



Determining Seat Time for a Shortened Section of a High-Stakes Standardized Test

Nicolas May, Patrick McLain, & Stephen O'Connell, Michigan Language Assessment

Introduction

Following feedback from Michigan English Test (MET) stakeholders that the number of items and duration of the test were onerous, a shortened version of the test was developed that maintained the high reliability and content validity of the MET. However, following these revisions to the number of items on the exam, the test developers faced the challenge of determining an appropriate new seat time for the reading and grammar section that would avoid introducing any undue speededness.

Study Design and Results

Using item response time data from a previous research project, two seat times were chosen to be trialed: 60 minutes and 75 minutes.

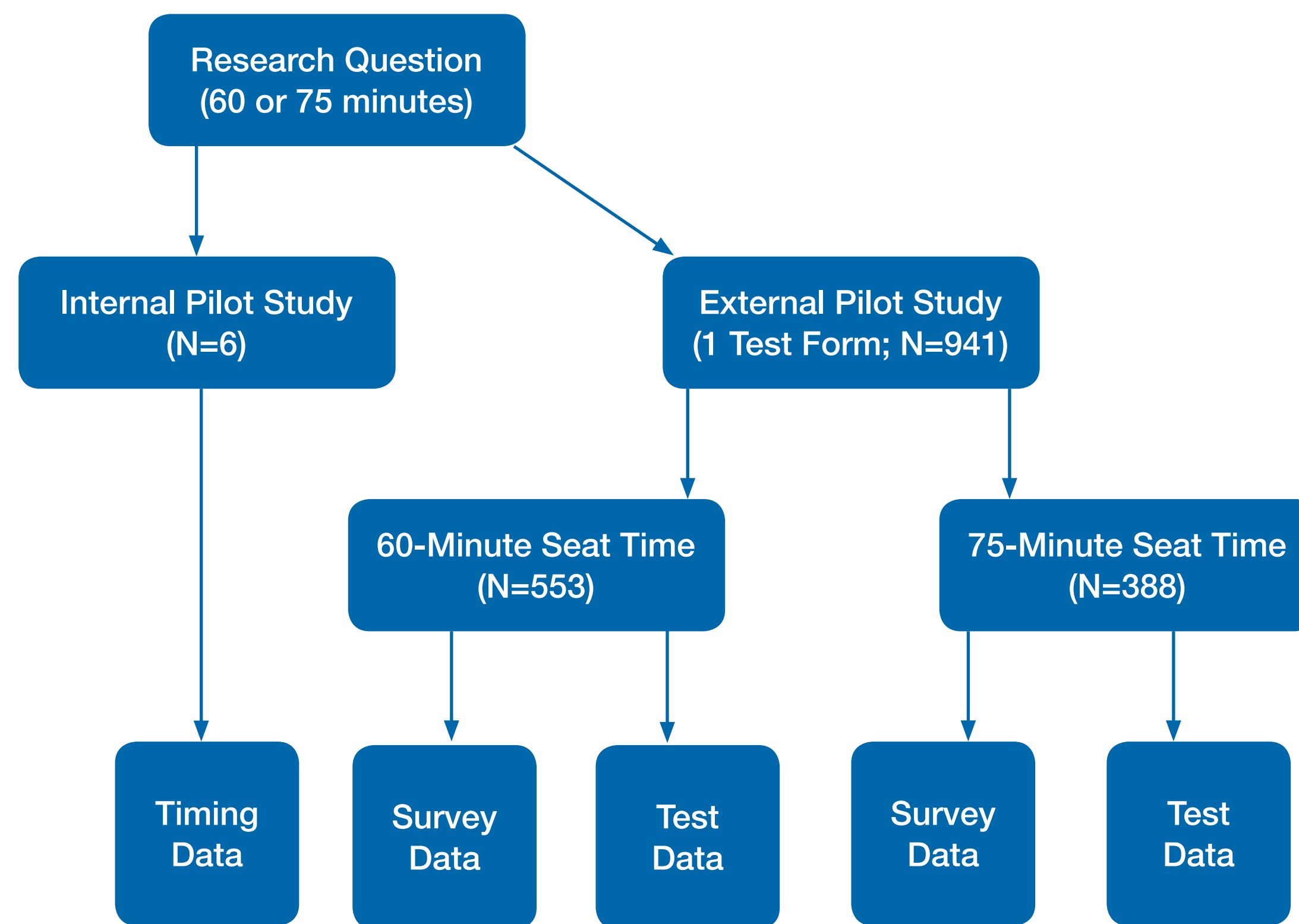


Figure 1: Study Design

Several sources of evidence were used to determine which seat time was most appropriate. The sources of evidence were:

