

The Missing Response Patterns in the Ontario Secondary School Literacy Test

Jingshun Zhang
Ruth A. Childs

Abstract

Large-scale assessments are often an important indicator of students' achievement for schools, states, and provinces. Missing responses can affect the appropriateness of our analysis models and the results of large-scale educational assessments. The study of missing response patterns (MRPs) can inform the design of a test and interpretation of test results. This study will examine the causes and effects of MRPs based on analyses of students' responses to the Ontario Secondary School Literacy Test (OSSLT) in 2006. This is a test with high stakes for students. With some preliminary statistical analyses in SPSS (descriptive statistics, plots, cross-tabs, and multinomial and logistic regressions), we are exploring possible causes of MRPs by examining the relationships between patterns of missing responses and responses to test items and background questionnaire items. All results will be helpful for us to understand more about the test's construct and internal validity to support improvement of the relevant large-scale assessment in the future.

Introduction

Large-scale assessments are often an important sign for schools, states, and provinces of students' achievement (Taylor & Tubianosa, 2001) and are valued by the general public for the accuracy and objectivity with which they can measure the effectiveness of students, schools, and teachers (Crundwell, 2005, p. 5). As Wolfe, Childs, and Elgie (2004) wrote, "When we talk about reporting assessment results, we assume that the students have responded to the test items. However, many students omit or provide unmarkable responses to one or more items" (p. 62). Missing Response Patterns (MRPs) in this research is a broad concept involving missing data, incomplete data, missing responses, omitted responses, neglected responses, nonresponse, and missingness. MRPs are important to consider in statistical analyses in educational measurement, social science, and medical studies (Toledano & Gatsonis, 1999), as they can significantly affect the results of the data analysis in large-scale assessment. The study of MRPs can inform our understanding of the design, validity, and uses of large-scale assessment.

Missing responses can affect the appropriateness of our analysis models and the results of large-scale educational assessments. For example, many analyses make strong assumptions about the causes of the missing data, such as assuming that a missing response is equivalent to an incorrect response, which may be

particularly problematic when subgroups of examinees (e.g., those grouped by language or gender) differ in their patterns of nonresponse (Emenogu, Falenchuk, & Childs, 2010). Also, much valuable information can be found from examining patterns of missing data. The study of MRPs can inform the design of a test and the interpretation of test results. This study will examine the causes and effects of MRPs based on analyses of students' responses to the 2006 Ontario Secondary School Literacy Test (OSSLT). The OSSLT is "a cross-curricular literacy test issued to all secondary school students in the province of Ontario" and "consists of a reading and a writing component, both of which must be successfully completed for secondary school graduation in Ontario" (Cheng, Fox, & Zheng, 2007, p. 67). In an information session before the test, teachers were required to read the following statement to students: "Answer all the test questions. Not answering questions or leaving a writing task blank will reduce your chances of success" (EQAO, 2006, p. 33). During the test administration, the students were reminded, "Answer all the questions in each section" (p. 35). Although the OSSLT is a high-stakes test required for students' graduation from high school, based on our study, there are nevertheless a few missing response patterns.

We have two research questions in this project: (1) How do the MRPs change across the test items and structure? (2) How are these changes related to students' characteristics? With some preliminary statistical analyses in SPSS (descriptive statistics, plots, and cross-tabulations or cross-tabs), we explore possible causes of MRPs by examining the relationships between patterns of missing responses and responses to test items and background questionnaire items. All results will be helpful for us to understand more about the test's construct and internal validity to support improvement of the relevant large-scale assessment in the future.

Literature Review

In 1987, Little and Rubin stated that the typical classifications of MRPs are *missing at random*, *missing completely at random*, and *not missing at random*. Although many analyses assume that data are either missing at random or missing completely at random, in fact, it is likely that most examinee nonresponse on large-scale assessments is not missing at random. For example, Ludlow and O'Leary (1999) suggested that different MRPs may indicate different test-taking strategies. These strategies may affect the number of items any individual examinee attempts. Also, examinees' ability estimates will be affected by their test-taking strategies and so will be less comparable across groups if the strategies differ across groups.

Some researchers have studied the effects of the treatments of MRPs (e.g., deleting cases that are incomplete, scoring missing responses as wrong, ignoring missing responses) on assessment results. For example, Emenogu, Falenchuk, and Childs (2010) studied the effect of the treatment of MRPs on Mantel-Haenszel DIF detection. Many researchers have used simulations to study the effects of MRPs in the models. Some researchers have studied the effect of MRPs on statistical and theoretical models such as regression and factor analysis. Kamakura and Wedel (2000) studied the estimation of factor models and the imputation of MRPs, and

proposed an approach that provided direct estimates of factor weights without the replacement of MRPs with imputed values.

Some studies have investigated the possible causes of MRPs, including the relationship between MRPs and examinees' characteristics or attitudes. For example, Grandy (1987) examined group differences in MRPs by gender. Zhu and Thompson (1995) also examined differences by racial/ethnic groups and the relationship between MRPs and performance on the items. Recently, Chuah and Linden (2008) studied examinees' aberrant responses by combining response-pattern and response-time data. Wise and DeMars (2008) also showed a relationship between examinee noneffort and the validity of program assessment results. On the other hand, MRPs may be related to test items (Draney & Wilson, 2004, p. 1). For example, Xu (2005) suggested that nonresponse rates are related to item format, gender, language, and culture. Choppin (1974) studied student response patterns on an international survey of academic standards and found that students from England and the US were more likely to guess than those from European countries such as France and Sweden. Zhu and Thompson (1995) found that White students were most likely to respond to all items on a multiple-choice assessment, followed by Asian students and Hispanic students; African-American students had the highest nonresponse rates. In TIMSS 1995, the students from Denmark and some East Asian countries tended to leave an item blank if they did not know how to answer it, while students from the US had a relatively larger proportion of random guesses (Xie, 2005). Students who omit more items on the test may not have less knowledge or fewer skills, but they may be more reluctant to try to answer when they are not certain of the answer. Different test-taking strategies can occur in a systematic manner due to differential cultural origin and instructional emphasis or some other unknown factors. Wise and DeMars (2008) also found a relationship between examinee noneffort and the validity of program assessment results. A demands-capacity model of test-taking effort proposed by Wise and Smith (2011) considered a test as a series of examinee-item encounters. Also, some researchers have considered MRPs, correct responses, and wrong responses together. For example, Schmidt, Wolfe, & Kifer (1993) used a triangle graphic to show the relationship among MRPs, correct responses, and wrong responses (p. 93).

