

Teenagers' ability to assess credibility in texts*

Modeling Canadian *Programme for International Student Assessment* 2018 data

Amy Farrow

10 January 2022

Abstract

This paper uses Programme for International Student Assessment (PISA) 2018 data to analyze Canadian students' ability to assess credibility of media texts. Predictors including demographic traits, subjects taught at school, reading frequency, and access to books and technology were used to create ordinal logistic and binary logistic models. English not being spoken at home, family wealth, and positive affect were associated with lower odds of scoring highly on the meta-cognitive assessing credibility scale. Higher parental occupation, perception of reading confidence, having books, reading long texts, and reading for enjoyment were associated with higher odds of scoring highly.

Contents

1	Part 1: Overview	2
1.1	Problem	2
1.2	Data	2
1.3	Results	6
2	Part 2: Detailed analysis	7
2.1	Models	7
2.2	Analysis and conclusions	7
A	Appendix A	10
B	Appendix B	10
C	Appendix C	10
	References	12

*Code and data are available at: github.com/amycfarrow/MAS.

1 Part 1: Overview

1.1 Problem

Today’s media landscape is characterized by large amounts of information from a wide range of sources and by increasing concern about misinformation (Cooke 2018). One solution to this information overload and misinformation is promoting information literacy, which encompasses a wide range of different literacies, including digital literacy, media literacy, and metaliteracy (Cooke 2018). With the increasing diversification of sources, one key information literacy skill is assessing credibility (Metzger 2007).

To successfully navigate the contemporary media landscape, today’s teenagers must become information literate at the same time as they learn more traditional literacies in school. Their ability to assess credibility of sources, amongst other skills, is a key tool they can use to protect themselves from misinformation. This paper explores the impact that demographics, education, and reading behaviors have on the ability to assess credibility for 15-year-olds in Canada.

1.2 Data

Every three years, the Organisation for Economic Co-operation and Development (OECD) tests teenagers globally to assess their demographic information as well as performance in reading, mathematics, and science (“PISA FAQ,” n.d.). The goal of the Programme for International Student Assessment (PISA), which has run for seven cycles, is to “assesses the extent to which 15-year-old students near the end of their compulsory education have acquired the knowledge and skills that are essential for full participation in modern societies” (OECD 2019). PISA tests students enrolled in educational institutions in grade 7 or higher, who are between the ages of 15 years 3 months and 16 years 2 months (“PISA FAQ,” n.d.). The 2018 PISA had a sample of 710,000 students complete the survey, representing 31 million students in 79 countries around the world (OECD 2019). Tests are administered on computers and take about two hours to complete (“PISA FAQ,” n.d.). PISA places great importance on breadth of coverage, representativeness of sample, and validity and reliability of measures (OECD 2019).

This paper uses the “Programme for International Student Assessment (PISA) 2018” (2020) data, specifically the subset of Canadian students. The dataset includes 19,422 observations, each representing an individual student who completed the survey.

For the first time, the 2018 PISA questionnaire included a scenario to assess student’s metacognitive reading processes related to assessing credibility (“PISA 2018 Technical Report” 2019). Students were given a reading task and a set of strategies and were asked to rate the usefulness of the strategies. The strategies were assessed by experts and a hierarchy was created. The students were given a score out of six based on how many times they chose a more useful over a less useful strategy (“PISA 2018 Technical Report” 2019). Figure 1 shows the distribution of scores received by Canadian PISA survey-takers. There is a polarization in the scores, with the majority of students either scoring 0 out of 6 or 6 out of 6.

Figure 2 shows the four predictor variables with a binary set of responses. They consist of student gender, language, and information literacy skills being taught at school.

Figure 3 shows the six predictor variables with an ordered set of categorical responses. They include the frequency of school reading activities, the longest texts read in school, reading for pleasure, and books in the home.

Figure 4 shows the six predictor variables with numerical values. They include index values to measure the parents’ occupational status, the student’s access to information and communication technology resources, and the parents’ education in years of schooling; a composite scale that measures the student’s positive affect; a composite scale that measures the student’s perception of their own reading competence; and a measure of family wealth that is assessed based on possessions in the family home.

More descriptive statistics, including missing values, are available in Appendix A.

