.

Javaras, K. N., and B. D. Ripley. 2007. "An "Unfolding" latent variable model for likert attitude data: Drawing inferences adjusted for response style." *Journal of the American Statistical Association* 454-463.

King, G., C. J. L. Murray, J. A. Salomon, and A. Tandon. 2004. "Enhancing the validity and cross-cultural comparability of measurement in survey research." *American Political Science Review* 341-355.

Kyllonen, Patrick C., and Jonas P. Bertling. 2014. "Innovative questionnaire assessment methods to increase cross-country comparability." Edited by Leslie Rutkowski, Matthias von Davier and David Rutkowski. *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis* (CRC Press).

Loveless, Tom. 2017. *The 2017 Brown Center report on American education: How well are American students learning?* The Brookings Institution, Washington D.C.: The Brookings Institution.

Ludeke, S. G., Y. J. Weisberg, and C. G. DeYoung. 2013. "Idiographically desirable responding: Individual differences in perceived trait desirability predict overclaiming." *European Journal of Personality* 580-592.

Ludeke, Steven, and Guido Makransky. 2015. "Does the over-claiming questionnaire measure overclaiming? Absent convergent validity in a large community sample." *Psychological Assessment.*

Marin, G., R. J. Gamba, and B. V. Marin. 1992. "Extreme response style and acquiescence among hispanics: The role of acculturation and education." *Journal of Cross-Cultural Psychology* 498-509.

OECD. 2014. *The PISA 2012 Technical Report.* OECD, Paris: OECD.

Paulhus, D. L. 2002. "Socially desirable responding: The evolution of a construct." Edited by H. I. Braun, D. N. Jackson and D. E. Wiley. *The role of constructs in psychological and educational measurement* 49-69.

Paulhus, D. L., P. D. Harms, M. N. Bruce, and D. C. Lysy. 2003. "The over-claiming technique: Measuring self-enhancement independent of ability." *Journal of Personality and Social Psychology* 890-904.

Rossi, P. E., Z. Gilula, and G. M. Allenby. 2001. "Overcoming scale usage heterogeneity: A Bayesian hierarchical approach." *Journal of the American Statistical Association* 20-31.

Thomson, Sue, Lisa De Bortoli, and Catherine Underwood. 2017. *PISA 2015: Reporting Australia’s results.* Australian Council for Educational Research, Camberwell, Victoria: Australian Council for Educational Research.

Watkins, D., and S. Cheung. 1995. "Culture, gender, and response bias: An analysis of re-sponses to the self-description questionnaire." *Journal of Cross-Cultural Psychology* 490.

Zimmerman, J., P.K.Broder, J.J. Shaughnessy, and B.J. Underwood. 1977. "A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition." *Intelligence* 5-31.

# **Appendix, Tables, and Figures**

1. The detailed categorization is summarized in Appendix B. A fifth group, “Others”, includes students whose scores of REAL\_MEAN and PSEUDO\_MEAN are either at the 2nd or the 3rd quartile. [↑](#footnote-ref-1)
2. Norway is not included in the analysis due to no data available for the mathematics familiarity questions. [↑](#footnote-ref-2)