In addition, some researchers are interested in the relationship between literacy and MRPs. For example, Brown and his colleagues (1996) studied the relationship between literacy performance and MPRs. They found that the lower the levels of adults' literacy skills, the more MRPs they have. In this paper we will apply relevant theories from previous studies to real data from the Ontario provincial large-scale assessment, the Ontario Secondary School Literacy Test, administered in 2006.

Research Methods

Combining broad theories in educational measurement, educational statistics, and pedagogy (curriculum, teaching, and learning), this project employed different methods for its different subquestions.

Data Sources

Davey and Savla (2010) wrote, “Data collection from human beings in the real world poses considerably greater challenges than in the laboratory setting” (p. 47). Nevertheless, this quantitative study mainly employed real data. This study analyzed the data from the 2006 OSSLT, developed by Ontario’s Education Quality and Accountability Office (EQAO). About a third of the population of Canada lives in Ontario and Ontario’s testing program “illustrate[s] the divergent uses of large-scale assessments in Canada throughout the twentieth century” (Klinger, DeLuca, & Miller, 2008, p. 3).

The 2006 OSSLT included an English version (190,758 test-takers) and a French version (6,539 test-takers). After data cleaning, we finalized the data for this study as 190,480 students, retaining only the students who answered the English-language version of the test, those who were taking the test when they were first eligible (students take the test for the first time when they are in Grade 10 and can retake it in subsequent years), and those who did not receive accommodations.

The 2006 OSSLT had 62 items (questions) and was structured as shown in Table 1. We selected different types of items for various studies.

Types of Items	Booklet 1	Booklet 2
Multiple Choice (MC) Items	23	25
Open Response (OR) Items	2	4
Short Writing (SW) Items	4	0
Long Writing (LW) Items	2	2
Total	31	31

Table 1. The Structure of the OSSLT

Our study focused on the 48 multiple-choice (MC) items. We created several indicators of missing response patterns for use in the analyses, such as number of missing responses in open-response (OR) items, number of missing responses in short writing (SW) items and long writing (LW) items, number of missing responses in multiple-choice items, number of missing responses in all types of items, number of items not reached at the end in Booklet 1, number of items not reached at the end in Booklet 2, number of items omitted in Booklet 1, and number of items omitted in Booklet 2. In addition, we computed the number of items of each type answered correctly.

Methods for Data Analysis

We analyzed the percentages and empirical distributions of MRPs, and drew some statistical graphs such as bar graphs, histograms, line graphs, and scatter plots. We analyzed the relationships between MRPs and students’ characteristics, demographic

information and their relevant survey responses. We also looked at the observed factors in one-way analysis of variance (*ANOVA*). We undertook statistical analysis of the whole database using the statistical software package SPSS 17.

We employed some preliminary statistical analyses that are related to MRPs, such as descriptive statistics, plots, and cross-tabs. We explored possible causes of the MRPs by examining the relationships between them and responses to test items and background questionnaire items. In the study of MRPs, there were two foci: the item level, and the student level.

In summary, this study probed the factors possibly affecting the MRPs. After descriptive analysis of MRPs and individual student characteristics, we looked for the following two types of MRPs, which are related to our two research questions:

- MRPs by item across students—we studied them for each individual item across all students; and
- MRPs by student across items—we studied them for each student across all items

Results

We will introduce our data analysis results in two parts, which are related to the MRPs by item and by student, as explained above. Based on our descriptive analysis, we obtained the following results: For all six open-response items, 87.1% of the students answered all the items; 7.7% did not answer just one item; 1.9% did not answer two items; and 1.5% did not answer three or more. For the four writing items, 4.0% of the students did not answer one of the items and 1.0% did not answer two or more. For the multiple-choice items, 95.7% of the students answered all the items and less than 0.5% did not answer more than five MC items.

Study 1: The Investigation of MRPs by Item across Students

According to the literature review and our previous studies, MRPs may be related to a test's items and its structure. Therefore, corresponding to our first research question, we undertook a series of analyses to investigate this.

1. MRPs differ by item

Figure 1 shows the percentage of students who did not answer each MC item. The items are ordered as they appeared on the test; those in Booklet 1 begin with B1 and those in Booklet 2 begin with B2. We found that the percentage of students who did not answer an item ranged from 0.2 to 1.2%. However, this trend was not very stable. At the beginning, the percentages were higher and then went down. They went up suddenly at the beginning of the second section in Booklet 1. The switching point between sections at the beginning of Booklet 2 brought the second significant rise in the percentage of students not answering each item. Also, we found that the trend in Booklet 1 was much less stable than that in Booklet 2.

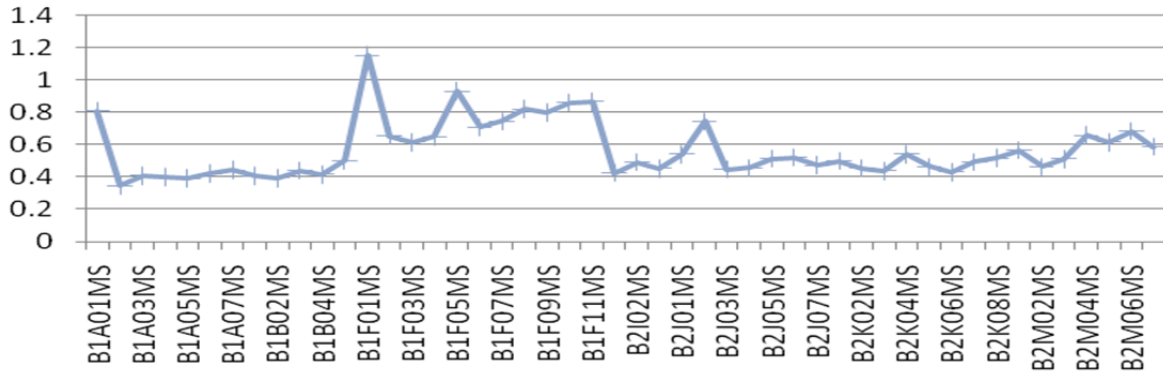


Figure 1. The percentage of students not responding to each item

2. MRPs – The gap between correct responses and incorrect responses

When we study the responses to items and tests with a holistic approach to large-scale assessment, we need to consider the relationships among correct responses, incorrect responses, and missing responses together.