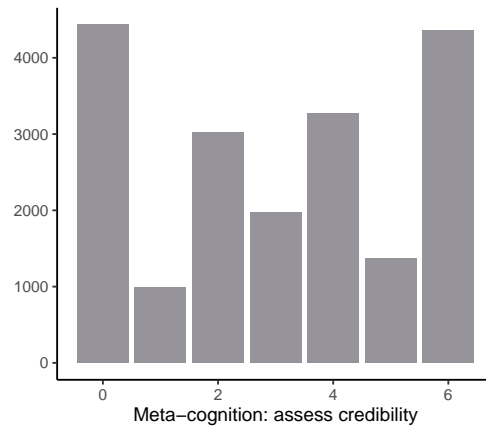


Figure 1: Response variable distribution

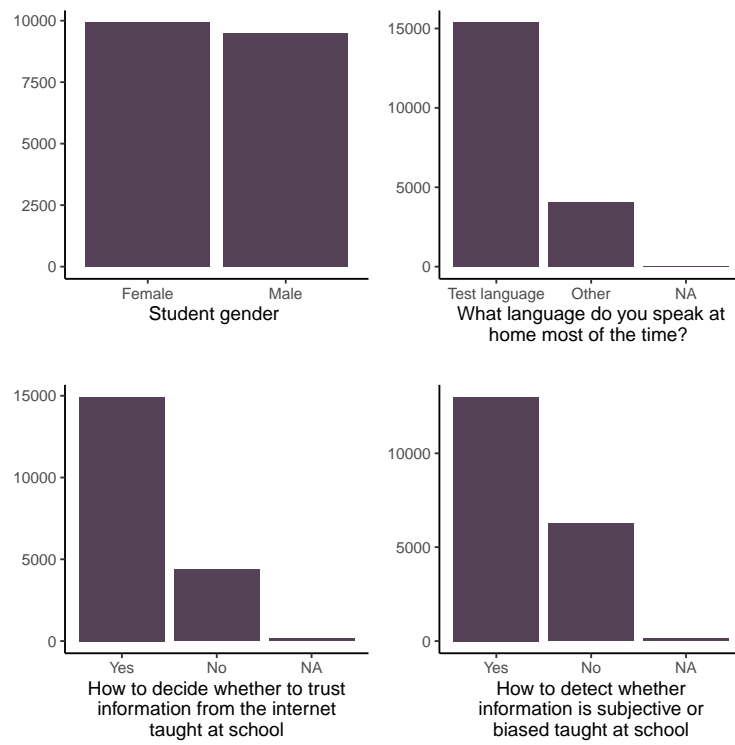


Figure 2: Binary predictor variable distributions

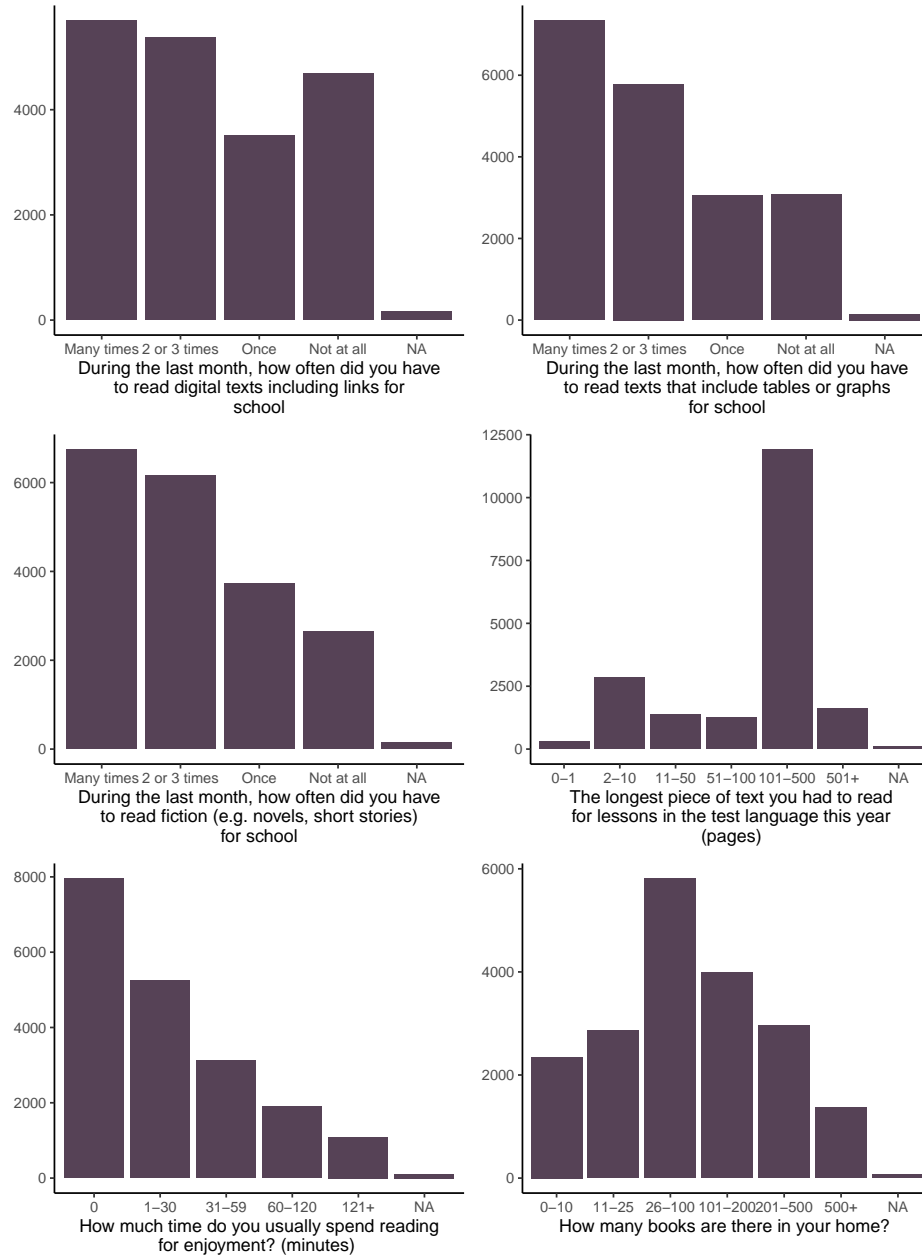


Figure 3: Ordinal predictor variable distributions

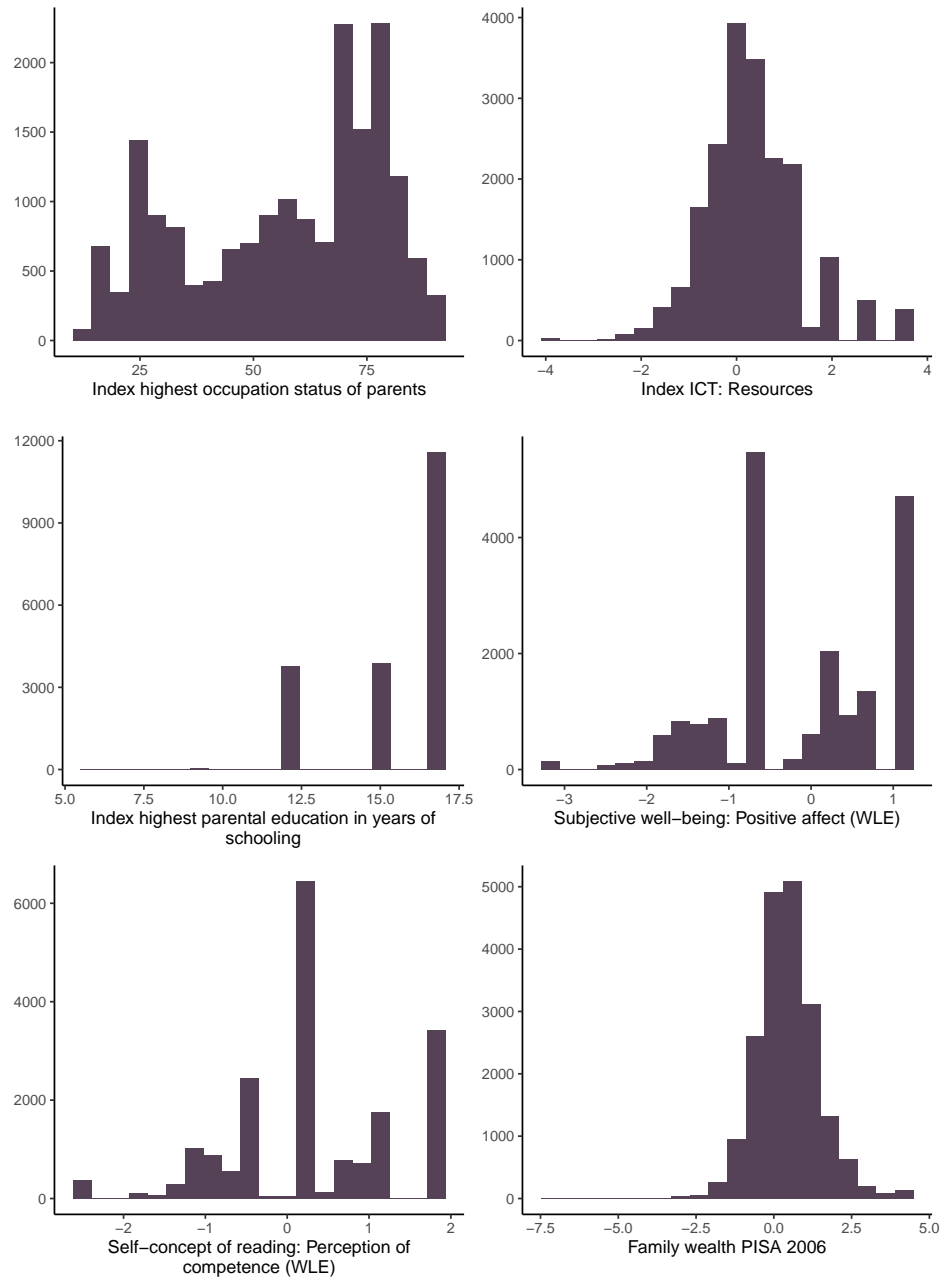


Figure 4: Interval and ratio predictor distributions

1.3 Results

Three models were used to analyse the relationship between the predictors and assessing credibility scores. The first two both used a model designed for ordered categories, like the assessing credibility score, but were created using different subsets of the data. The third model considered assessing credibility score as a dichotomous outcome: either low or high, splitting the scale in half. All predictors described above were considered for all models.

By looking at the outcome of all models, the predictors that are significant regardless of the model type can be identified. Each significant predictor is associated with an increase or decrease in the odds of a student scoring higher on the assessing credibility scale. Talking about each individual predictor, we are assuming that all other predictors are being kept the same when we talk about an increase or decrease in odds.

