Do Language Spoken at Home and Reading Literacy Associate with Inconsistent Responding to Mixed-Worded Scales?

Shampa Roseline Rozario

UNIVERSITY OF OSLO

Master of Science in Assessment, Measurement and Evaluation

30 credits master thesis

CEMO: Centre for Educational Measurement
Faculty of Educational Sciences

Spring 2022

Popular Abstract

Questionnaires sometimes include statements like *I feel like an outsider in school* and *I feel like I belong at school* near each other. These two statements are similar in meaning but have opposite wording. The intention is to stop respondents from answering carelessly. If a respondent agrees with both these statements or disagrees with both, they are considered inconsistent. In this thesis, I look into whether a respondent's reading ability is associated with the likelihood of responding inconsistently. I find that in more than half of the seventy-six worldwide education systems that participated in PISA 2018, it seems to hold that a student with more reading ability is less likely to be an inconsistent responder. I also ask if a respondent is more or less likely to respond inconsistently when they speak a different language at home rather than the language of the questionnaire. I find that a student who speaks a different language at home seems to be more likely to respond inconsistently in five education systems, but less likely to respond inconsistently in seven education systems. Looking at the association among reading ability, language spoken at home, and inconsistent responding together, it appears that when language spoken at home seems to be related to inconsistent responding, it is so for both higher and lower levels of reading ability. To understand these relationships better, future researchers can look into more variables like the socioeconomic and immigration status of the families of the student. They can also study different processes of reading ability to better understand its relationship to inconsistent responses.

Acknowledgements

This thesis would not have been possible without the support of certain people. I would first like to thank my thesis supervisors Professor Dr. Johan Braeken and Dr. Isa Steinmann for their patience, guidance, and support. I have benefitted greatly from their deep knowledge in this field and detailed feedback.

Thank you to my partner, Ontiq Dey, for constantly listening to me even when I became too emotional, repeatedly proofreading (even after work and during difficult times), and for the sacrifices he has made for me to pursue my Master's degree.

I am grateful to my parents and my younger brother for their constant love and support, which keep me motivated and confident. I am grateful of their belief in me.

Do Language Spoken at Home and Reading Literacy Associate with Inconsistent Responding to Mixed-Worded Scales?

Mixed worded scales contain both positively and negatively worded items. These are introduced in some questionnaires to reduce response style, which is a tendency to answer items regardless of their content. When a respondent fails to switch their response on facing a negatively worded item, the response is inconsistent. By analyzing data of 15-year-old students from 76 education systems that participated in PISA 2018, I investigate whether reading literacy or language spoken at home are associated with the probability of being an inconsistent responder. I expected to find that a respondent is more likely to respond inconsistently if they speak a different language at home rather than the language of the assessment. I also expected that students with higher reading literacy are less likely to respond inconsistently. I found that the proportion of inconsistent responders ranges from 9% to 33% across the education systems with an average of 17%. Using separate logistic regressions for each of the education systems with both the predictors in the model, I found that in 45 out of 76 education systems, as expected, reading literacy showed a negative association with inconsistent response. In 5 education systems, as expected, an inconsistent response is found to be more probable when a different language is used at home. However, contrary to expectation, in 7 education systems the opposite was found. I conclude by discussing limitations regarding the generalizability of the study and discussing how future research can go deeper into understanding how language spoken at home and reading literacy associate with inconsistent responses.

Keywords: mixed-worded scales, inconsistent respondents, reading literacy, language spoken at home, PISA 2018

In educational and social sciences research, the Likert scale is a fundamental psychometric tool that is often used where respondents indicate their level of agreement on a graded scale (Joshi et al., 2015). Response bias can be a threat to the validity of data collected using this tool. Response bias occurs when the answer patterns on the questionnaires do not reflect respondents' actual opinion (Baumgartner & Steenkamp, 2001). Response bias can occur for several reasons; two of them are widely known: response set and response style. Response set occurs when respondents do not answer according to their opinion but rather to social desirability. Response style is a tendency to answer items regardless of their content (Sonderen et al., 2013). Inattention is a response style where bias comes from respondents being less careful to read and understand the questions and the response options, and thus responding away from the truth. Some respondents have the tendency to fall into response styles when items are phrased in a consistent way. This can, for example, happen when a respondent becomes habituated with similarly structured items before facing one with a different wording (Sonderen et al., 2013).

It is a usual practice to use mixed-worded scales to try to reduce response styles (Dalal & Carter, 2014). Mixed-worded scales put both positively and negatively worded questions in the same questionnaire. Negatively worded items are "cognitive speed bumps" (Podsakoff et al., 2003) that stop respondents from responding in an automatic pattern. Negatively worded items are made by reversing the direction of a positively worded item. This can be done in two ways (Sonderen et al., 2013; Swain et al., 2008). One way to do this is by introducing words like 'not' or 'no' or affixes like 'un-', 'non-', 'dis-' or '-less'. This method changes the direction of the item but does not change the wording substantially. Another way is to use antonymic words such as using 'I hate this' to negate 'I love this'. When mixed-worded scales are introduced a method effect regarding the choice of wording arises (Dalal & Carter, 2014). The authors point out that researchers should be concerned about the possibility of a method effect if the items are confusing or unclear in any way. They advise to carefully review the type and quality of the negatively worded items. One reason for method effects arising out of use of negatively worded items is that the respondent must also reverse their thinking flow when the wording reversal occurs, which means that the respondent needs to tick the opposite side of the response scale to stay consistent with their responses to the positively worded items in the mixed-worded questionnaire (Steinmann et al., 2021a). For example, in a mixed-worded scale (see figure 1) where a respondent wants to express positive belongingness, they have to agree with an item such as “I feel like I belong at school” and disagree with an item such as “I feel like an outsider (or left out of things) in school”. Respondents who fail to switch their response to the opposite side when items are reversed are considered inconsistent responders.

Figure 1

Examples of Consistent and Inconsistent Responders to Positively and Negatively Worded Items.

Consistent response

	Strongly agree	*Agree*	*Disagree*	*Strongly disagree*
I feel like I belong at school.	☐01	☐02	☐03	☒04
I feel like an outsider (or left out of things) at school.	☒01	☐02	☐03	☐04

Inconsistent response

	Strongly agree	*Agree*	*Disagree*	*Strongly disagree*
I feel like I belong at school.	☐01	☐02	☐03	☒04
I feel like an outsider (or left out of things) at school.	☐01	☐02	☐03	☒04

Note. The figure shows consistent and inconsistent response to a positively and a negatively worded item. Items are taken from the *st034* scale of the PISA 2018 student questionnaire.

Two possible reasons for inconsistent responses appear in the literature. The first one arises due to a lack of reading proficiency or cognitive skills. In this case, responders fail to notice the change in item wording or fail to reverse their thinking flow (Bolt et al., 2020; Steinmann et al., 2021a; Weems et al., 2003). In the second case, inconsistency arises due to a lack of careful handling of the questionnaire due to distraction, hurry, or less commitment (Kam & Meyer, 2015; Quilty et al., 2006; Steinmann et al., 2021a; Weems et al., 2003). In this case, responders fail to detect the changing of the item wording and fail to reverse their thinking flow due to lack of attention.

Why do we need to find out Inconsistent Responders?

Inattentive responding is a source of measurement error that hinders meaningful results. Therefore, identifying and removing inattentive respondents before data analysis is vital for reducing error variances and increasing the self-reported data's statistical power (Maniaci & Rogge, 2014). McGrath et al. (2010) pointed out that, inattention negatively affects both individual and aggregate levels. At the individual level, inattention invalidates the interpretation and the use of the scores. However, the aggregate effects of insufficient effort in responding on construct validity evidence are uncertain. Inattentive responses might result in a weak correlation with other measures, reduce reliability, distort factor structure (McGrath et al., 2010), and reduce the power in multiple regression (Maniaci & Rogge, 2014) and t-tests (DeSimone & Harms, 2018).

Steedle et al. (2019) examined how insufficient effort in responding affects the construct validity evidence in accessing social-emotional learning competencies. First, the authors administered nine methods of detecting insufficient effort responding to a social-emotional learning assessment. These methods identified 0.9 to 20.3 percent of responders as insufficient effort responders. Next, the authors removed the flagged responders from the data to examine the effects of insufficient effort responding on several types of construct validity evidence for a self-report measure of social-emotional learning competencies. Unlike the previous studies, Steedle et al. (2019) found negligible or small improvements in criterion-related validity, coefficient alpha, concurrent validity, and confirmatory factor analysis model-data fit.

Steinmann et al. (2021a) asked how the mean score, dimensionality and reliability of mixed-worded scales is affected when inconsistent responders are removed. They also looked into how the interplay between mixed-worded sales and external variables is affected when inconsistent responders are removed. Consistent with their hypothesis, they found that the mean scores of the three self-concept scales (reading, mathematics, and science) increased upon the removal of inconsistent responders, indicating that including inconsistent responders biases means toward the midpoint of the scales. The authors hypothesized that inconsistent responders cause wording-related covariance, and thus, factor analysis is likely to suggest multiple underlying factors. Consistent with this, they find a one-dimensional structure in most education systems when inconsistent responders are removed, apart from those with a high to a moderate proportion of inconsistent responders. For the reliability measure, the authors expected that removal of inconsistent responders would lead to higher reliability. For the reading self concept scale, they found that reliability only slightly increase when inconsistent responders are removed. But for the mathematics and science self concept scale, they found that for education systems that have a larger proportion of inconsistent responders, their removal leads to stronger reliability. The authors find no difference in the relationship between achievement scores and the mixed-worded scales when comparing groups with and without inconsistent responders.

Reading Proficiency

Compared to readers in higher school grades, beginning readers are perhaps more likely to miss the mixed wording or fail to reverse their thinking flow when encountering a mixed worded scale. They might struggle more with decoding and handling the change between positively and negatively worded items. Steinmann et al. (2021b) for instance showed that inconsistent responding was more common among students that scored low on reading achievement tests among primary and secondary school students. Similarly, Steinmann et al. (2021a) show that inconsistent responders scored almost one standard deviation lower on a reading test, than consistent responders at the end of primary school. Looking at data of 15-year-old students at the country level, Montazerikafrani (2021) finds that reading ability explains a large part of the variation in inconsistency.

Language spoken at home

Using mixed-worded scales, researchers can look into the characteristics of inconsistent responders, remove the inconsistent responders to make a clean dataset and then run their analysis (Steinmann et al., 2021a; Steinmann et al., 2021b), try to understand reasons behind inconsistent reading, recommend how questionnaires can be resigned, or look into what affects the probability that a responder responds inconsistently. Steinmann et al. (2021a) find that the share of inconsistent responders range from below 5% to around 33% across different education systems around the world for reading, mathematics, and science self-concept scales. They identify that future research can hypothesize that this variation can be attributed to difference in achievement scores, to what language is spoken at home, and to language characteristics. Montazerikafrani (2021) looked into how the culture in an education system, the GDP of the country, and achievement levels affects the probability of responding inconsistently. In this paper I will look into how the probability of responding inconsistently is affected when a responder speaks a different language at home than the language they answer the questionnaire in.