- Preliminary Seat Time Estimate (Internal Pilot Data)
- Test Taker and Proctor Survey Results (External Pilot Survey Data)
- Test Summary Statistics (External Pilot Test Data)
- Lu & Sireci Speededness Indices (External Pilot Test Data)
- Differential Item Functioning Analysis (External Pilot Test Data)

Internal Pilot Data

Prior to piloting the exam on the test takers, a small internal pilot of the revised exam was run with 6 CaMLA employees (3 native speakers and 3 highly proficient non-native speakers). The purpose of this pilot was to use the information on the amount of time the internal pilot participants needed to complete the 50 and 75 item exams and the current seat time for the 75 item exam to estimate an appropriate seat time for the 50 item exam. This was done using the following ratio:

$$\frac{\text{Internal Pilot Average Time (75 items)}}{\text{Current Seat Time (75 Items)}} \times \frac{\text{Internal Pilot Average Time (50 items)}}{\text{Estimated Seat Time (50 minutes)}}$$

This resulted in a seat time estimate of 71.25 minutes for the revised reading and grammar section. (! Small sample size).

External Pilot Survey Data

Seat Time	N	Too Much	OK	Too Little
60-Minutes	550	3.27	71.27	25.45
75-Minutes	385	5.71	87.79	6.49

Pearson's Chi-Square Test (Independence): $\chi^2=57.21$, $df=2$, **p-value<0.001**

Seat Time	N	Too Much	OK	Too Little
60-Minutes	33	0.00	78.79	21.21
75-Minutes	16	6.25	81.25	12.50

Fisher's Exact Test (Independence): **p-value=0.3887**

Seat time difference had a statistically significant impact on the test takers' perception of having "too much", "OK", or "too little" time to complete the exam. By contrast, seat time difference did not have a statistically significant impact on the proctors' perception of the test takers having "too much", "OK", or "too little" time to complete the exam.

External Pilot Test Data

Seat Time	N	Mean	SD	Minimum	First Quartile	Median	Third Quartile	Maximum
60-Minutes	553	31.43	11.01	0	22	32	41	50
75-Minutes	388	33.55	10.44	10	26	34	43	50

Welch Two Sample T-Test: $T=-2.9975$, $df=859.54$, **p-value=0.0028**

Lu & Sireci Indices

Three indices described in Lu & Sireci (2007) were used to determine if either the 60-minute or 75-minute seat time resulted in any undue test speededness for the shortened exam.

Indices	Formula	Description	Interpretation
Power Ratio	$\frac{S_U}{S_X}$	Ratio of the standard deviations of the number of items not reached (U) to the total number of items not given a correct answer (X).	Values less than 0.25 are indicative of an unspeeeded test
Speededness Ratio	$\frac{S_W}{S_X}$	Ratio of the standard deviations of the number of items incorrectly answered or omitted (W) to the total number of items not given a correct answer (X).	Values less than 0.1 are indicative of a speeeded test.
Speededness Quotient	$SQ = \frac{\sum U}{\sum W + \sum U}$	Proportion of items not reached (U) to the total number of items not given correct answers (W+U) summed across all test takers.	Values close to 0 are indicative of a power test. Values close to 1 are indicative of a speeeded test.

Indices	60-Minutes	75-Minutes
Power Ratio	0.14	0.03
Speededness Ratio	0.97	1.00
Speededness Quotient	0.012	0.001

Differential Item Functioning Analysis

Differential item functioning (DIF) analysis was performed to determine if there was a significant difference in item difficulties between test takers who took the 60-minute and 75-minute seat times.