Speaking a language other than the test language at home is associated with a 32% decrease. Considering the credibility strategies are assessed in the test language, this result makes sense. Reading texts with graphs once over the past month, when compared with reading them many times, is associated with an 18% decrease. Similarly, never reading texts with graphs is associated with a 19% decrease. We can expect that students who are not asked to read texts with graphs at school on a regular basis will have lower odds of a high assessing credibility score.

Some decreasing results are more surprising. Family wealth increasing by one unit (where each unit is equivalent to one standard deviation in the data distribution) is associated with an 11% decrease. Positive affect increasing by one unit (where each unit is equivalent to one standard deviation in the data distribution) is also associated with an 11% decrease.

Other variables are associated with increases in odds. Parental occupation index increasing by one unit (on a scale of 0 to 100) is associated with a 1% increase.

Many reading-related predictors are associated with increases. A student's self-perception of reading competence increasing one unit (where each unit is equivalent to one standard deviation in the data distribution) is associated with a 23% increase. Having 101-500 books in the home, compared to having 0-10 books, is associated with a 69% increase. Compared to a longest text read for school of 0-1 pages, the longest text being 101-500 pages is associated with an increase of 153%. The text being 501+ pages is associated with a 117% increase. Compared to never reading for fun, reading for 1-30 minutes, 31-59 minutes, 60-120 minutes, and 121+ minutes are associated with increases of 42%, 51%, 49%, and 40%, respectively.

If we want to investigate ways to improve teenagers' ability to assess the credibility of sources, measures associated with traditional literacy should be considered: reading texts with graphs at school, reading longer texts for school, and reading for fun are all related to the assessing credibility score. Family circumstances, while less manipulable, are also worth consideration if only to understand what may give some students an advantage over others when developing critical information literacy: language spoken at home, family wealth, parental occupation, and books in the home are all significant.

It is also important to note that a student's assessing credibility score cannot be predicted with a high degree of accuracy based on the predictors described above. These models are helpful for identifying significant factors, but not for determining an individual student's outcome.

2 Part 2: Detailed analysis

2.1 Models

Prior to analysis, an 80/20 train/test split was done.