When we use our native language in our daily lives, we fluently use numerous lexicalised sequences from our memory that have formed from close collocational links of words over time (Foster, 2013). We do this as a routine rather than always using our grammatical knowledge. However Foster (2013) finds that non native speakers use the local language more with the help of rules rather than as routines.

In a study on 102 third and fourth grade students in China Zhou and McBride (2018) finds "cognitive similarities and differences in reading among native and non-native Chinese speakers". Native Chinese speakers performed significantly better than non-native speakers on reading skills with Chinese words while the latter performed better in English vocabulary and English working memory. Whereas attributes like Chinese vocabulary Chinese working memory were found to be positively correlated to Chinese word reading, for the non-native speakers an additional attribute, visual skills, was also found to be positively correlated. In a study with 884 native English sixth-graders and 284 sixth-graders who speak English as a second language, Low and Siegel (2005) finds that the latter group lags behind the former in terms of the ability to manipulate and reflect on the grammatical structure of language (Cain, 2007).

Given that mixed worded scales include items with reverse wording, inconsistency in responses can arise when there is a change in syntactic awareness. I hypothesize that non-native speakers, who are responding to questions formulated not in their native language, may show a lower syntactic awareness and thus are more likely to answer inconsistently to items that have reverse wording.

Research questions

By analyzing a mixed-worded scale where there are three positively and three negatively worded items about the sense of belongingness in school, I will look into how the probability of responding inconsistently vary across different education systems for native vs non-native speakers. I am asking the following:

1. How the proportion inconsistent respondents vary across the education systems?
2. Is a respondent less likely to respond inconsistently if they score higher in reading?
3. Is a respondent more likely to respond inconsistently if they speak a different language at home?

Method

For this study, I used the data from PISA (Programme for International Student Assessment) 2018. PISA is an international survey conducted every three years in many regions worldwide. PISA is an assessment for the 15-year-old students to assess their acquired ability for full participation in social and economic life. PISA focuses on any one of the core subjects, namely maths, reading, or science, each year. PISA alters the core subjects every three years and presents a thorough analysis of achievement in each of the three core subjects every nine years (OECD, 2019b).

Sample

In the 2018 cycle, the primary domain of assessment in PISA was reading. In many regions of the world, PISA is considered the assessment tool for educational systems (OECD, 2019b). The number of participant countries and economies varies in each assessment cycle. For example, 80 education systems participated in the 2018 assessment cycle. The sample size varied for each country, ranging between 2,016 from Moscow region (Russia) to 35,943 from Spain. I have excluded Vietnam (because plausible values were missing in the data), North Macedonia, Lebanon, and Israel (because data for all the 6 items in the belong scale were not available). I analyze the remaining 76 education systems.

Measures

Attempting to explain the inconsistency in responses, Montazerikafrani (2021) looked into the following predictors: cultural aspects like individualism and tightness, gross domestic product, and average reading literacy of the countries in PISA 2018. When reading literacy was included as a predictor of inconsistency, the author found that the other predictors became insignificant, leaving reading literacy the lone significant predictor in the model. The inclusion of reading literacy also boosted the R-squared statistic of the regression. Using this information, I consider reading literacy an acceptable control variable as I attempt to understand how speaking another language at home rather than the test's language alters the probability of responding inconsistently. Thus, my outcome variable is whether a student is an inconsistent responder. The predictors are reading literacy and language spoken at home.

Inconsistent responders

To identify inconsistent respondents for my study, I used the sense of belonging at school scale (ST034) from the PISA 2018 student questionnaire. This scale contains a balanced number of positively and negatively worded items, three of each. The sense of belonging at school shows how accepted, respected, and supported a 15-year-old student feels in their social context at school (Goodenow & Grady, 1993; OECD, 2019c). In addition, this scale helps us to understand school connectedness, school attachment, school engagement, school identification, and school bonding (OECD, 2019c; Slaten et al., 2016). This scale contains three positively worded and three negatively worded items. Table 1 shows the original items according to original order, the same as in PISA 2018 questionnaire. In this scale the negations were made by using general antonymic words and phrases. When negations are formed in this way, it requires more attention to notice them (Montazerikafrani, 2021).

Table 1

Items in the Belong scale

Thinking about your school: to what extent do you agree with the following statements?		
Item wording	Wording direction	Item name
I feel like an outsider (or left out of things) in school	–	ST034Q01TA
I make friends easily at school.	+	ST034Q02TA
I feel like I belong at school.	+	ST034Q03TA
I feel awkward and out of place in my school.	–	ST034Q04TA
Other students seem to like me.	+	ST034Q05TA
I feel lonely at school.	–	ST034Q06TA

Note. For each item, the response categories included 1 = agree a lot, 2 = agree a little, 3 = disagree a little, and 4 = disagree a lot. Positively worded items are indicated by '+' and negatively worded items by '–'.

The predictors: Reading Literacy and Language spoken at home

Large-scale assessment surveys like PISA use plausible values for reading literacy scores. Since such surveys involve large number of items and there is time limitation for conducting them (Laukaityte & Wiberg, 2017), all students do not answer the same set of test items; instead a rotation system of question booklets is used (Gilleece, 2015). This leads to measurement error in terms of assessing individual proficiency of a student (Von Davier et al., 2009). Multiple plausible values are then generated by imputing multiple achievement scores using background data and students' responses to test items (Gilleece, 2015). Using plausible values allows for estimating scores for the missing items which leads to an overall estimate of achievement (Gilleece, 2015). PISA 2018 uses 10 plausible values for reading. Table.... shows their means and standard deviation across the education systems. In PISA 2018, language spoken at home is a binary variable where the respondent answers whether *language spoken at home is the same as the language of the assessment* or *not*.

Statistical Analysis [1]

Inconsistent Responder Detection

There are several different methods to identify inconsistent responders of self-reported data (Hong et al., 2020; Steedle et al., 2019). In this study, the mean absolute difference method has been used to identify inconsistent responders. The mean absolute difference method primarily detects responders who did not distinguish between positively and negatively worded items. The average score difference between positively and negatively worded items is calculated in the mean difference method. For items scored in the 1-6 range, if the mean absolute difference between the average scores on positively and negatively worded items is more than or equal to 2.0 (ACT, 2016; Steedle et al., 2019), then a respondent is flagged as inconsistent (Steedle et al., 2019). Respondents who cannot pay attention to the negatively worded items are expected to show larger average score differences between the positively worded and negatively worded items after coding all items in the same direction (Hong et al., 2020; Steedle et al., 2019). For this study, I fist take the average response of the negatively worded items (NWM). Then I reverse code the positively worded items (PWM) and take its average $(5-PWM)$. The mean absolute difference is therefore $(|NWM-(5-PWM)|)$. This quantifies the extent to which a responder does not change to the opposite side of a likert scale with four levels when faced with an item with reversed

[1] All calculations and analysis for this thesis have been conducted using the R software (R Core Team, 2020).

wording. A respondent is considered to be more inconsistent when this mean absolute difference is higher. Like Steinmann et al. (2021a), I set the cut-off point for the mean absolute difference at 1.75 to consider a respondent as inconsistent. I use the same cut-off point for all the educations systems in this study.

Logistic Regression

I used separate logistic regressions for each education system to explore if the probability of 15-year-old students responding inconsistently may be explained by what language they speak at home and their reading literacy scores. PISA 2018 reports 10 plausible values for reading literacy. As suggested in the PISA Data Analysis Manual by OECD (2009), 10 separate regressions needs to be carried out one for each of the plausible values. The average of the parameter estimates from these separate regressions are calculated and reported. This process is carried out with the help of the *withPV* command from the R package 'survey' version 4.1-1 (Lumley, 2020). This package has also been used to incorporate student level weights in the model design.

Results

The outcome variable for this study is whether a student is an inconsistent responder or not. Figure 2 shows the prevalence of inconsistent responders across the education systems with lowest in Philippines at around 9% to highest in Georgia at around 33%. Average number of inconsistent responders across all the education systems is around 17%.

One of the predictors explored in this paper is the reading score of the students. Table 2 shows the mean and the standard deviation of the plausible values (from 1 to 5) of reading score across education systems sorted the mean of the first plausible value of reading. B-S-J-Z (China) has the highest average reading score at around 560, while Philippines has the lowest at around 339. Table 3 continues with the next five plausible values (from 6 to 10).

The other predictor explored in this study is whether a student speaks the language of the test at home or speaks another one. Figure 3 shows the percentage of students who speak a language other than language of the test at home. In Philippines around 95% of the students speak a different language at home other than the language of the test. In B-S-J-Z (China) almost all student speak the language of the test at home.

Taking reading score and inconsistent respondents together, it can be seen from Figure 4 that in 18 out of the 76 education systems, average reading score is higher for inconsistent responders. It can also be observed that these 18 education systems are quite evenly spread from top to bottom showing that inconsistent responders can be found in education systems with both higher and lower average reading proficiency.

Taking reading score and language spoken at home together, it can be seen from Figure 5 that in 7 out of the 76 education systems, average reading score is higher for students that speak a different language. In most of the remaining education systems, the gap in average reading score is greater when looking from the language spoken at home scenario. The gap is lower when looking at reading scores for the consistent and inconsistent responders groups.