For each item, we know that the sum of the percentages of correct responses, incorrect responses, and missing responses should be 100%. Graphically, if we plot the percentage correct for each item as the distance from the bottom and the percentage incorrect as the distance from the top, then the gap between these lines is the percentage missing. Appendix I provides such a graph, but the gaps are small and so are difficult to see.

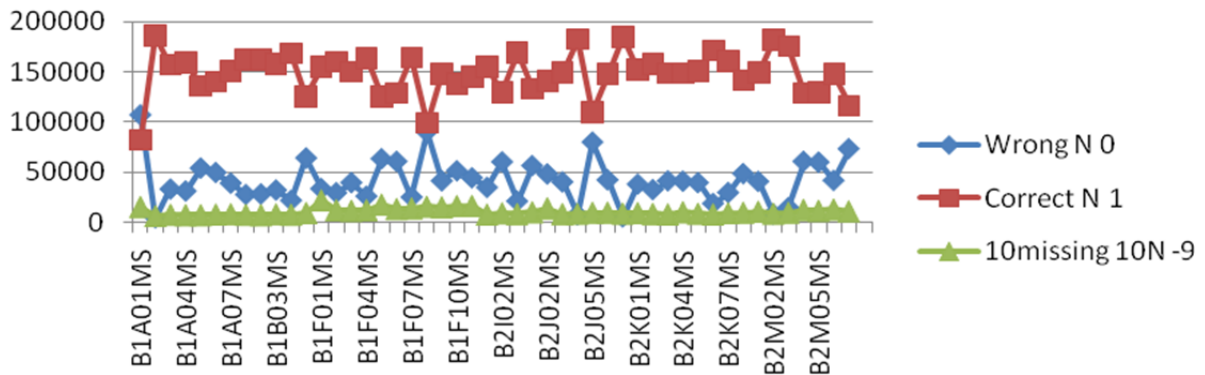


Figure 2. For each item, the percentage of students answering correctly, the percentage answering incorrectly, and 10 times the percentage with missing responses

In Figure 2, both the percentage of students answering correctly and the percentage of students answering incorrectly are plotted vertically from the bottom of the graph. To make the pattern of missing responses clearer, the missing responses were amplified by multiplying the percentage by 10. A correlation analysis confirms the patterns seen in this figure. The correlation between correct responses and incorrect responses is $r = -1.000$, $p < .01$. The percentage of students missing responses to each item is also significantly correlated with both the percentage of students answering correctly ($r = -.419$, $p < .01$) and the percentage of students answering incorrectly ($r = .407$, $p < .01$).

3. Item difficulty and MRPs

According to the results above, the percentage of students answering an item correctly is negatively correlated to the percentage of students not answering the item. Here, we wanted to analyze further to see how the percentage of students not answering an item was related to the item's difficulty. Item difficulty is simply the percentage of students taking the test who answered the item correctly. The larger the percentage of respondents getting an item right, the easier the item. The higher the difficulty index, the easier the item is understood to be (Crocker & Algina, 1986; Matlock-Hetzel, 1997; Wood, 1960). We computed the item difficulty by dividing the number of people answering the item correctly by the total number of responses to the item, ignoring the missing responses. Based on our data, we obtained Figure 3.

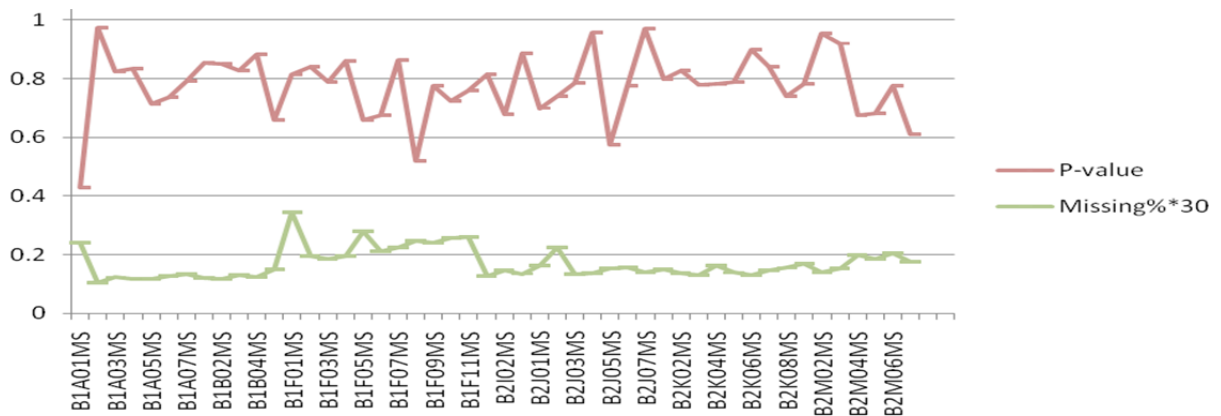


Figure 3. For each item, the item difficulty (percentage of students answering correctly out of students who responded to the item) and 30 times the percentage with missing responses.