Because the response variable, assessing credibility score, is an ordered categorical variable, an ordinal logistic model was selected. The ordinal logistic model can be represented as:

$$\text{logit}(P(Y \leq j)) = \alpha_j - \beta X$$

where j is a specific outcome on the ordinal scale, X is the set of predictors chosen for the model, α_j is the intercept specific to the level in question, and β is the model's coefficients.

Backward elimination was used to refine the model. The model with all possible predictors was reviewed for the least significant predictor, and the models with and without that predictor were compared using an ANOVA. If the null hypothesis $H_0 : \beta = 0$ was not able to be rejected, then the predictor was removed. This process was repeated until the null hypothesis for tested predictors could be rejected. The model with all predictors and the final model are presented in Table 1.

The model diagnostics (shown in Appendix B) demonstrated that the ordinal logistic model was predicting exclusively values of 0 and 6. This was true of all iterations of the ordinal logistic model. This is possibly due to the imbalanced nature of the dataset. The most frequent values of the assessing credibility score are 0 and 6. Two approaches were taken to address this imbalance.

First, the training set was randomly undersampled to create a balanced dataset, where each value of the assessing credibility score was represented at an equal frequency. This reduced the number of observations the model was trained on from 13,736 to 4,940.

The same procedure with ordinal logistic models, p-values, and ANOVA was used to refine a model on this balanced dataset. The model with all predictors and the final model are presented in Table 1.

The model diagnostics (shown in Appendix C) show that the model trained on the balanced dataset predicted a full range of assessing credibility score values, unlike the model trained on the imbalanced dataset.

The second approach to the polarized assessing credibility score was converting the score to a binary outcome. Scores below 4 were reassigned to 0, and all other scores were reassigned to 1. With this binary outcome, logistic regression was applicable. The logistic model can be represented as:

$$\log\left(\frac{p}{1-p}\right) = \beta X$$

where p is the probability of the binary outcome being a score of 4 or above, X is the set of predictors chosen for the model, and β is the model's coefficients.

The same procedure with p-values and ANOVA was used to refine a logistic model. The model with all predictors and the final model are presented in Table 1.

Both the ordinal logistic model trained on the complete training data and the ordinal logistic model trained on the balanced subsample were used to predict values for the test data, and the RMSE for both was calculated. The model trained on the full training set had an RMSE of 3.386. Despite being trained on fewer values, the model trained on the balanced data performed better with an RMSE of 2.989. While this is an improvement, both RMSE values indicate a poor predictive performance when considering that the range of response is only 6.

The logistic model was assessed using Area Under the Curve (AUC) on the test set, which was 0.668. This is not a strong predictive performance.

2.2 Analysis and conclusions

While the models are not predictively strong, they are useful for identifying variables that significantly relate to a student's ability to assess credibility in a text.