Figure 2

Prevalence of inconsistent responders across education systems

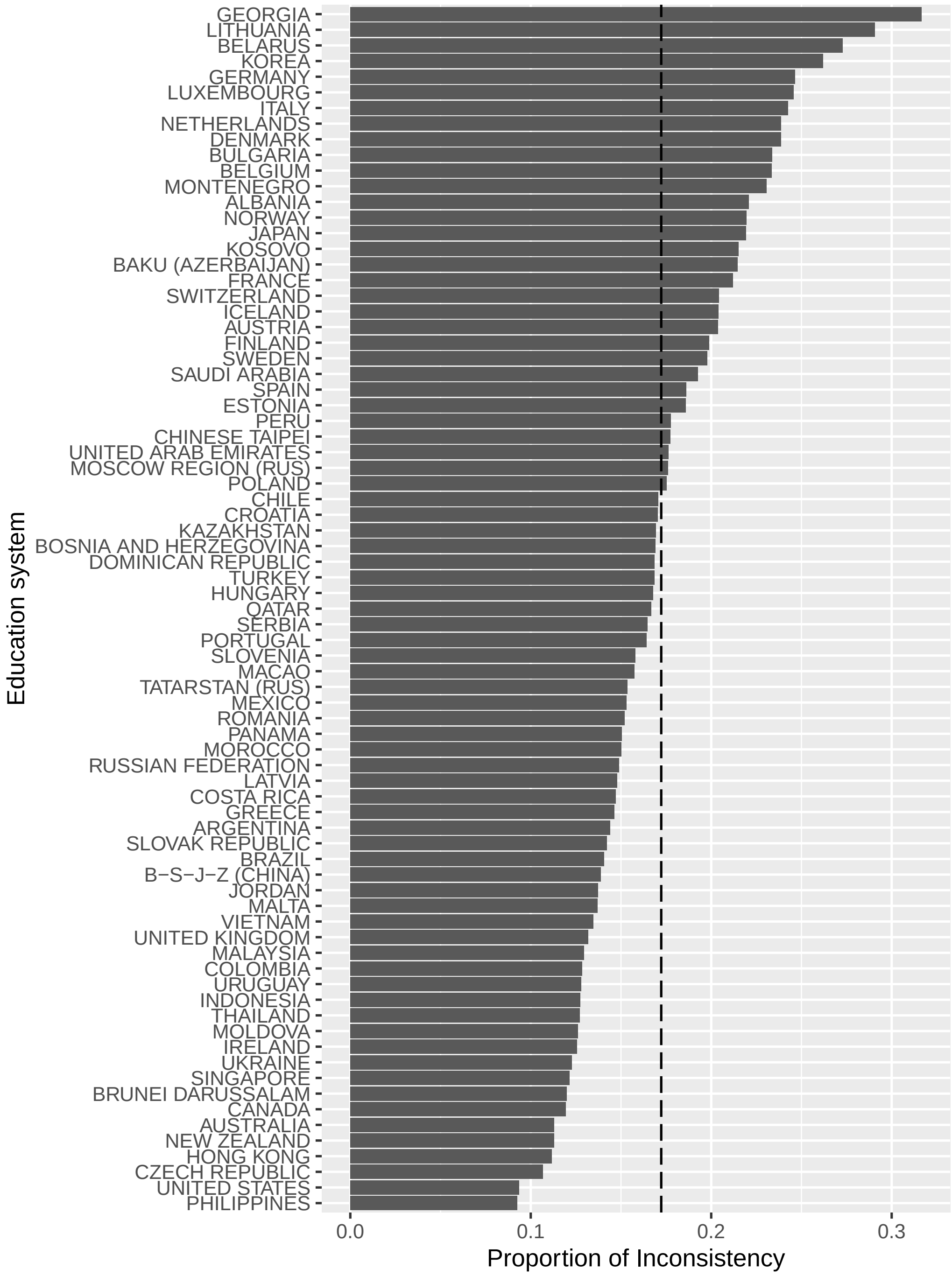


Table 2

Mean and standard deviation of plausible values (1 to 5) of reading

Education system	pv1read_Mean	pv1read_SD	pv2read_Mean	pv2read_SD	pv3read_Mean	pv3read_SD	pv4read_Mean	pv4read_SD	pv5read_Mean	pv5read_SD
B-S-J-Z (CHINA)	561.03	90.34	560.27	90.05	560.80	89.83	560.27	89.80	560.39	90.02
SINGAPORE	548.58	110.34	548.00	109.79	549.51	111.38	548.32	110.32	548.27	109.84
HONG KONG	527.33	97.90	527.36	97.87	527.55	97.42	527.15	97.19	527.26	97.08
MACAO	525.08	92.51	525.55	91.64	524.45	91.35	525.35	92.38	524.86	92.65
ESTONIA	523.70	92.60	524.04	93.75	522.80	92.65	523.41	93.77	523.02	93.37
FINLAND	519.94	99.58	520.40	99.33	521.12	100.09	520.37	99.47	520.16	98.82
IRELAND	518.04	90.45	517.53	90.30	518.23	89.93	517.43	91.42	517.47	89.72
KOREA	515.58	101.22	514.76	100.77	515.67	101.61	516.23	101.42	516.11	100.97
POLAND	513.13	96.95	512.80	96.35	513.48	96.90	512.27	96.53	512.25	96.51
CANADA	509.47	101.87	509.15	102.18	508.41	101.64	508.28	101.70	508.68	101.63
NEW ZEALAND	507.89	104.49	507.72	105.45	508.46	105.88	508.26	105.44	508.57	106.09
CZECH REPUBLIC	506.77	99.63	507.44	100.49	506.31	100.44	506.98	101.02	506.77	100.52
SWEDEN	505.44	106.44	506.32	106.52	505.55	107.77	505.33	106.64	505.15	107.52
JAPAN	502.91	97.15	504.04	97.40	503.28	97.64	504.00	96.95	503.73	97.81
AUSTRALIA	502.47	108.43	503.62	108.92	503.06	108.88	502.26	109.21	502.67	108.70
GERMANY	500.85	106.27	499.88	105.55	499.82	105.49	500.09	105.68	501.17	104.89
UNITED KINGDOM	500.50	97.81	499.31	98.46	499.28	98.24	500.24	98.08	499.23	97.83
UNITED STATES	500.15	108.45	500.79	107.95	500.30	107.90	501.10	108.48	500.48	108.08
CHINESE TAIPEI	497.66	102.08	497.57	101.48	498.70	102.31	498.02	102.01	498.08	102.40
NORWAY	497.39	105.97	498.11	106.25	497.57	105.83	498.35	106.25	497.40	106.14
BELGIUM	495.31	101.67	495.82	102.26	495.64	102.50	495.51	101.55	495.72	101.66
PORTUGAL	490.88	96.18	490.06	95.69	490.19	95.68	490.88	95.78	490.14	95.51
DENMARK	489.06	94.92	488.38	94.23	488.42	94.95	488.03	94.16	488.18	94.31
MOSCOW REGION (RUS)	487.29	90.21	487.16	90.61	484.84	89.57	486.00	91.22	486.66	90.79
AUSTRIA	486.40	98.46	487.42	98.58	486.99	98.16	486.99	98.39	487.44	98.31
SWITZERLAND	484.51	101.05	483.93	100.89	484.20	100.67	484.46	100.39	483.53	100.41
FRANCE	484.27	105.40	484.22	105.46	484.13	105.84	484.01	105.50	484.10	105.72
SPAIN	483.15	91.82	482.20	91.46	482.48	91.14	482.98	91.79	482.97	91.64
HUNGARY	483.04	96.63	482.78	96.86	482.63	97.19	483.29	96.91	482.68	97.22
RUSSIAN FEDERATION	480.71	93.55	479.93	93.45	480.48	93.31	480.07	93.64	480.42	92.84
ITALY	480.55	93.76	480.89	94.47	480.85	94.13	480.59	93.71	481.43	93.86
SLOVENIA	479.70	94.55	480.17	93.59	479.88	92.86	480.02	93.77	480.08	94.01
NETHERLANDS	479.53	106.05	479.87	106.23	479.51	105.82	479.43	105.56	481.15	106.39
CROATIA	477.53	89.23	477.75	89.35	477.47	88.82	477.16	89.29	477.69	89.25
LATVIA	475.62	88.65	476.61	89.56	476.35	89.23	476.21	88.69	475.55	88.57
BELARUS	475.57	89.13	475.08	88.72	475.29	88.81	475.78	89.45	475.61	88.61
ICELAND	473.07	103.34	472.43	103.55	473.43	103.83	473.61	105.09	472.56	103.55
LITHUANIA	470.52	94.23	471.61	94.39	470.92	95.02	471.39	93.94	470.83	93.65
LUXEMBOURG	470.22	108.62	470.76	108.42	470.43	109.12	470.63	108.36	471.26	109.12
CHILE	469.80	93.28	469.33	93.65	469.16	93.58	469.06	93.17	470.00	93.63
UKRAINE	468.12	92.87	468.58	92.96	467.80	92.48	468.80	92.61	467.88	92.53
TATARSTAN (RUS)	465.05	92.08	465.30	91.36	465.47	91.41	465.14	91.81	464.87	92.08
TURKEY	464.23	87.78	464.42	87.70	464.71	87.08	464.61	87.40	464.20	87.21
GREECE	460.48	96.85	461.39	96.44	460.32	96.63	460.46	95.65	460.40	96.13
SLOVAK REPUBLIC	460.31	100.70	459.76	100.48	460.40	100.84	459.85	100.91	460.56	100.86
MALTA	450.19	112.50	450.30	112.29	450.00	112.33	449.97	112.42	449.78	111.59
SERBIA	439.65	96.51	440.41	95.95	439.54	95.39	440.18	95.51	439.98	95.94
MEXICO	427.81	82.17	427.79	81.69	427.74	81.70	427.66	81.68	427.82	81.97
ROMANIA	427.57	95.60	428.40	96.71	427.30	95.73	428.94	96.24	428.31	96.38
URUGUAY	426.02	96.42	427.68	96.83	427.14	96.26	426.88	96.26	426.49	95.06
MOLDOVA	425.69	93.10	424.93	93.98	425.63	93.06	425.62	93.60	426.40	93.59
COSTA RICA	425.58	79.09	424.19	78.61	425.75	78.48	424.84	78.34	425.14	78.58
BULGARIA	423.33	101.04	423.45	100.69	423.68	101.35	423.78	101.00	422.87	101.05
COLOMBIA	422.32	89.86	422.48	89.83	422.29	89.72	422.40	89.94	422.91	89.53
MONTENEGRO	421.76	85.91	421.71	86.27	421.89	86.48	421.51	85.40	421.84	85.54
UNITED ARAB EMIRATES	421.20	113.32	421.67	113.12	422.35	112.84	421.99	113.39	422.04	112.53
JORDAN	419.60	85.64	419.31	85.26	419.46	85.93	419.81	85.62	419.90	85.37
MALAYSIA	416.23	83.09	416.42	83.39	416.44	83.41	415.83	83.27	415.98	83.36
BRAZIL	415.79	98.46	415.60	97.88	415.99	98.42	415.72	98.20	415.43	98.58
ARGENTINA	415.03	97.01	415.90	96.75	414.40	96.51	414.95	95.90	415.22	96.19
THAILAND	409.49	88.90	408.40	88.01	408.58	88.14	409.00	88.67	409.38	88.70
BRUNEI DARUSSALAM	409.17	97.16	407.67	97.67	407.93	98.02	407.86	97.58	408.57	97.93
QATAR	408.27	109.38	407.97	109.15	408.05	109.16	408.36	108.87	408.52	109.28
ALBANIA	406.82	80.10	407.47	80.32	406.86	80.90	406.78	80.13	406.74	80.31
KAZAKHSTAN	404.49	87.08	404.50	86.70	404.63	87.25	404.59	87.46	404.62	87.77
BOSNIA AND HERZEGOVINA	403.17	80.10	403.37	79.41	403.92	79.68	403.03	79.45	402.84	79.58
PERU	402.66	91.15	402.58	90.64	402.80	90.62	402.70	91.85	402.91	91.53
SAUDI ARABIA	402.15	83.09	402.86	84.36	402.13	84.13	402.02	84.02	402.71	83.64
INDONESIA	390.10	81.64	390.39	82.39	390.57	82.17	390.35	81.86	389.97	82.15
BAKU (AZERBAIJAN)	389.32	73.80	389.09	74.06	388.81	74.23	389.25	73.98	389.37	73.84
GEORGIA	381.16	83.77	381.07	84.93	380.30	85.41	381.39	85.12	382.62	84.79
PANAMA	378.90	85.98	377.82	86.86	377.92	86.11	377.93	86.19	378.34	86.45
MOROCCO	358.40	74.82	358.34	74.74	358.55	75.05	358.20	74.07	357.84	74.67
KOSOVO	349.79	68.56	350.65	68.38	350.91	68.36	350.92	67.70	350.81	68.95
DOMINICAN REPUBLIC	344.30	81.27	344.75	81.13	343.69	81.85	344.21	81.84	345.69	81.50
PHILIPPINES	338.56	78.51	339.06	79.31	338.78	77.55	338.85	79.05	338.61	78.93

Note. The education systems are sorted by the mean of plausible value 1.