4. The skill requirements of items and MRPs

The 2006 OSSLT was created to have different skill requirements for each question. Reading was defined as the process through which the reader actively makes meaning for a variety of written texts. Students were expected to understand the texts used in the OSSLT according to the expectations in the Ontario Curriculum across all subjects up to the end of Grade 9. Writing was defined as the constructive process of communicating in the written forms expected of students in the Ontario Curriculum across all subjects up to the end of Grade 9. Writing skills were evaluated on the 2006 OSSLT through a combination of multiple-choice questions and short and long writing tasks (EQAO, 2006). Table 2 summarizes the three reading and four writing skills.

Targeted Skill	Code	Skill Category
Reading Skill 1	R1	understanding explicitly stated information and ideas
Reading Skill 2	R2	understanding implicitly stated information and ideas (making inferences)
Reading Skill 3	R3	making connections between information and ideas in a reading selection and personal knowledge and experience (interpreting reading selections by integrating information and ideas in a reading selection with personal knowledge and experience)
Writing Skill 1	W1	developing a main idea with sufficient supporting details
Writing Skill 2	W2	organizing information and ideas in a coherent manner
Writing Skill 3	W3	using conventions (spelling, grammar, punctuation) in a manner that does not distract from clear communication
Writing Skill 4	W4	topic development (main idea, supporting details and organization)

Table 2. Targeted Skills. Note. Adapted from EQAO (2005), p. 4; W4 is not measured by the MC items.

We analyzed whether the MRPs showed significant differences among these skills. Appendix II summarizes the data for analysis, from which we drew Figure 4.

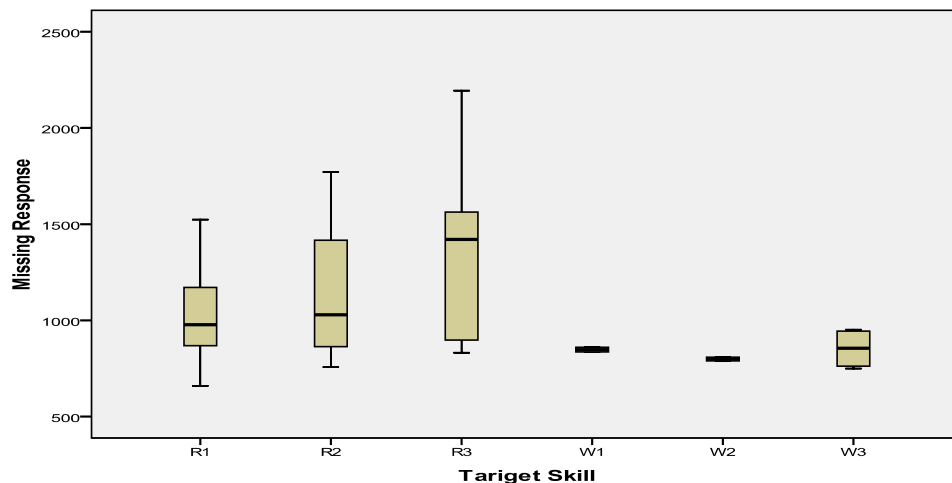


Figure 4. Percentage of missing responses by six targeted literacy skills in reading and writing

An ANOVA of the targeted literacy skills classified as reading and writing found a significant difference between the missing response rates for reading ($M = 0.294\%$, $SD = 0.028\%$) and writing ($M = 0.123\%$, $SD = 0.059\%$) skills, $F(1,42) = 5.096$, $p = .029$. The interaction between skill and type of skill (reading or writing) was not significant, $F(4,42) = 1.669$, $p = .175$. The variance was much larger for the three reading skills than for the writing skills. The writing skills were measured by far fewer items than the reading skills, however.

5. “Not reached” items and MRPs

We also studied another interesting trend of MRPs: not reached by the end. Shin (2009) studied the impact of different scoring methods on IRT-based true score equating. The study recommended that omitted and not-reached items should be ignored or left blank, and not scored as incorrect. The study also found that the benefits of treating omitted and not-reached items as blank or missing increased as the sample size increased. Going over the whole test, we found that some students stopped answering questions from a certain point forward. Along with their growing numbers of MRPs, examinees skip more items by the end of the two booklets. Based on this list, we have many questions to answer. For example, there were some items and sequences of items that showed a high frequency of omits. Why? Was there evidence of correlation of omitting for nonadjacent items? Carefully studying this list helped us find many interesting results. Looking back at these patterns, we found most of them belonged to “not reached by the end” patterns. Those examinees with more MRPs skip more items by the end of the two booklets combined.

Number of Items “Not Reached” at the End of the Booklet	Booklet 1		Booklet 2	
	Number of Students	%	Number of Students	%
0	186,449	97.88377%	178,255	93.58200%
1	2,331	1.22375%	8,230	1.53664%
2	205	0.10762%	2927	1.53664%
3	115	0.06037%	124	0.06510%
4	74	0.03885%	44	0.02310%
5	61	0.03202%	26	0.01365%
6	62	0.03255%	20	0.01050%
7	58	0.03045%	79	0.04147%
8	42	0.02205%	19	0.00997%
9	24	0.01260%	756	0.39689%
10	34	0.01785%		0.00000%
11	1,025	0.53811%		0.00000%
Total	190,480	100.00000%	190,480	100.00000%

Table 3. “Not Reached” Patterns

We can see these trends in Figure 5: Booklet 2 had significantly more students who did not reach items than Booklet 1.