Table 1: Model comparison

	Ordinal logistic.	Ordinal logistic final	Ordinal logistic, balanced	Ordinal logistic final, balanced	Logistic	Logistic final
gender2	−0.146***	−0.145***	−0.060		−0.140***	−0.135***
language2	−0.332***	−0.330***	−0.396***	−0.387***	−0.327***	−0.325***
parent_occup	0.006***	0.006***	0.007***	0.007***	0.007***	0.007***
parent_ed	0.000		−0.010		0.001	
family_wealth	−0.203***	−0.204***	−0.161***	−0.118***	−0.192***	−0.192***
num_books2	0.172**	0.171**	−0.012	−0.005	0.158*	0.158*
num_books3	0.262***	0.262***	0.133	0.140	0.263***	0.264***
num_books4	0.424***	0.425***	0.232*	0.239*	0.448***	0.448***
num_books5	0.610***	0.610***	0.511***	0.525***	0.595***	0.595***
num_books6	0.410***	0.411***	0.227+	0.241+	0.443***	0.443***
ict_resources	0.104***	0.106***	0.055		0.086**	0.087**
trust_internet2	−0.062		−0.074		−0.075	−0.100*
detect_bias2	−0.061+	−0.078*	−0.077		−0.066	
longest_text2	0.498**	0.503***	0.643**	0.642**	0.212	0.209
longest_text3	0.517**	0.522***	0.689**	0.666**	0.254	0.252
longest_text4	0.502**	0.505**	0.548*	0.549*	0.251	0.250
longest_text5	0.899***	0.905***	0.940***	0.929***	0.604***	0.605***
longest_text6	0.636***	0.642***	0.788**	0.777**	0.375*	0.380*
read_fun2	0.350***	0.349***	0.344***	0.351***	0.301***	0.301***
read_fun3	0.441***	0.442***	0.397***	0.413***	0.420***	0.422***
read_fun4	0.457***	0.456***	0.380***	0.402***	0.436***	0.437***
read_fun5	0.299***	0.299***	0.311**	0.338**	0.336***	0.338***
school_read_fic2	−0.065+	−0.072+	−0.067		−0.077+	−0.085+
school_read_fic3	−0.050	−0.054	0.029		−0.083	−0.089+
school_read_fic4	0.164**	0.163**	0.032		0.152**	0.155**
school_read_graphs2	−0.161***	−0.173***	−0.062	−0.078	−0.140**	−0.155***
school_read_graphs3	−0.170***	−0.178***	−0.193*	−0.198*	−0.225***	−0.238***
school_read_graphs4	−0.227***	−0.227***	−0.206*	−0.214**	−0.211***	−0.208***
school_read_digital2	−0.058		−0.118+	−0.128+	−0.068	
school_read_digital3	−0.054		−0.059	−0.069	−0.067	
school_read_digital4	−0.004		0.092	0.074	0.008	
positive_affect	−0.119***	−0.118***	−0.112***	−0.111***	−0.134***	−0.133***
read_competence	0.198***	0.199***	0.203***	0.206***	0.177***	0.179***
Num.Obs.	13 736	13 736	4940	4940	13 736	13 736
AIC	48 964.6	48 960.1	18 796.9	18 790.7	18 001.8	17 997.9
BIC	49 258.1	49 216.0	19 050.6	18 992.4	18 257.7	18 216.2
Log.Lik.	−24 443.282	−24 446.038	−9359.466	−9364.372	−8966.892	−8969.941
edf	39.000	34.000	39.000	31.000		

Some predictors have fairly consistent coefficients across all models. This analysis refers to the numbers in the balanced ordinal logistic model due to its performance on the test set, but only coefficients that are significant and similarly valued across all models are discussed.

First, speaking a language other than the test language at home has a coefficient of -0.387. To interpret the coefficient, we exponentiate it to get the proportional odds ratio. In this case, the proportional odds ratio is approximately 0.68, indicating that for students who speak a language other than the test language in the home, the odds of scoring higher is 0.68 times that of students who do speak the same language at home, holding all other variables constant, a decrease of 32%.

Family wealth has an odds ratio of 0.89, indicating that for every one unit increase in the normalized family wealth scale (where 1 = 1 standard deviation), the odds of scoring higher is multiplied by 0.89, a decrease of 11%.

Positive affect has an odds ratio of 0.89, indicating that for every one unit increase in the normalized positive affect scale (where 1 = 1 standard deviation), the odds of scoring higher is multiplied by 0.89, a decrease of 11%.

Reading texts including graphs or tables once over the last month has an odds ratio of 0.82, indicating that for students who read these once, the odds of scoring higher is 0.82 times that of students who read them many times, a decrease of 18%. Similarly, never reading texts has a decrease of 19%.

Parental occupation has an odds ratio of 1.01, indicating that for every one unit increase in the parental occupation scale (0-100), the odds of scoring higher is 1.01 times higher.

Perception of reading competence has an odds ratio of 1.23, indicating that for every one unit increase in the normalized perception scale (where 1 = 1 standard deviation), the odds of scoring higher is multiplied by 1.23, an increase of 23%.

101-500 books in the home has an odds ratio of 1.69, indicating that for students who have 101-500 books, the odds of scoring higher is 1.69 times that of students who have 0-10 books, an increase of 69%.

Longest text read in the last month being 101-500 pages long has an odds ratio of 2.53, indicating that for students who have read a 101-500 page text, the odds of scoring higher is 2.53 times that of students who only read a text of 0-1 pages, an increase of 153%. Similarly, longest text read in the last month being 501+ pages long has an odds ratio of 2.17, indicating that for students who have read a 500+ page text, the odds of scoring higher is 2.17 times that of students who only read a text of 0-1 pages, an increase of 117%.