Table 3

Mean and standard deviation of plausible values (6 to 10) of reading

Education system	pv6read_Mean	pv6read_SD	pv7read_Mean	pv7read_SD	pv8read_Mean	pv8read_SD	pv9read_Mean	pv9read_SD	pv10read_Mean	pv10read_SD
B-S-J-Z (CHINA)	560.84	89.75	560.77	90.24	559.40	89.44	560.30	89.73	561.08	90.13
SINGAPORE	547.78	110.63	548.23	110.18	547.72	110.29	548.94	110.05	549.24	110.55
HONG KONG	527.01	97.59	526.73	97.25	526.78	97.27	527.78	97.32	526.70	97.23
MACAO	525.30	92.92	525.59	91.62	524.77	91.81	524.63	92.11	525.14	92.43
ESTONIA	522.44	92.35	523.25	93.26	522.92	92.72	523.76	93.43	523.42	92.88
FINLAND	521.00	99.43	519.89	99.24	519.93	99.03	520.44	99.24	520.02	98.90
IRELAND	517.80	89.76	518.25	90.64	517.08	89.84	517.03	90.66	518.21	90.43
KOREA	515.66	101.68	516.68	101.57	515.55	101.56	515.83	101.32	515.41	101.50
POLAND	512.73	96.09	512.04	96.29	512.55	97.00	512.38	96.10	512.70	95.99
CANADA	508.57	102.27	508.79	101.61	508.89	101.89	508.23	101.86	507.98	102.10
NEW ZEALAND	508.23	105.49	508.57	105.77	508.25	105.98	508.14	105.51	508.06	105.75
CZECH REPUBLIC	506.90	99.88	507.09	100.22	506.94	100.62	506.50	100.39	506.85	99.84
SWEDEN	505.94	107.13	505.38	107.57	505.60	107.41	504.90	106.43	505.63	107.41
JAPAN	503.10	96.55	503.64	97.20	503.35	97.54	503.52	97.29	503.24	97.67
AUSTRALIA	503.17	108.89	502.31	108.92	503.21	108.48	502.88	109.12	502.94	108.08
GERMANY	499.86	105.89	500.11	106.04	500.21	106.42	500.42	106.18	499.92	106.53
UNITED KINGDOM	499.68	97.60	500.05	97.98	499.02	97.68	499.78	98.15	498.79	97.80
UNITED STATES	500.95	108.19	499.89	107.71	500.30	108.11	501.08	107.43	500.63	107.93
CHINESE TAIPEI	498.14	102.18	498.38	102.05	498.47	102.67	497.61	101.70	497.40	102.38
NORWAY	498.76	106.09	498.01	105.60	497.47	105.71	497.53	105.96	497.50	105.36
BELGIUM	495.68	101.98	495.43	102.15	494.73	101.64	495.50	101.68	495.19	102.03
PORTUGAL	491.50	96.25	490.62	96.30	489.84	95.91	490.90	96.65	490.97	96.58
DENMARK	488.00	93.86	487.73	94.65	487.82	94.12	487.74	94.30	487.43	94.62
MOSCOW REGION (RUS)	486.35	90.92	486.79	91.45	487.15	91.42	486.29	90.33	486.38	92.48
AUSTRIA	487.71	98.23	487.58	98.93	487.34	98.57	486.41	98.45	487.08	98.47
SWITZERLAND	484.14	100.20	485.35	101.29	484.84	100.54	484.89	100.64	484.39	100.61
FRANCE	483.37	104.62	482.97	105.84	483.74	105.45	483.22	105.36	483.86	105.52
SPAIN	482.59	91.26	482.84	91.66	482.14	91.73	482.97	91.78	482.99	91.45
HUNGARY	482.75	96.30	483.15	96.86	482.60	96.59	483.56	96.61	482.59	96.43
RUSSIAN FEDERATION	479.77	93.43	480.65	93.64	480.30	94.47	480.29	93.38	480.72	94.25
ITALY	481.75	94.16	481.16	94.32	480.73	93.78	481.18	94.31	481.77	94.87
SLOVENIA	479.56	93.94	480.55	93.72	479.92	93.82	480.43	93.65	480.70	93.75
NETHERLANDS	479.71	106.74	479.49	105.74	480.02	106.44	478.83	106.22	480.67	105.97
CROATIA	476.73	88.16	477.26	89.86	477.88	89.37	477.46	88.81	476.87	89.30
LATVIA	476.06	88.34	476.29	88.46	476.65	89.52	476.36	89.37	476.01	88.39
BELARUS	475.36	89.00	474.72	88.96	474.67	89.53	475.10	89.23	475.55	89.20
ICELAND	473.80	104.89	473.29	105.13	474.37	104.74	473.44	103.86	473.09	104.66
LITHUANIA	470.91	94.00	471.05	94.23	470.70	94.13	470.89	93.61	471.77	94.49
LUXEMBOURG	470.92	108.50	470.46	107.81	470.82	108.50	469.38	108.38	470.33	107.56
CHILE	469.52	93.79	469.09	93.33	469.06	92.62	469.02	93.12	469.69	93.95
UKRAINE	469.21	92.52	468.57	92.70	468.32	92.57	469.15	93.22	468.30	92.00
TATARSTAN (RUS)	465.49	91.72	464.83	91.76	465.40	92.20	464.61	91.14	464.64	92.46
TURKEY	465.35	88.09	464.74	87.45	464.80	86.65	464.78	87.56	464.06	87.38
GREECE	460.37	95.85	460.29	96.61	461.97	97.23	460.06	96.13	460.46	96.41
SLOVAK REPUBLIC	460.89	101.07	460.74	101.73	461.51	100.48	460.62	101.14	461.09	100.36
MALTA	450.19	112.88	450.44	112.35	450.09	112.78	450.20	112.25	450.12	111.25
SERBIA	439.78	95.60	440.22	96.28	439.23	95.93	439.76	95.21	440.05	96.03
MEXICO	427.63	81.38	427.44	81.55	427.44	81.70	426.84	82.48	428.23	82.21
ROMANIA	428.42	96.57	428.19	97.03	428.08	96.13	429.19	96.31	428.43	95.94
URUGUAY	426.90	96.37	427.12	95.26	427.15	96.51	426.99	96.43	426.56	95.88
MOLDOVA	425.83	92.99	425.82	93.12	425.39	93.04	426.95	93.45	425.88	91.19
COSTA RICA	424.51	79.42	425.03	78.39	425.53	78.34	425.04	78.05	425.01	78.10
BULGARIA	422.66	101.87	423.45	101.21	423.31	102.07	423.37	101.58	423.36	101.71
COLOMBIA	423.65	89.86	422.49	89.53	421.90	89.79	422.12	89.51	422.41	89.42
MONTENEGRO	421.79	86.04	422.45	85.86	422.20	85.75	422.16	85.55	421.30	85.68
UNITED ARAB EMIRATES	422.60	113.44	421.84	113.20	421.91	113.29	422.12	113.75	421.51	113.89
JORDAN	419.85	85.59	419.40	86.07	419.18	85.81	419.16	86.13	419.55	86.39
MALAYSIA	416.29	83.44	416.73	83.01	415.99	83.31	416.43	83.00	416.07	83.73
BRAZIL	415.63	98.26	415.78	98.32	415.62	97.89	415.49	98.57	415.46	98.09
ARGENTINA	415.66	97.09	415.63	96.54	415.64	96.13	415.19	96.33	414.99	96.97
THAILAND	409.23	88.87	408.98	89.06	409.32	89.32	409.51	89.22	409.31	88.29
BRUNEI DARUSSALAM	408.44	96.34	408.37	97.43	408.31	97.26	407.57	97.55	407.98	97.57
QATAR	408.85	109.24	408.36	108.94	408.31	108.91	408.02	109.93	407.90	109.36
ALBANIA	406.98	80.67	406.70	80.73	407.09	81.15	407.14	80.46	407.03	80.75
KAZAKHSTAN	404.46	86.52	403.61	86.39	404.72	86.81	404.09	87.02	404.54	86.77
BOSNIA AND HERZEGOVINA	402.81	78.97	402.61	79.31	403.46	80.05	402.96	78.91	403.26	79.66
PERU	403.91	90.94	403.53	91.25	402.55	90.92	403.60	90.77	402.29	90.26
SAUDI ARABIA	401.40	84.29	402.84	83.54	402.71	82.97	402.09	84.12	402.21	83.49
INDONESIA	389.99	82.06	389.91	81.27	389.58	81.64	389.99	81.57	390.17	82.42
BAKU (AZERBAIJAN)	389.56	73.31	388.63	73.82	389.09	73.58	389.76	73.44	389.57	73.61
GEORGIA	381.13	84.59	381.80	84.30	381.30	85.15	380.65	84.73	381.58	85.18
PANAMA	378.68	86.64	377.85	85.93	379.04	86.38	377.80	86.81	377.93	85.99
MOROCCO	357.95	73.71	357.99	74.84	358.46	73.94	358.41	74.93	357.83	74.43
KOSOVO	349.78	67.98	350.61	68.28	350.28	68.46	350.20	68.95	350.01	68.67
DOMINICAN REPUBLIC	344.29	81.42	344.90	82.30	345.01	81.18	344.82	82.08	345.96	82.11
PHILIPPINES	339.19	78.81	338.35	79.21	338.52	79.27	339.05	78.91	339.21	79.10

Note. The education systems are sorted by the mean of plausible value 1.

