**Who tend to overclaim in the PISA student background questionnaire?**

* Call them pseudo-concepts instead of foil
* Consider having a master country level spreadsheet that has many columns, including math score, score by gender, %, % by gender, blablabla?

# **Objectives**

Comparing the United States with 63 other education systems that participated in the 2012 Program for International Student Assessment (PISA), this analysis examines the background questionnaire items designed to estimate students’ tendency to overstate what they know. In particular, this analysis (a) explores students’ tendency to overclaim their familiarity to mathematics concepts across participating education systems, (b) attempts to adjust such overclaiming response pattern, (c) identifies students based on their response patterns, and (d) compares their percentage distribution and mathematics performances.

# **Perspectives and Theoretical Framework**

Self-reported Likert scale is commonly used in social science research to understand 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 the Program for International Student Assessment (PISA), and the Trends in International Mathematics and Science Study (TIMSS). These large-scale international studies 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 many cross-cultural research[[1]](#footnote-1).

However, literature points out many problems facing self-reported Likert scale data[[2]](#footnote-2). Social desirability bias is one of the problems, as participants tend to choose answers that are viewed as desirable by society at large[[3]](#footnote-3). There are also evidence of individuals overstating the levels of traits that they themselves see as desirable[[4]](#footnote-4). 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[[5]](#footnote-5), race[[6]](#footnote-6), socioeconomic status[[7]](#footnote-7), and cultures in general.[[8]](#footnote-8) For example, 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. [[9]](#footnote-9) 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[[10]](#footnote-10), forced-choice comparisons[[11]](#footnote-11), situational-judgement-test, behavioral items, bi-factor models[[12]](#footnote-12), and the overclaiming technique[[13]](#footnote-13). They all help alleviate the comparability issue, although there is no consensus among the limited literature on which method is the best.[[14]](#footnote-14) 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[[15]](#footnote-15). However, it has not received much research attention. No previous studies have examined the overclaiming questions at the item-level, nor have they identified 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 coordinated by the Organization for Economic Cooperation and Development (OECD). This analysis uses PISA 2012 mathematics assessment data and links the assessment data to data from the PISA student background questionnaire, which asked students about their familiarity with mathematical concepts. There are 13 actual mathematical concepts such as “exponential function”, “complex number”, and “polygon” (see appendix A). 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.[[16]](#footnote-16)

The response options for both the real concepts and the foils 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. This analysis calculates three indexes 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 (FOIL\_MEAN) is the mean score of students’ responses to the 3 foils. The last index (REAL\_MEAN\_ADJUSTED) is the difference between the first index and the second index, which is a conventional way of correcting overclaiming[[17]](#footnote-17).

Based on students’ score of the first two indexes, this analysis identifies students in each education system who fits into one of the following four groups.

* 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.

Table 1. xxxxx

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **REAL\_MEAN** | | |
|  |  | **Bottom quartile** | **2nd and the 3rd quartiles** | **Top quartile** |
| **PSEUDO\_MEAN** | **Top quartile** | **Irrational respondents** |  | **Over claimers** |
| **2nd and the 3rd quartiles** | **Others** | | |
| **Bottom quartile** | **Low claimers** |  | **Ideal students** |

Results are presented at the education system level for the United States and 63 other education systems around the world[[18]](#footnote-18). Two-tailed t-tests were performed for comparisons of percentages and mathematics performance. Test results with p-values under 0.05 are considered statistically significant.

# **Results and Conclusions**

**Q1: How do students responded 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[[19]](#footnote-19). 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 falls into this category.
* When presented with the foil 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.
* This is a clear indication of the existence of the overclaiming response style among PISA participating students. Given that students tend to overclaim with varying degrees across education systems, interpreting students’ familiarity with real mathematical concepts without any adjustment would be misleading.

**Q2: Could overclaiming 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 mathematic performance for these 64 education systems are 474 out of 1000.
* Table 2 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 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[[20]](#footnote-20).

**Q3: Which country tend to have the highest overclaim rate**

* 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. [Figure 2] Across 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 are low performers in mathematics, defined as having an average mathematic 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?**

* [Figure 2-a] Depending on the education system, there may be a gender gap in the percentage of over claimers. 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 percentage of boys than girls who are over claimers. In 3 education systems, which all have below OECD average mathematic performance, higher percentage of girls than boys are identified as over claimers.

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

* 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 moves 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 exploring their characteristics. 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 other ways of remedy the lack of comparability concerns. Large-scale assessment data could only be fully utilized for basic research and evidence-based policy making when more efforts are put in alleviating measurement bias in different cultural contexts.

[call for consistent overclaiming measure across cycles?]