Reading 1-30 minutes for fun has an odds ratio of 1.42, indicating that for students who have read 1-30 minutes, the odds of score higher is 1.42 times that of students who never read for pleasure, and increase of 42%. Reading for 31-59 minutes, 60-120 minutes, and 121+ minutes have similar odds ratios, marking increases of 51%, 49%, and 40%, respectively.

Some predictors (gender, access to ICT resources, reading fiction in school) were only significant in the polarized ordinal logistic and logistic models, indicating that they may be useful for distinguishing between the middle of the scale but not for sorting students in high- and low-performing groups. Others, like parental education and being taught how to decide to trust internet information and how to detect bias, did not appear to be significant factors in relation to a student's assessing credibility score.

This analysis indicates that predictors related to reading more (at school or for pleasure) are positively associated with higher assessing credibility scores. Demographic predictors are more mixed in outcome, with parental occupation increasing the proportional odds and family wealth decreasing the proportional odds.

These results are specific to Canadian 15-year-olds and may not transfer to other contexts, including other countries or other ages of students. Further, the students are tested in a very restricted, artificial environment, and their performance on PISA may not be comparable to their ability (and inclination) to assess the credibility of sources that they encounter in the real world. Assessing practical information literacy would require different measures. Finally, a test for assessing credibility that resulted in a more normally-distributed, granular score would be beneficial to future modeling.

A Appendix A

Table 2: Predictor variable summary table

assess_credibility	Total N	Missing N		value
gender	19422 (100.0)	0	1	3.2 (2.2)
			2	2.9 (2.2)
language	19407 (99.9)	15	1	3.1 (2.2)
			2	2.7 (2.2)
trust_internet	19274 (99.2)	148	1	3.1 (2.2)
			2	2.9 (2.2)
detect_bias	19268 (99.2)	154	1	3.1 (2.2)
			2	2.9 (2.2)
school_read_digital	19263 (99.2)	159	1	3.2 (2.2)
			2	3.0 (2.2)
			3	2.9 (2.2)
school_read_graphs	19273 (99.2)	149	4	3.0 (2.2)
			1	3.3 (2.2)
			2	2.9 (2.2)
			3	2.8 (2.2)
school_read_fic	19277 (99.3)	145	4	2.9 (2.2)
			1	3.2 (2.2)
			2	2.9 (2.2)
			3	2.8 (2.2)
longest_text	19331 (99.5)	91	4	3.1 (2.2)
			1	1.7 (1.9)
			2	2.6 (2.2)
			3	2.6 (2.1)
			4	2.6 (2.2)
			5	3.3 (2.2)
read_fun	19315 (99.4)	107	6	3.1 (2.2)
			1	2.5 (2.2)
			2	3.2 (2.2)
			3	3.5 (2.1)
			4	3.6 (2.1)
			5	3.5 (2.1)
num_books	19356 (99.7)	66	1	2.2 (2.1)
			2	2.7 (2.2)
			3	3.0 (2.2)
			4	3.3 (2.2)
			5	3.6 (2.2)
			6	3.4 (2.2)
parent_occup	18171 (93.6)	1251	[11.0,89.0]	3.0 (2.2)
ict_resources	19398 (99.9)	24	[-3.8,3.6]	3.0 (2.2)
parent_ed	19307 (99.4)	115	[6.0,17.0]	3.0 (2.2)
positive_affect	19018 (97.9)	404	[-3.1,1.2]	3.0 (2.2)
read_competence	19153 (98.6)	269	[-2.4,1.9]	3.0 (2.2)
family_wealth	19411 (99.9)	11	[-6.9,4.5]	3.0 (2.2)

B Appendix B

Figure 5 shows diagnostic plots for the first ordinal logistic model.

C Appendix C

Figure 6 shows diagnostic plots for the second ordinal logistic model, trained on a balanced undersampled subset of the training data.