Figure 3

Language spoken at home across education systems

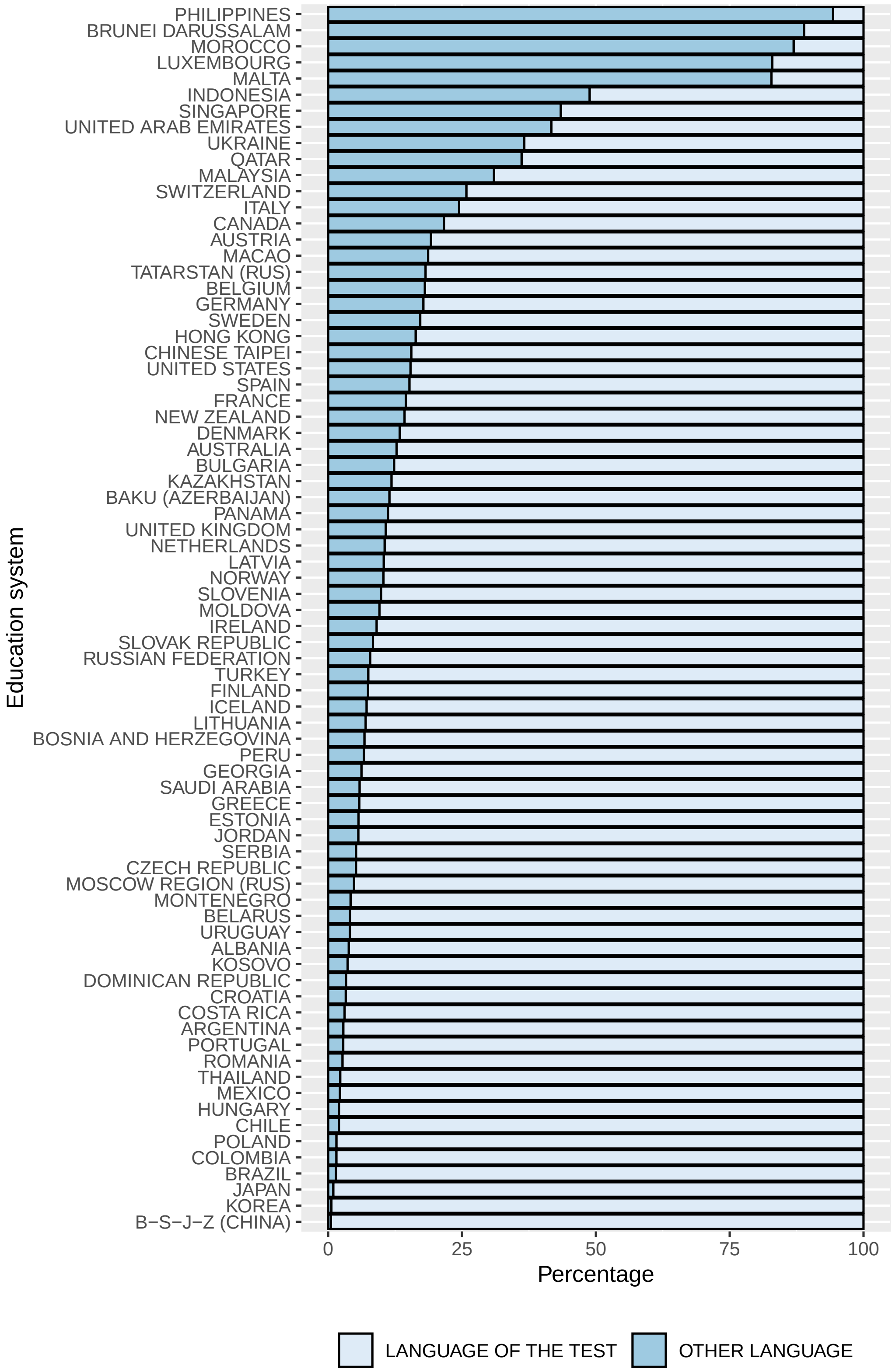


Figure 4

Box plots of plausible value 1 of reading for consistent and inconsistent responders

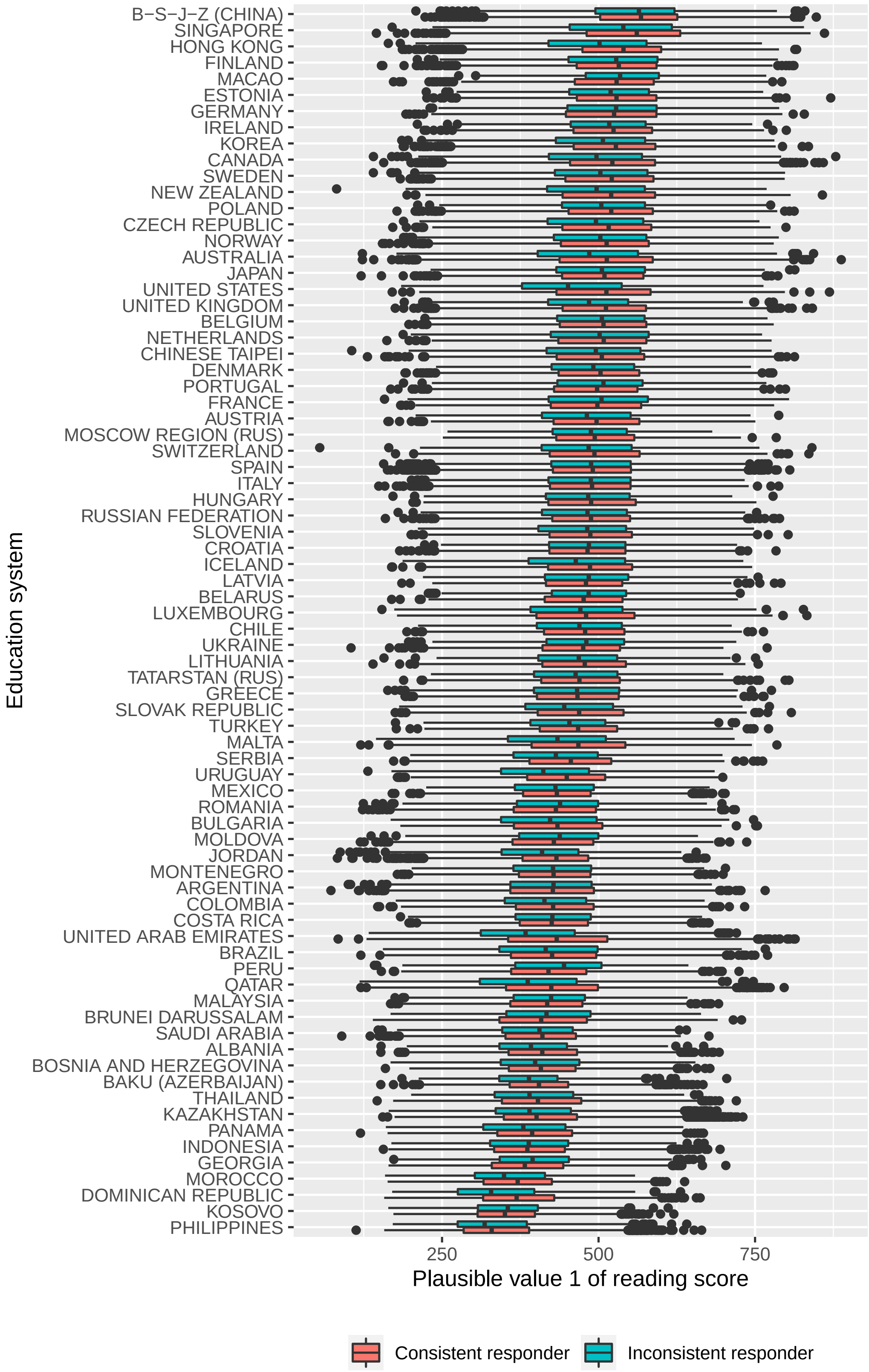
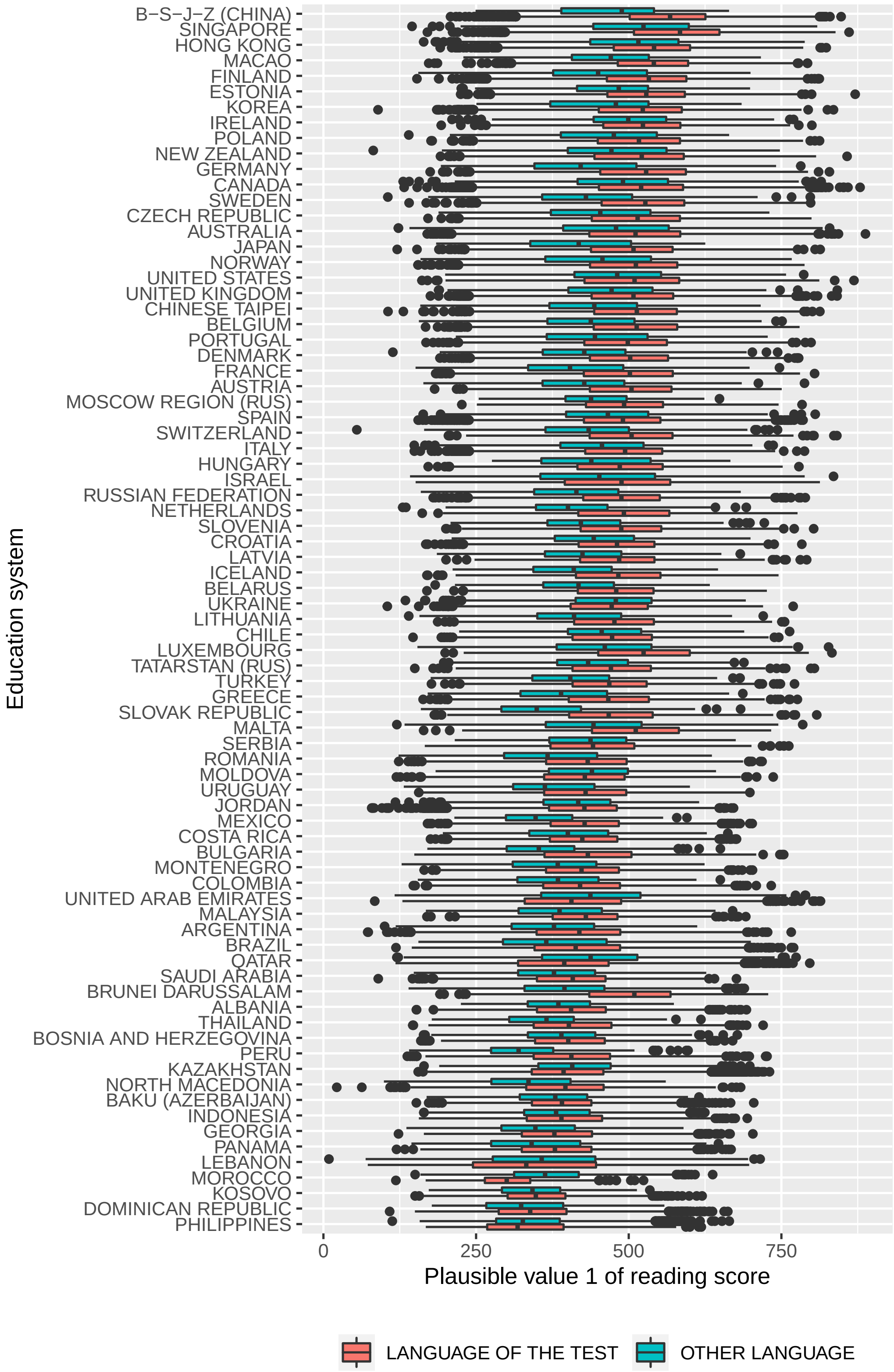