# **References**

# **Appendix**

* + Knowledge about above could help understand students’ behavior better, they also remind us the importance of carefully interpreting students’ self-report.
    - These findings are important for increasing the quality of data obtainable from ILSA quesitonnaires, and may be even more generally useful in addressing cross-cultural (Bartram 2013) and even cross-sub-group (Mickelson 1990) comparability issues. – Bertling 2014
  + Field Trial and initial Main Survey analyses (Bertling and Roberts, 2011) showed increased cross-cultural comparability of correlations with achievement for the adjusted compared to the unadjusted index
  + Jasmine: As Markus mentioned, one of the NAEP Validity Panel research studies I’ve been doing (Computer Familiarity and Access Study), also used “overclaiming” items to adjust students’ response on the confidence items. What we found is that if we adjust students’ confidence level (Students’ confidence level on the “real concepts”-students’ confidence level on the “pseudo concepts”, overclaiming), the predictive power of students’ confidence for predicting achievement is much higher.
* Secondary analysts of these attitudinal data, however, generally ignore the issue of cultural differences in response style or scale usage heterogeneity, leading to descriptive statistics and inferences that may be biased and misleading. In this paper I explore this issue several ways and illustrate the extent, consequences, and some possible solutions using data from PISA 2006.[[21]](#footnote-21)
* Secondary analysts who use attitudinal data from international education assessments are at risk of reaching erroneous conclusions if they do not consider the issue of cultural differences in survey response style.[[22]](#footnote-22)
* Although individual respondents' idiosyncratic usage of different response styles adds noise to attitude survey data, systematic differences in response style across nations or cultures can introduce far more serious biases in both descriptive statistics and inferential results from more complex models. Unfortunately, there is much empirical evidence of such systematic biases between cultures. (Javaras and Ripley 2007; King, Murray, Salomon, and Tandon 2004; Johnson 2003; Rossi, Gilula, and Allenby 2001; Baumgartner and Steenkamp 2001; Heine, Takata, and Lehman 2000; de Vijver and Leung 1997; Chen, Lee, and Stevenson 1995; Mullen 1995; Greenleaf 1992; Poortinga 1989).[[23]](#footnote-23)
* These findings, particularly the cross-national research on secondary school populations, suggest that heterogeneity in response style could be a potential source of bias in the secondary analysis of PISA and other international assessment data. In this paper I investigate the extent, form, and consequences of cross-cultural differences in response style or scale usage using data from the PISA 2006 student questionnaire and science assessment (OECD 2007).[[24]](#footnote-24)

1. (e.g.

   Loveless 2007; Haahr, Nielsen, Hansen, and Nielsen 2005; Woessmann 2001) (from Buckely paper), note that I could probaby search for some different/newer papers of these authers [↑](#footnote-ref-1)
2. Buckley paper; Does the Over-Claiming Questionnaire Measure Overclaiming?; [↑](#footnote-ref-2)
3. Paulhus, 2002 (from the Does the Over-Claiming Questionnaire Measure Overclaiming?;) [↑](#footnote-ref-3)
4. (Ludeke, Weisberg, & DeYoung, 2013 (from the Does the Over-Claiming Questionnaire Measure Overclaiming?;) [↑](#footnote-ref-4)
5. Watkins, D. and S. Cheung (1995). From Buckley paper [↑](#footnote-ref-5)
6. Marin, Gamba, and Marin (1992) from Buckley paper; The Attitude-Achievement Paradox Among Black Adolescents [↑](#footnote-ref-6)
7. Bachman and O'Malley (1984) from buckley paper; Correlates of Respondent Accuracy in the Denver Validity Survey [↑](#footnote-ref-7)
8. Bertling 2014, Chen, Lee, and Stevenson (1995) from Buckely paper; also from the Buckley paper: Unfortunately, there is much empirical evidence of such systematic biases between cultures. (Javaras and Ripley 2007; King, Murray, Salomon, and Tandon 2004; Johnson 2003; Rossi, Gilula, and Allenby 2001; Baumgartner and Steenkamp 2001; Heine, Takata, and Lehman 2000; de Vijver and Leung 1997; Chen, Lee, and Stevenson 1995; Mullen 1995; Greenleaf 1992; Poortinga 1989) [↑](#footnote-ref-8)
9. Chen, Lee, and Stevenson (1995) from Buckely paper. [↑](#footnote-ref-9)
10. Cite something [↑](#footnote-ref-10)
11. Cite something [↑](#footnote-ref-11)
12. (e.g., Brown & Maydeu-Olivares, 2011; Cheung & Rensvold, 2000; Rutkowski et al., 2014) from Jia He paper [↑](#footnote-ref-12)
13. Cite something [or cite the whole sentence: bertleing 2014; OECD technical report?] (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy, and Underwood, 1977). From the 2012 tech report [↑](#footnote-ref-13)
14. Jia he paper [↑](#footnote-ref-14)
15. 2012 tech report [↑](#footnote-ref-15)
16. Tech report page 367 [↑](#footnote-ref-16)
17. (OCT; Paulhus, Harms, Bruce and Lysy, 2003; see also Zimmerman, Broder, Shaughnessy and Underwood, 1977) from tech report page 367 [↑](#footnote-ref-17)
18. Norway is kicked out for sure (no data for familiarly questions), other countries remain watching [↑](#footnote-ref-18)
19. Students’ response for each item could be found in the appendix [↑](#footnote-ref-19)
20. Bertling 2014 page 283 (or PISA 2012 tech report) [↑](#footnote-ref-20)
21. buckley [↑](#footnote-ref-21)
22. buckley [↑](#footnote-ref-22)
23. buckley [↑](#footnote-ref-23)
24. buckley [↑](#footnote-ref-24)