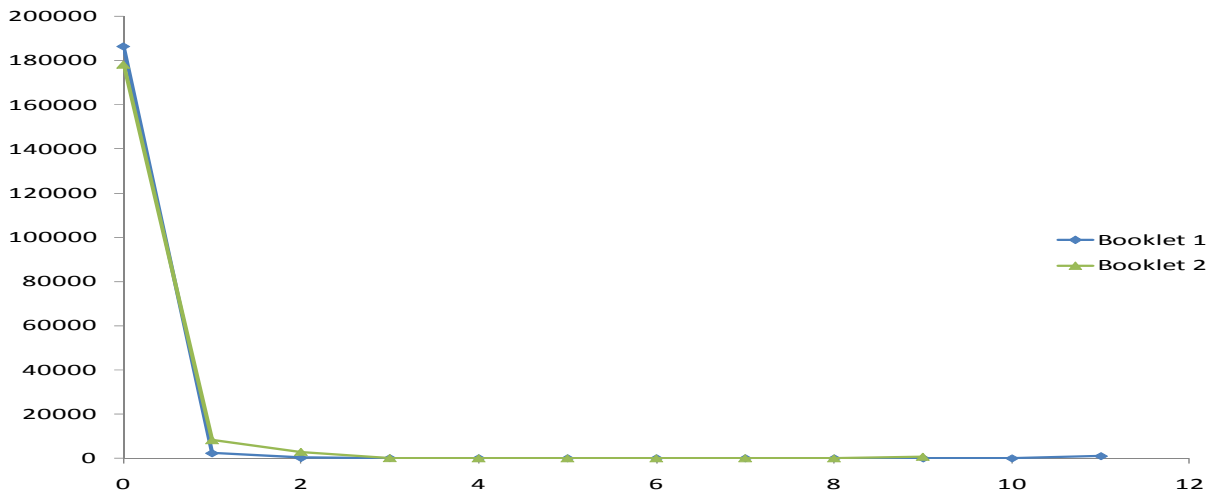


Figure 5. Numbers of students not reaching items in the two booklets

We also suspected that the MRPs had some relationship with an item's wording and length. We will continue to study these areas in the future.

Study 2 – The investigation of MRPs by student across items

The second part of this study focused on the individual students' MRPs and probed broadly the relationships between MRPs and other factors.

1. Distribution of MRPs by student

For all students, we computed their total number of missing responses for all items; the distribution is shown in Figure 6. Overall, 86.2% of students did not have any missing items, 95.5% of the examinees missed less than 3 items, and only 1% of examinees had more than 10 items missing. Also, we found that 427 examinees (.02%) missed all 62 items and they gave us only their names.

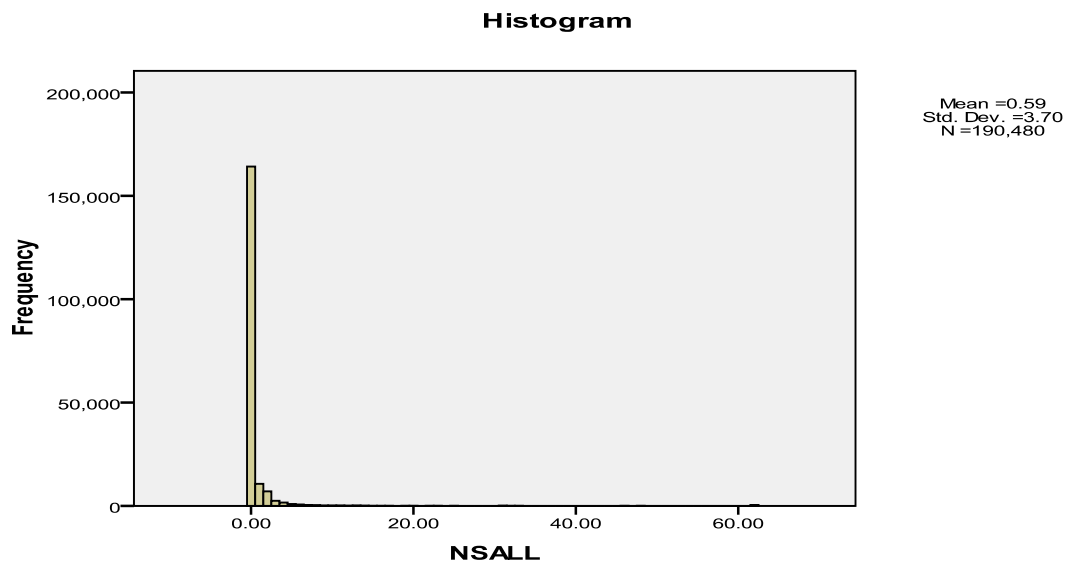


Figure 6. Distribution of students' total number of missing responses

A student's total score as reported by the EQAO and his or her total number of missing responses had a significant negative correlation, $r = -.316$, $p < .001$. This means that the more items a student missed, the lower his or her total score.

2. MRPs and skills

In order to investigate the relationships between each student's total number of missing responses and the targeted skills R1, R2, R3, W1, W2, and W3, we created six variables that indicated the total number of items measuring each skill that were answered correctly. All skills were significantly negatively correlated with the total number of missing responses. Students' total score in Reading Skills 1 and Reading Skills 2 had the largest correlations ($r = -.229$, $p < .001$ and $r = -.248$, $p < .001$) and Writing Skill 3 had the smallest correlation ($r = -.118$, $p < .001$) with students' total number of missing responses.

3. MRPs and individual characteristics (group differences)

Based on previous studies, we know that MRPs may be related to students' background information. Many researchers have also found that MRPs have significant differences among different groups. The 2006 OSSLT obtained relevant background information from students using a survey. We undertook broad identification of these variables. Here, we show some of them.

a) Gender and MRPs

We investigated the relationship between MRPs and student background information. The number of MC items not answered was not significantly different between male and female students. However, the number of items of any type not reached at the ends of the booklets was significantly different by gender. Males ($M =$

0.69, $SD = 3.91$) had significantly higher numbers of missing items than females ($M = 0.48$, $SD = 3.45$), $t(190,342) = 12.14$, $p < .001$, $d = 0.06$.

b) Language background

Ontario has many students who have immigrated there with their parents. Therefore, these students have various original language backgrounds. We were interested in whether this language background and MRPs had a significant correlation. Other studies have found that the performance of ESL/ELD students is consistently and similarly lower across item formats, reading text types, skills and strategies, and the four writing tasks (Cheng, Klinger, & Zheng, 2007; Fox & Cheng, 2007). Therefore, this study examined the MRPs for students who were classified by their schools as English as a Second Language (ESL) learners or English Literacy Development (ELD) learners, that is, students who were receiving instruction designed to help them improve their skills in reading, writing, and oral communication in English (Ontario Ministry of Education, 2002, p. 2).