Figure 5

Box plots of plausible value 1 of reading for different languages spoken at home

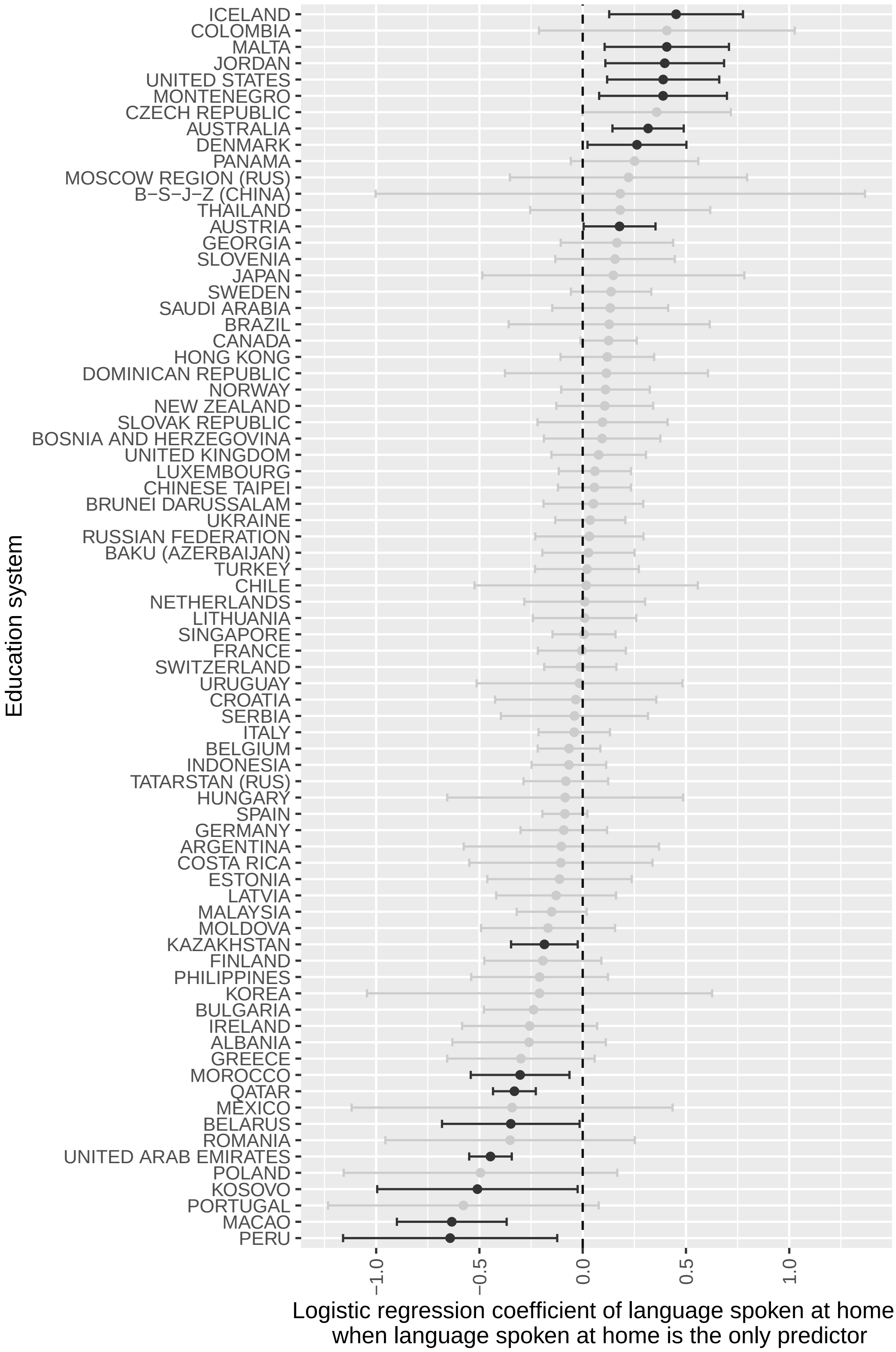


I run three sets of logistic regressions separately on each of the 76 education systems: inconsistent responder on a single predictor of reading score, inconsistent responder on a single predictor of language spoken at home, and finally inconsistent responder on both the predictors. When moving from the single predictor model (see figure 8) to the two predictor model (see figure 9), number of education systems with a negative association between reading score and inconsistent responder remains the same at 45. And the number of education systems with a positive association falls from 5 to 3. Portugal and Macao fall out from having statistically significant coefficient, while Belarus, Georgia and Peru remain. This implies that the number of education systems that are found to have statistically insignificant relation between reading score and inconsistent responder rises from 26 to 28.

When moving from the single predictor model (see figure 6) to the two predictor model (see figure 7), number of education systems with a positive association between another language (not test's language) spoken at home and inconsistent responder falls from 8 to 5. Montenegro, Jordan, Iceland, United States, and Australia reaming with statistically significant coefficient, while Malta, Denmark, and Austria fall out. The number of education systems with a negative association falls from 8 to 7. Kazakhstan, Qatar, Bulgaria, United Arab Emirates, Kosovo, Peru, and Macao continue to have statistically significant coefficient, while Morocco falls out.

Figure 6

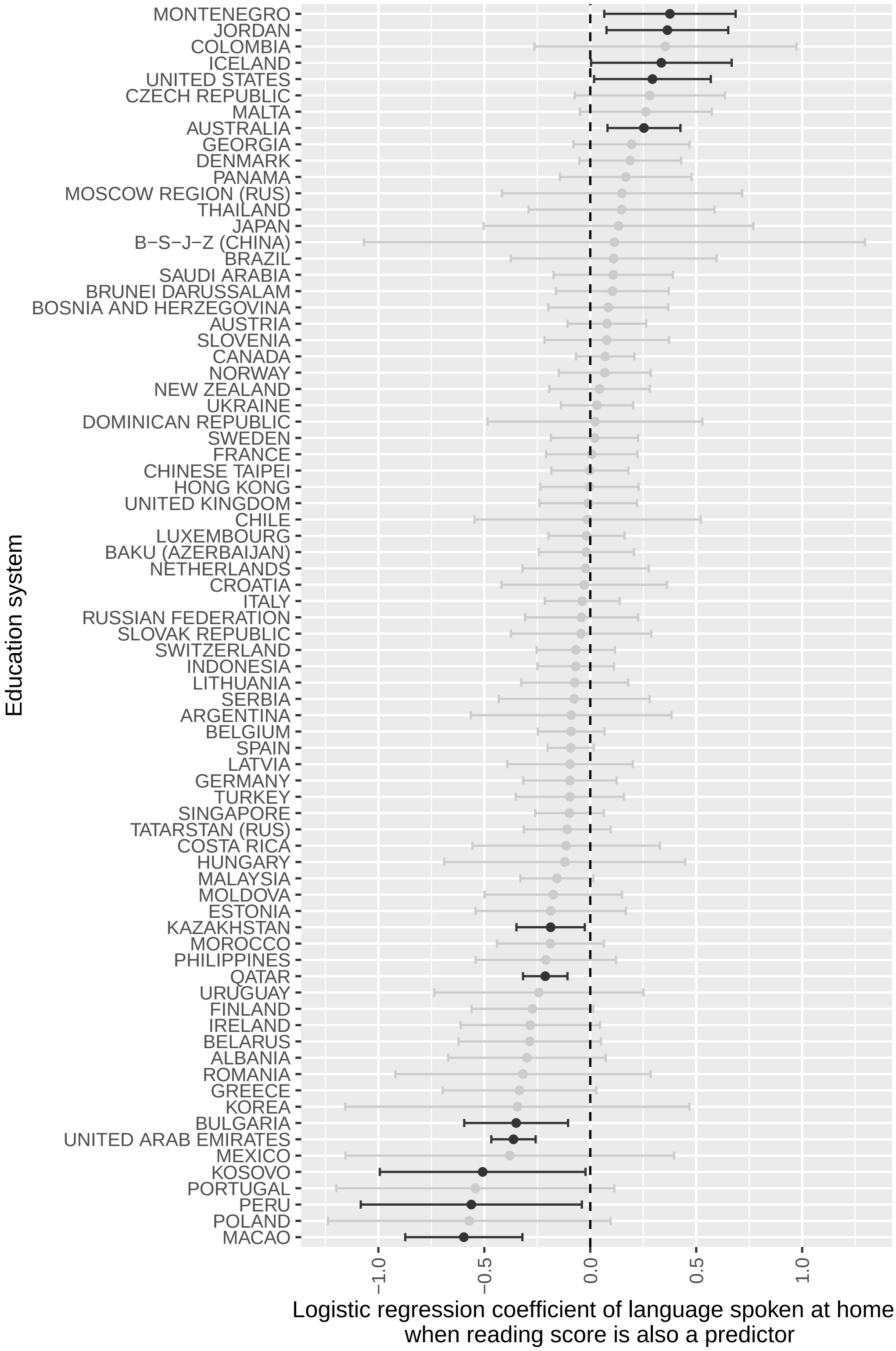
Regression coefficients for language spoken at home (single predictor model)



Note. At 5% significance level, black = significant coefficient, grey = not significant coefficient. Binary code: language of the test = 0; other language =1.