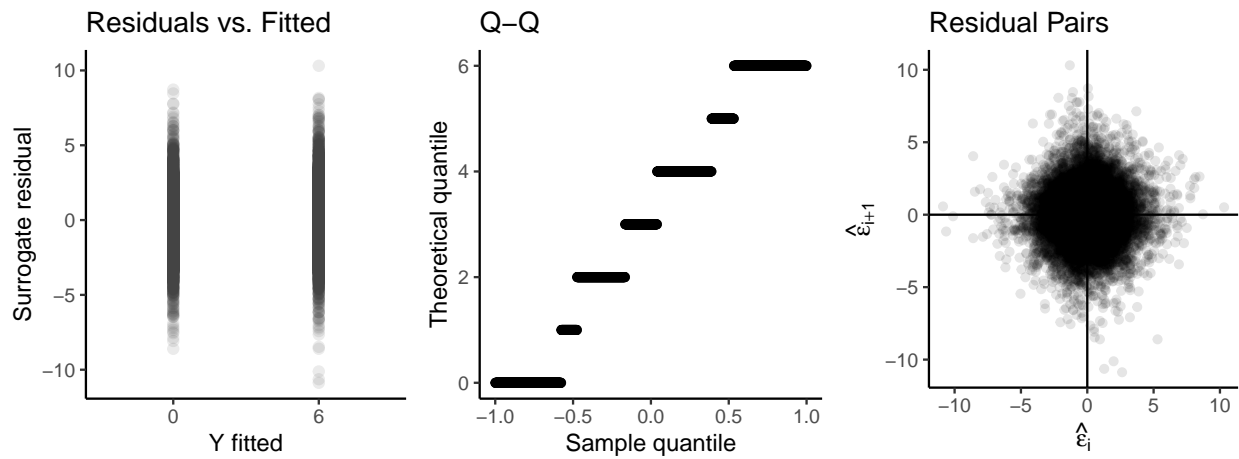


Figure 5: Diagnostic plots for first ordinal logistic model

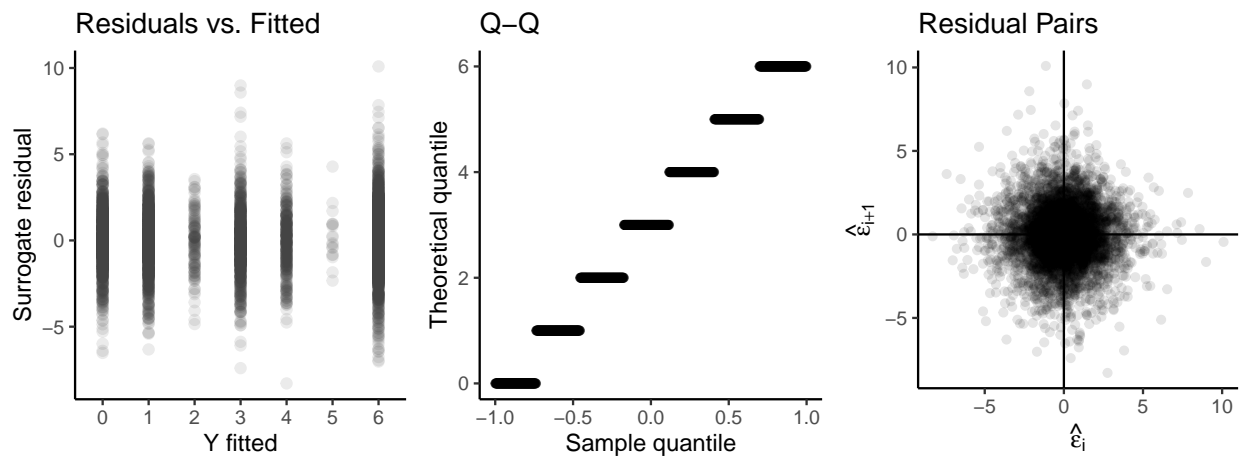


Figure 6: Diagnostic plots for second ordinal logistic model

References

- Cooke, Nicole A. 2018. *Fake News and Alternative Facts: Information Literacy in a Post-Truth Era*. ALA Editions Special Reports. ALA Editions.
- Metzger, Miriam J. 2007. “Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research.” *Journal of the American Society for Information Science and Technology* 51, 58 (13, 14): 2078–91. <https://doi.org/10.1002/asi.20672>.
- OECD. 2019. *PISA 2018 Assessment and Analytical Framework*. Organisation for Economic Co-operation; Development. https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_b25efab8-en.
- “PISA 2018 Technical Report.” 2019. <https://www.oecd.org/pisa/data/pisa2018technicalreport/>.
- “PISA FAQ.” n.d. *Organisation for Economic Co-Operation and Development*. <https://www.oecd.org/pisa/pisafaq/>.
- “Programme for International Student Assessment (PISA) 2018.” 2020. Organisation for Economic Co-operation and Development. <https://www.oecd.org/pisa/data/2018database/>.