Students who were classified as ESL/ELD had significantly more missing items ($N = 8,482$, $M = 1.16$, $SD = 4.49$) than other students ($N = 181,998$, $M = 0.56$, $SD = 3.66$)—in fact, about twice as many: $t(9,013) = 12.17$, $p < .001$, $d = 0.15$. Furthermore, the variance in ESL students' MRPs was much larger than for other students.

Family language background (responses to the question “Languages you speak at home”) also had significant differences for MRPs, $t(72,246) = 26.67$, $p < .001$, $d = 0.15$. Examinees who spoke only English at home ($M = 0.31$, $SD = 1.53$) had less missing data than others ($M = 0.59$, $SD = 2.13$).

c) IEPs and MRPs

Students who had an Individual Education Plan (IEP) had special needs (e.g., behavior difficulties, autism, deaf or hard of hearing, blind or low vision, learning disabilities); some of these students also required accommodations (e.g. time, seating, and so on). Students with IEPs ($M = 0.90$, $SD = 4.56$) had significantly higher MRPs than others ($M = 0.53$, $SD = 3.50$), $t(39664) = 13.89$, $p < .001$, $d = 0.09$.

d) Home computer use and MRPs

Finally, students who used a computer every day at home ($N = 60,152$, $M = 0.27$, $SD = 1.33$) were compared with students who did not ($N = 127,159$, $M = 0.44$, $SD = 1.87$). The students who did not use computers at home daily had significantly more missing data, $t(159,527) = 21.89$, $p < .001$, $d = 0.10$.

4. Rates of Nonresponse and Their Relationship to Personal Characteristics

To investigate the relationship between the number of items to which a student did not respond and the student's characteristics, including gender and whether the student was taking applied or academic programs and others, while controlling for the students' ability estimate based on the 48 MC items, we performed a sequence of multinomial and logistic regressions. Table 4 provides the results of an analysis in

which the outcome variable represents five levels of missing responses: 0, 1, 2, 3, or 4 or more.

Parameter	<i>B</i>	<i>SE</i>	Wald Chi-Square
Thresholds			
Missing = 0	3.364	.0626	2891.809***
Missing = 1	4.674	.0651	5161.948***
Missing = 2	5.187	.0673	5949.595***
Missing = 3	5.479	.0690	6297.108***
Theta (θ)	-.429	.0151	805.018***
Program: Applied	.216	.0310	48.516***
Program: Academic	-.112	.0312	12.875***
Program: Unknown	0	.	.
IEP or Disability: No	-.164	.0666	6.042*
IEP or Disability: Yes	0	.	.
Accommodations: No	.267	.0699	14.597***
Accommodations: Yes	0	.	.
Only English at Home: No	.448	.0259	299.222***
Only English at Home: Yes	0 ^a	.	.
English Language Learner: No	.030	.0464	.419
English Language Learner: Yes	0	.	.
Gender: Female	-.122	.0231	27.890***
Gender: Male	0	.	.
(Scale)	1		

Table 4. Multinomial regression of number of missing responses (categorized) on students' characteristics. Note. *df* for all chi-square tests is 1. * $p < .05$; ** $p < .01$; * $p < .001$.**

We applied a 3-PL IRT model in the relevant analysis. The ability estimate, θ , was included in the model as a covariate because analyses reported earlier indicated that many variables in the model had strong relations with θ . The *B* parameter of -.429 confirmed that θ was negatively related to the number of missing responses.

Controlling for θ , a student's academic program, whether he or she had an IEP (except for gifted programs) or disability, whether he or she received accommodations on the test, and gender were significantly correlated with the categorized number of missing responses. In particular, students taking courses in the applied program had more missing responses than those with an unknown program, while students in the academic program had fewer missing responses. Students without an IEP or identified disability had fewer missing responses. Students without accommodations on the test had more missing responses. Students who spoke a language other than or in addition to English at home had

more missing responses. Females had fewer missing responses. Identification of a student by their school as an English Language Learner when the OSSLT was administered was not a significant predictor of number of missing responses.

Conclusion

We believe a holistic approach to assessment should include studies of correct responses, incorrect responses, and nonresponse altogether. Given that MRPs are underaddressed in research, we hope our focus on them can help fill this gap.

We believe that many benefits can be expected from this study. We investigated the MRPs in the OSSLT, a test with high stakes for students. We broadly investigated the MRPs and MC items in the OSSLT and found some interesting results in areas such as MSPs, IRT, item difficulty level, item orders, and test structure. These results will help test-makers improve this test. For each item on the OSSLT, the rate of MRPs was lower than 1.2%. According to traditional treatment, the MRPs could be considered as missing at random. However, based on our results, MRPs have a very strong relationship with different groups such as gender, IEP, whether students were taking the test for the first time, computer use at home, and other variables. Therefore, even though the rates of MRPs were very low, student characteristics still had some effects on them. All these results will be helpful for us to understand more about the test's construct and internal validity. Furthermore, the results of this study (for instance, the causes of unanswered questions in the test) can also have meaningful pedagogical implications for the teaching and learning process. For example, gender differences are related to some differences in MRPs. Therefore, in classroom learning, a teacher might often remind boys to stay on task. Also, we found that the first items in each test's section have higher MRPs. We need to take care of students' psychological performance when they start any test, project, or assignment, for instance by reminding them to calm down and focus on their work as soon as possible at the beginning of their task.

Based on our current analyses of MRPs, we will continue our study and try to get more useful results to support improvement of the relevant large-scale assessments. This study focused on the analysis of missing response patterns in the OSSLT 2006. However, this is not the final goal for our studies. It is necessary to apply these results to the development of large-scale assessments, the improvement of teaching and learning, and the informing of policy. These are future directions for study.