Figure 7

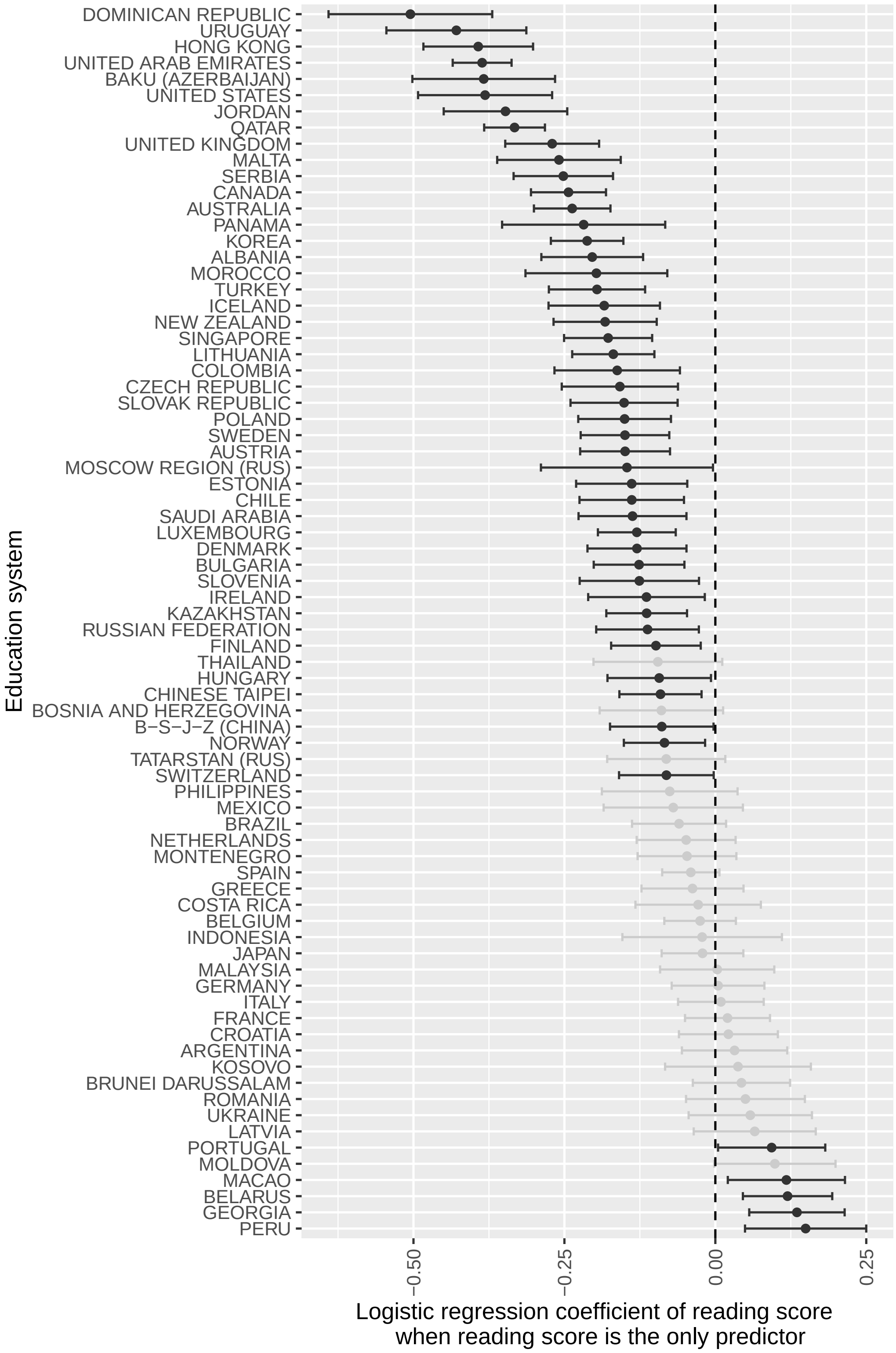
Regression coefficients for language spoken at home (two predictor model: when reading score is also a predictor)



Note. At 5% significance level, black = significant coefficient, grey = not significant coefficient. Binary code: language of the test = 0; other language =1.

Figure 8

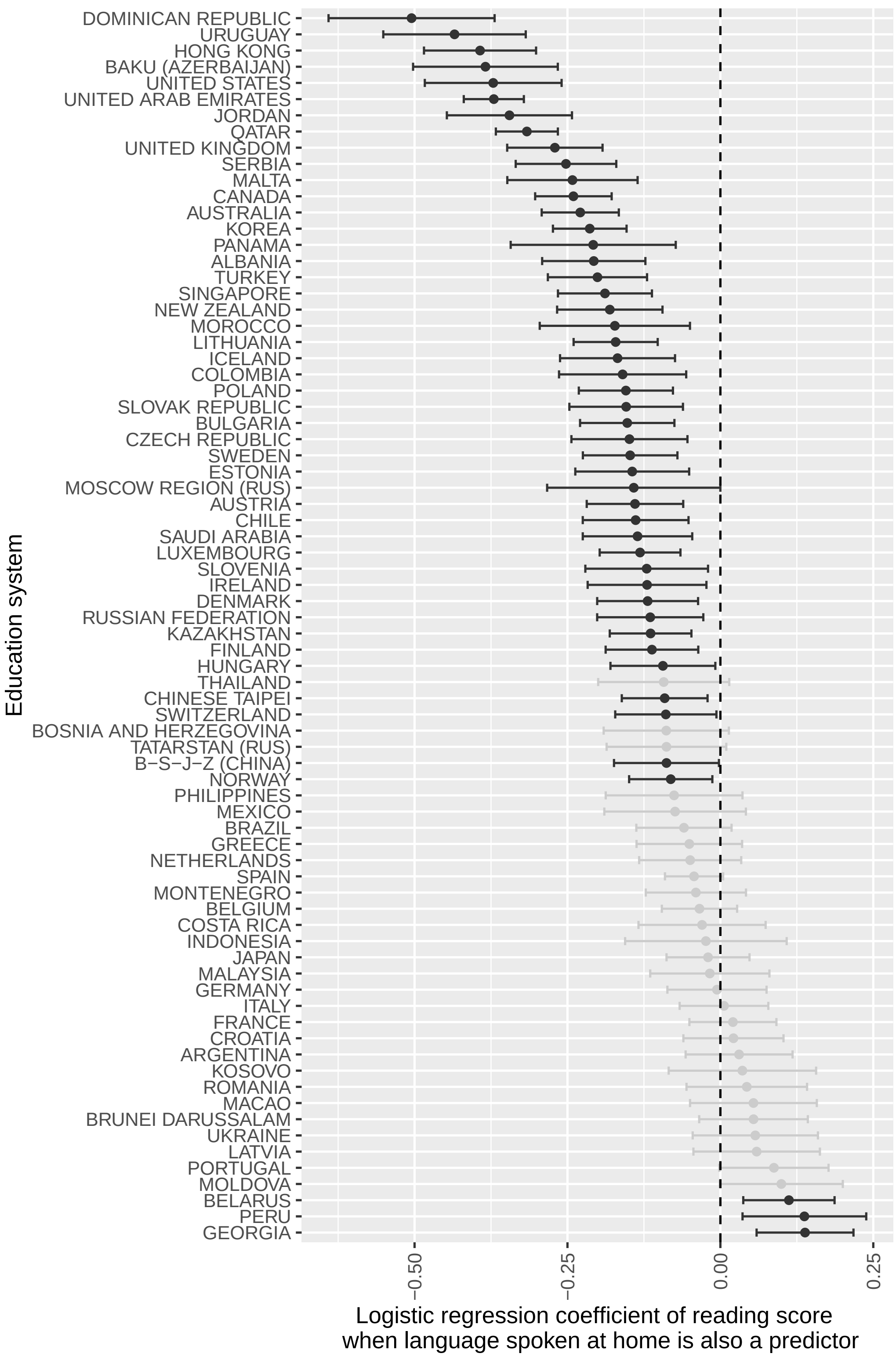
Regression coefficients for reading score (single predictor model)



Note. At 5% significance level, black = significant coefficient, grey = not significant coefficient.

Figure 9

Regression coefficients for reading score (two predictor model: when language spoken at home is also a predictor)



Note. At 5% significance level, black = significant coefficient, grey = not significant coefficient.

Limitations

Students may not take the PISA test seriously as it is a low-stake survey (Akyol et al., 2021). Findings from low-stakes surveys cannot be generalized to other scales, responder groups, or high-stakes scenarios (Steinmann et al., 2021a). Like Steinmann et al. (2021a) who looked at self concept scales of reading, mathematics and science, and like Montazerikafrani (2021) who also looked at the belong scale, generalizing such findings to high-stakes assessments would not be appropriate. Looking at students ranging from grade 3 to grade 12, Bolt et al. (2020) found that the proportion of students who get confused when facing reverse worded items goes down for students studying in higher grades. The participants in my study were high school students. Therefore the conclusion of my study cannot be generalized to all student age groups.

There are various methods for identifying inconsistent responders (Hong et al., 2020; Steedle et al., 2019). In this study, I have used the mean absolute difference method to identify the inconsistent responders. Different methods of identification of inconsistent responders can provide different results. Therefore, generalizing this paper's results across all the methods is not recommended.

Discussion and conclusion

In this study I wanted to see how the proportion of inconsistent responders vary across education systems, if a student is more likely to be an inconsistent responder if they score lower in reading, and if a student is more likely to be an inconsistent responder if they speak a different language at home rather than the language of the test. First, I regressed inconsistent responder on only language spoken at home, then on only reading score, and finally on both of them together as predictors of inconsistent responder. As mentioned earlier, research has found differences in how native and non-native speakers use the local language in terms of grammar, vocabulary, working memory, and visual cues (Cain, 2007; Foster, 2013; Low & Siegel, 2005; Zhou & McBride, 2018). I thus formed the hypothesis that non-native speakers who speak a different language at home is more likely to respond inconsistently when faced with mixed worded scales where some items have reverse wording that requires the respondents to notice the reversal and respond accordingly.

With regards to language spoken at home, it was expected that someone who speaks a different language at home rather than the language of the test is more likely to be an inconsistent responder. However, I find split results for the association between language spoken at home and the probability of responding inconsistently, both when this predictor enters into the model as a single predictor and when it is accompanied by reading score. In the two predictor model, while the estimates for Montenegro, Jordan, Iceland, United States, and Australia seem to conform to expectation, the estimates from Kazakhstan, Qatar, Bulgaria, United Arab Emirates, Kosovo, Peru, and Macao do not. Results from the model with both the predictors point towards association between language spoken at home and being an inconsistent responder for education systems with higher average reading score and those also ones with lower average reading scores. In the latter group, for whom the use of another language at home other than the language of the test is associated with lower chance of being an inconsistent responder, there are middle-eastern countries like Qatar and United Arab Emirates. Studying Qatar's data, Cheema (2014) finds that students from both first and second generation immigrant families outperform native Qatari students in reading, mathematics and science for each gender group, for each grade level they studied, and for different levels of socioeconomic status. For Macao, as per PISA 2003 data, the proportion of students at different reading proficiency levels is similar for the three groups: native, first generation immigrant, and second generation immigrant (Schleicher, 2006). One interesting observation is that the education systems where speaking a different language at home seems to be associated with greater likelihood of inconsistent response, the proportion of inconsistent reponders is more than the average of all the education systems together (17%). For the former group some fall below the average and some above.

To better understand the split results of the association between language spoken at home and inconsistent response found in this study, understanding the characteristics of who speaks what language at home is needed. Data can be analysed for different generations of immigrant families where it can be expected that there will be a difference in reading ability between children from first and from second generation immigrant families even though they may speak their mother tongue at home. It is expected that the socioeconomic status of new immigrant families are on average different in different countries based on the reason why the families immigrated. A model that takes into account these variables will add more explanatory power and help to isolate the individual effect of language spoken at home and inconsistent response.