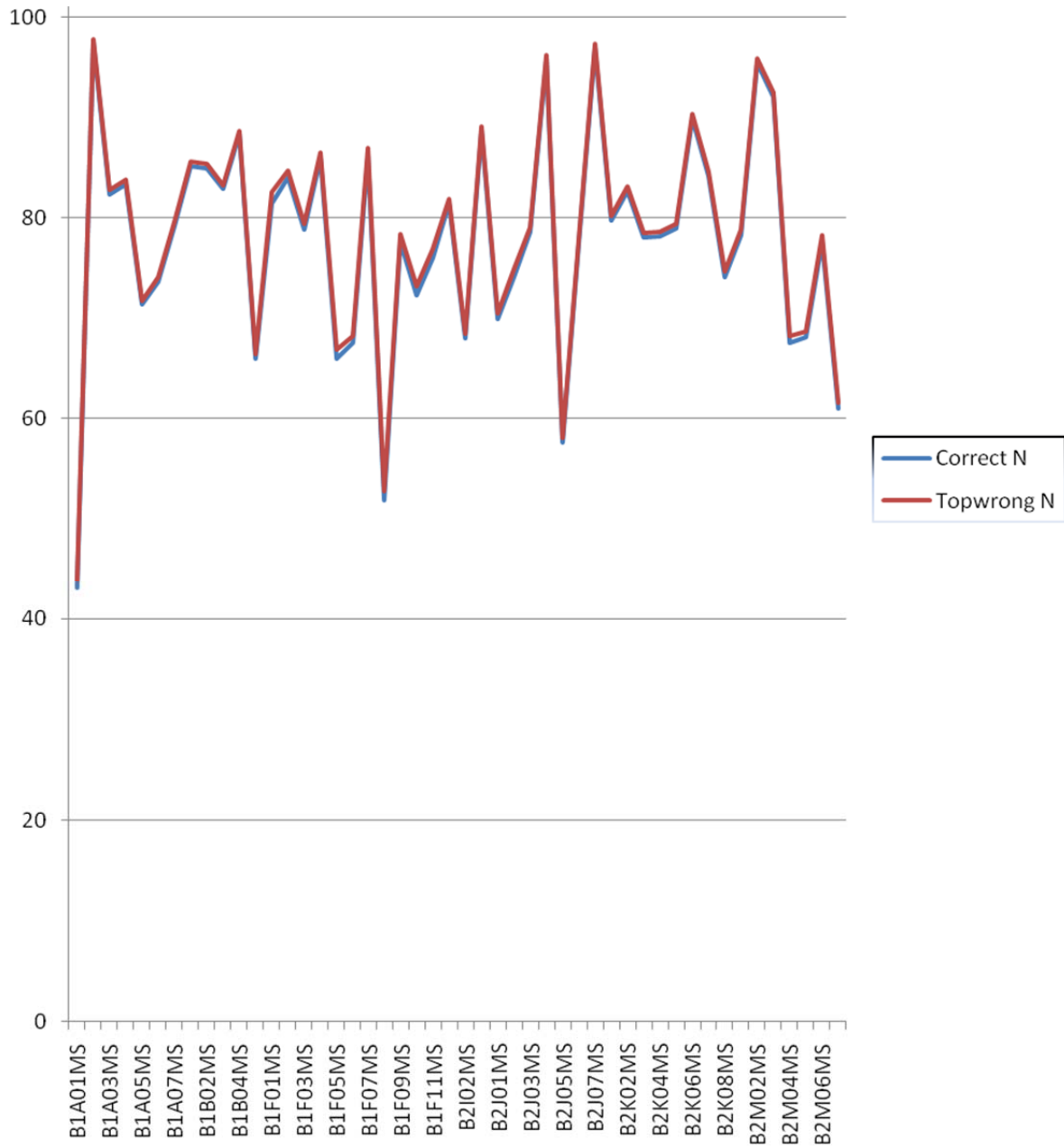
References

- Brown, H., Prisuta, R., Jacobs, B., & Campbell, A. US Department of Education. *National Center for Education Statistics. Literacy of Older Adults in America: Results from the National Adult Literacy Survey*, NCES 97-576. Washington, DC: 1996.
- Cheng, L., Fox, J., & Zheng, Y. (2007). Student accounts of the Ontario secondary school literacy test: A case for validation. *Canadian Modern Language Review*, 64(1), 69-98.
- Cheng, L., Klinger, D., & Zheng, Y. (2007). The challenges of the Ontario Secondary School Literacy Test for second language students. *Language Testing*, 24, 185-208.
- Choppin, B. H. (1974). *The correction for guessing on objective tests* (IEA Monograph Studies, No. 4). Stockholm: The International Association for the Evaluation of Educational Achievement.
- Chuah, S. C., & Linden, W. J. (2008, March). *Detection of aberrant candidate response: Improving detection by combining response pattern and response time data*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Crundwell, R. M. (2005). Alternative strategies for large-scale student assessment in Canada: Is value-added assessment one possible answer? *Canadian Journal of Educational Administration and Policy*, 41, 1-21.
- Davey, A., & Savla, J. (2010). *Statistical power analysis with missing data: A structural equation modeling approach*. New York: Routledge/Taylor & Francis.
- Draney, K., & Wilson, M. (2004). Missing data: Issues and solutions (paper commissioned by the EQAO Assessment Review Project). In R. Wolfe, R. Childs, & S. Elgie (Eds.), *Final Report of the External Evaluation of EQAO's Assessment Processes: External evaluation of EQAO's assessment processes, Volume 3* (pp. 1-5). Toronto: Ontario Institute for Studies in Education, University of Toronto.
- EQAO (Education Quality and Accountability Office). (2005). *Ontario Secondary School Literacy Test Framework*. Toronto: Queen's Printer. Retrieved May 1, 2011 from <http://www.ontla.on.ca/library/repository/mon/10000/250328.pdf>.
- EQAO (Education Quality and Accountability Office). (2006). *The complete guide for administering the Ontario Secondary School Literacy Test*. Toronto: Queen's Printer. Retrieved May 1, 2011 from <https://ozone.scholarsportal.info/bitstream/1873/2233/1/259658.pdf>.