With regards to reading literacy, it was expected that someone scoring higher in reading maybe is less likely to be an inconsistent responder. In the two predictor model, apart from Belarus, Peru, and Georgia where the estimate came out to be opposite the expectation, estimates for 45 out of the 76 education systems came out with the expected sign. In the remaining 28 education systems the parameter estimate for reading were not significantly different to zero. The composite reading proficiency scale (that is used in this study) comprises of five subscales: retrieving information process, interpreting texts process, reflection and evaluation process, continuous text format, and non-continuous text format (OECD, 2019a). As mentioned earlier, possible reasons for responding inconsistently are lack of reading proficiency or cognitive skills (Bolt et al., 2020; Steinmann et al., 2021a; Weems et al., 2003), and lack of careful handling of the questionnaire due to distraction, hurry, or less commitment (Kam & Meyer, 2015; Quilty et al., 2006; Steinmann et al., 2021a; Weems et al., 2003). The reading composite scale is designed to capture different aspects of reading that relate to why inconsistency in responses can occur. And the results of this study appears to substantiate this design. However, to better dissect the association between reading literacy and inconsistent response, plausible values of the aforementioned sub scales of reading literacy in PISA 2018 data can be used. Instead of the composite scores of reading literacy, using the sub scales that look at separate reading processes and text formats can explain how reading ability is associated to inconsistent response in greater detail.

References

ACT. (2016). Development and validation of act engage: Technical manual.

Akyol, P., Krishna, K., & Wang, J. (2021). Taking pisa seriously: How accurate are low-stakes exams? *Journal of Labor Research*, *42*(2), 184–243.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of marketing research*, *38*(2), 143–156.

Bolt, D., Wang, Y. C., Meyer, R. H., & Pier, L. (2020). An irt mixture model for rating scale confusion associated with negatively worded items in measures of social-emotional learning. *Applied Measurement in Education*, *33*(4), 331–348.

Cain, K. (2007). Syntactic awareness and reading ability: Is there any evidence for a special relationship? *Applied Psycholinguistics*, *28*(4), 679–694. https://doi.org/10.1017/S0142716407070361

Cheema, J. R. (2014). The migrant effect: An evaluation of native academic performance in qatar. *Research in Education*, *91*(1), 65–77. https://doi.org/10.7227/RIE.91.1.6

Dalal, D. K., & Carter, N. T. (2014). Negatively worded items negatively impact survey research. *More statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences*, *112*.

DeSimone, J. A., & Harms, P. (2018). Dirty data: The effects of screening respondents who provide low-quality data in survey research. *Journal of Business and Psychology*, *33*(5), 559–577.

Foster, P. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In: *Researching pedagogic tasks*. Routledge, 2013, pp. 85–104.

Gilleece, L. (2015). Parental involvement and pupil reading achievement in ireland: findings from PIRLS 2011. *International Journal of Educational Research*, *73*, 23–36. https://doi.org/10.1016/j.ijer.2015.08.001

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *The journal of experimental education*, *62*(1), 60–71.

Hong, M., Steedle, J. T., & Cheng, Y. (2020). Methods of detecting insufficient effort responding: Comparisons and practical recommendations. *Educational and Psychological Measurement*, *80*(2), 312–345.

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British journal of applied science & technology*, *7*(4), 396.

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541.

Laukaityte, I., & Wiberg, M. (2017). Using plausible values in secondary analysis in large-scale assessments. *Communications in Statistics - Theory and Methods*, *46*(22), 11341–11357. https://doi.org/10.1080/03610926.2016.1267764

Low, P. B., & Siegel, L. S. (2005). A comparison of the cognitive processes underlying reading comprehension in native english and esl speakers. *Written Language and Literacy*, *8*(2), 131–155. https://doi.org/10.1075/wll.8.2.09low

Lumley, T. (2020). Survey: Analysis of complex survey samples [R package version 4.0].

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological bulletin*, *136*(3), 450.

Montazerikafrani, F. (2021). *Do culture and reading literacy associate with inconsistent responding on mixed-worded scales?* (Master's thesis). University of Oslo. http://urn.nb.no/URN:NBN:no-93748

OECD. (2009). *Analyses with plausible values*. https://doi.org/https://doi.org/https://doi.org/10.1787/9789264056275-9-en

OECD. (2019a). *Pisa 2018 reading framework*. https://doi.org/https://doi.org/https://doi.org/10.1787/5c07e4f1-en

OECD. (2019b). *Pisa 2018 results (volume i): What students know and can do*. OECD.

OECD. (2019c). *Pisa 2018 results (volume iii): What school life means for students' lives*. OECD.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Quilty, L. C., Oakman, J. M., & Risko, E. (2006). Correlates of the rosenberg self-esteem scale method effects. *Structural Equation Modeling*, *13*(1), 99–117.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Schleicher, A. (2006). Where immigrant students succeed: A comparative review of performance and engagement in pisa 2003. *Intercultural Education*, *17*(5), 507–516. https://doi.org/10.1080/14675980601063900

Slaten, C. D., Ferguson, J. K., Allen, K.-A., Brodrick, D.-V., & Waters, L. (2016). School belonging: A review of the history, current trends, and future directions. *The Educational and Developmental Psychologist*, *33*(1), 1–15.

Sonderen, E. v., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one*, *8*(7), e68967.

Steedle, J. T., Hong, M., & Cheng, Y. (2019). The effects of inattentive responding on construct validity evidence when measuring social–emotional learning competencies. *Educational Measurement: Issues and Practice*, *38*(2), 101–111.

Steinmann, I., Sánchez, D., van Laar, S., & Braeken, J. (2021a). The impact of inconsistent responders to mixed-worded scales on inferences in international large-scale assessments. *Assessment in Education: Principles, Policy & Practice*, 1–22.

Steinmann, I., Strietholt, R., & Braeken, J. (2021b). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological methods*.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed likert items. *Journal of Marketing Research*, *45*(1), 116–131.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series*, *2*(1), 9–36.

Weems, G. H., Onwuegbuzie, A. J., Schreiber, J. B., & Eggers, S. J. (2003). Characteristics of respondents who respond differently to positively and negatively worded items on rating scales. *Assessment & Evaluation in Higher Education*, *28*(6), 587–606.

Zhou, Y., & McBride, C. (2018). The same or different: An investigation of cognitive and metalinguistic correlates of chinese word reading for native and non-native chinese speaking children. *Bilingualism: Language and Cognition*, *21*(4), 765–781. https://doi.org/10.1017/S1366728915000279

Appendix A

GDPR Documentation & Ethical Approval

About us (/personvernombud/en/about_us.html)
Norwegian (/personvernombud/meld_prosjekt/meldeplikttest.html)

NSD (/) > Personverntjenester (/personvernombud/) > Data Protection Services (/personvernombud/en/) > Notify project (/personvernombud/en/notify/) > Notification Test

Denne siden på norsk (/personvernombud/meld_prosjekt/meldeplikttest.html)

Will you be processing personal data?

Are you unsure whether your project is subject to notification? Feel free to try our informal Notification test. Note that the test is intended as a guidance and is not a formal assessment.

Will you be collecting/processing directly identifiable personal data? ○ Yes ◉ No

A person will be directly identifiable through name, social security number, or other uniquely personal characteristics.

Read more about personal data (/personvernombud/en/help/vocabulary.html?id=8) and notification (/personvernombud/en/notify/index.html).

NB! Even though information is to be anonymized in the final thesis/report, check the box if identifying personal data is to be collected/processed in connection with the project.

Will directly identifiable personal information be linked to the data (e.g. through a reference number which refers to a separate list of names/scrambling key)? ○ Yes ◉ No

Note that the project will be subject to notification even if you cannot access the scrambling key (/personvernombud/en/help/vocabulary.html?id=11), as the procedure often is when using a data processor (/personvernombud/en/help/vocabulary.html?id=3), or in register-based studies (/personvernombud/en/help/research_methods/register_studies.html).

Will you be collecting/processing background information that may identify individuals (indirectly identifiable personal data)? ○ Yes ◉ No

A person will be indirectly identifiable if it is possible to identify him/her through a combination of background information (such as place of residence or workplace/school, combined with information such as age, gender, occupation, etc.).

Will there be registered personal data (directly/indirectly/via IP or email address, etc.) using online surveys? ○ Yes ◉ No

Please note that the project will be subject to notification even if you as a student/researcher cannot access the link to the IP or email address, as the procedure often is when using a data processor.

Read more about online surveys (/personvernombud/en/help/research_methods/online_surveys.html).

Will there be registered personal data using digital photo or video files?

○ Yes ● No

Photo/video recordings of faces will be regarded as identifiable personal data. In order for a voice to be considered as identifiable, it must be registered in combination with other background information, in such a way that people can be recognized.

Show results

Notify project

- Do I have to notify my project? (/personvernombud/en/notify/index.html)
- Notification Form (/personvernombud/en/notify/meldeskjema_link)
- Notifying changes (/personvernombud/en/notify/notifying_changes.html)

Get help notifying your project

- Processing the notification (/personvernombud/en/help/index.html)
- Frequently asked questions (/personvernombud/en/help/faq.html)
- Vocabulary (/personvernombud/en/help/vocabulary.html)
- Research topics (/personvernombud/en/help/research_topics/)
- Research methods (/personvernombud/en/help/research_methods/)
- Information and consent (/personvernombud/en/help/information_consent/)
- Other approvals (/personvernombud/en/help/other_approvals/)

© NSD - Norsk senter for forskningsdata • Kontakt NSD (/om/kontakt.html) • Personvern og informasjonskapsler (cookies) (/om/personvern.html)

Result of Notification Test: Not Subject to Notification

You have indicated that neither directly or indirectly identifiable personal data will be registered in the project.

If no personal data is to be registered, the project will not be subject to notification, and you will not have to submit a notification form.

Please note that this is a guidance based on information that you have given in the notification test and not a formal confirmation.

For your information: *In order for a project not to be subject to notification, we presuppose that all information processed using electronic equipment in the project remains anonymous.*

Anonymous information is defined as information that cannot identify individuals in the data set in any of the following ways:

> *- directly, through uniquely identifiable characteristic (such as name, social security number, email address, etc.)*
> *- indirectly, through a combination of background variables (such as residence/institution, gender, age, etc.)*
> *- through a list of names referring to an encryption formula or code, or*
> *- through recognizable faces on photographs or video recordings.*

Furthermore, we presuppose that names/consent forms are not linked to sensitive personal data.

Kind regards,
NSD Data Protection

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Appendix B

Data Management and Analysis Code

The R-script and RData files used for all the calculation and data analysis can be found in this link: https: //drive.google.com/drive/folders/14w-Jxxxkp21bUqMk0p57yMiUXFr-bSVa?usp=sharing. It contains one R-script (dataprep.R) that starts with downloading data from the OECD server and ends with a data frame that contains the calculated inconsistent responder (ICR) variable using the mean absolute difference method. The next three R scripts(onlyread.R, onlylang.R, and bothreadlang.R) starts with this data frame, followed by extraction relevant variables, writing a function that runs the education system wise logistic regression, and ends with generating plots of the regression coefficients. The next R-script (other plots) generates plots for descriptive statistics. The original data can be found in the following link to the OECD database: https://www.oecd.org/pisa/data/2018database/.