- Emenogu, B. C., Falenchuk, O., & Childs, R. A. (2010). The effect of missing data on Mantel-Haenszel DIF detection. *Alberta Journal of Educational Research*, 56, 458-468.
- Fox, J., & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. *Assessment in Education*, 14, 9-26.
- Grandy, J. (1987). *Characteristics of examinees who leave questions unanswered on the GRE General Test under rights-only scoring* (GRE Board Professional Rep. No. 83-1P). Princeton, NJ: Educational Testing Service.
- Kamakura, W. A., & Wedel, M. (2000). Factor analysis and missing data. *Journal of Marketing Research*, 37, 490-498.
- Klinger, D. A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in Canadian education. *Canadian Journal of Educational Administration and Policy*, 76. Retrieved Nov. 16, 2009 from <http://www.umanitoba.ca/publications/cjeap/articles/klinger.html>
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.
- Ludlow, L. H., & O'Leary, M. (1999). Scoring omitted and not-reached items: Practical data analysis implications. *Educational and Psychological Measurement*, 59, 615-630.
- Matlock-Hetzel, S. (1997, January). *Basic concepts in item and test analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. Retrieved Nov. 1, 2010, from <http://ericae.net/ft/tamu/Espy.htm>.
- Ontario Ministry of Education. (2002). *Ontario curriculum unit planner: ESL/ELD companion*. Toronto: Queen's Printer for Ontario.
- Schmidt, W. H., Wolfe, R. G., & Kifer, E. (1993). The identification and description of student growth in mathematics achievement. In L. Burstein (Ed.), *The IEA study of mathematics III: Student growth and classroom processes* (pp. 59-99). New York: Pergamon.
- Shin, Seon-Hi. (2009). How to treat omitted responses in Rasch model-based equating. *Practical Assessment, Research & Evaluation*, 14(1). Retrieved Nov. 1, 2010 from <http://pareonline.net/pdf/v14n1.pdf>.
- Taylor, A. R., & Tubianosa, T. (2001). *Student assessment in Canada: Improving the learning environment through effective evaluation*. Kelowna, BC: Society for the Advancement of Excellence in Education.
- Toledano, A. Y., & Gatsonis, C. (1999). Generalized estimating equations for ordinal categorical data: Arbitrary patterns of missing responses and missingness in a key covariate. *Biometrics*, 55, 488-496.

- Wise, S. L., & DeMars, C. E. (2008, March). *Examinee non-effort and the validity of program assessment results*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
- Wise, S. L., & Smith, L. F. (2011). A model of examinee test-taking effort. In J. A. Bovaird, K. F. Geisinger, & C. W. Buckendahl (Eds.), *High stakes testing in education: Science and practice in K-12 settings* (pp. 139-154). Washington, DC: American Psychological Association.
- Wolfe, R., Childs, R., & Elgie, S. (2004). *Ensuring quality assessments: A project to refine and affirm assessment processes. Final report of the external evaluation of EQAO's assessment process*. Retrieved July 2, 2009 from <https://ozone.scholarsportal.info/bitstream/1873/3608/1/244807.pdf>.
- Wood, D. A. (1960). *Test construction: Development and interpretation of achievement tests*. Columbus, OH: Charles E. Merrill.
- Xie, Y. (2005). *Three studies of person by item interactions in international assessments of educational achievement*. Unpublished doctoral dissertation, University of California at Berkeley, CA.
- Xu, Y. (2005). *Examining group differences in omission rates in a mathematics assessment*. Unpublished master's thesis, University of Toronto, Ontario, Canada.
- Zhu, D., & Thompson, T. D. (1995, April). *Gender and ethnic differences in tendencies to omit responses on multiple-choice tests using number-right scoring*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Appendix I



Appendix 2

Item	Correct	Incorrect	Missing	Skill
B1A01MS	106982	81960	1538	R3
B1A02MS	4118	185702	660	R1
B1A03MS	32793	156909	778	R1
B1A04MS	30874	158849	757	R2
B1A05MS	53771	135964	745	R1
B1A06MS	49375	140303	802	R2
B1A07MS	38774	150861	845	R2
B1B01MS	27401	162304	775	W3
B1B02MS	27925	161805	750	W3
B1B03MS	31842	157802	836	W1
B1B04MS	21550	168140	790	W2
B1B05MS	63963	125565	952	W3
B1F01MS	33243	155043	2194	R3
B1F02MS	29130	160105	1245	R1
B1F03MS	39215	150095	1170	R2
B1F04MS	25647	163598	1235	R1
B1F05MS	63117	125592	1771	R2
B1F06MS	60463	128669	1348	R1
B1F07MS	24838	164225	1417	R2
B1F08MS	90017	98900	1563	R3
B1F09MS	41146	147810	1524	R1
B1F10MS	51096	137744	1640	R2
B1F11MS	44066	144765	1649	R2
B2I01MS	34508	155163	809	W2
B2I02MS	60036	129508	936	W3
B2I03MS	20770	168850	860	W1
B2J01MS	56209	133242	1029	R2
B2J02MS	47915	141148	1417	R2
B2J03MS	39951	149677	852	R1
B2J04MS	7217	182394	869	R1
B2J05MS	79792	109710	978	R1
B2J06MS	42002	147485	993	R2

B2J07MS	4970	184612	898	R3
B2K01MS	37682	151848	950	R2
B2K02MS	32251	157365	864	R2
B2K03MS	41034	148614	832	R3
B2K04MS	40643	148807	1030	R2
B2K05MS	39173	150426	881	R1
B2K06MS	18368	171291	821	R2
B2K07MS	29334	160208	938	R1
B2K08MS	48259	141237	984	R1
B2M01MS	40416	148987	1077	R1
B2M02MS	7824	181773	883	R2
B2M03MS	14280	175226	974	R1
B2M04MS	60544	128683	1253	R2
B2M05MS	59629	129680	1171	R1
B2M06MS	41429	147748	1303	R3
B2M07MS	73148	116214	1118	R1