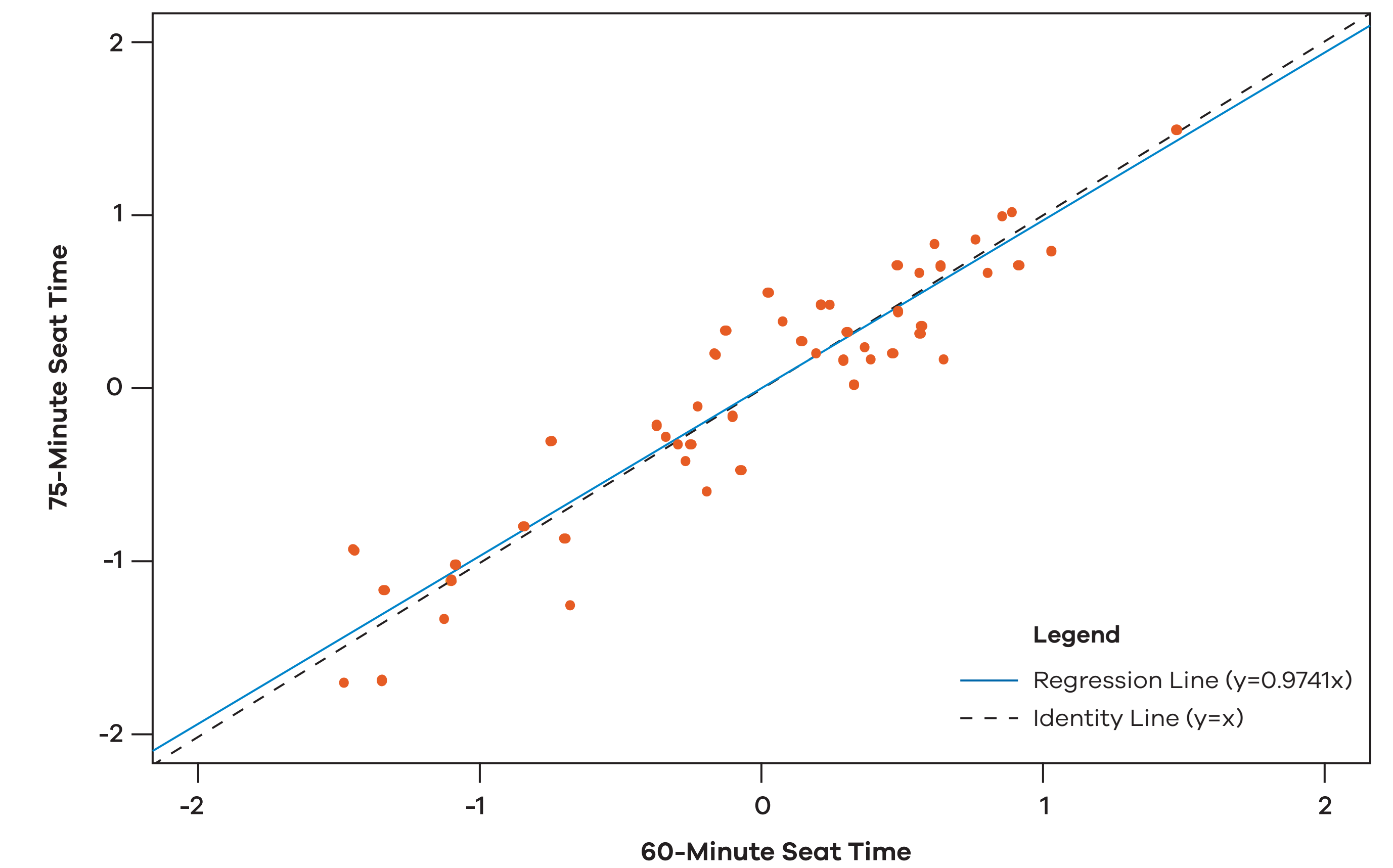


Figure 2: Item Difficulty Comparison

The clustering of the points near the identity line, and the similarity of the regression line to the identity line suggest that the item difficulties were not substantially different, regardless of the seat time allotted. Additionally, the high coefficient of determination ($r^2=0.8771$) shows that most of the variation in the item difficulties for the 75-minute seat time (87.71%) can be explained by the item difficulties for the 60-minute seat time.

Conclusion

Overall, while the analysis of the survey data revealed that the seat time did have a significant impact on the test takers' perception of the amount of time they had to complete the exam, the results of the differential item function analysis indicate that there was not a significant difference in test taker performance on the 60-minute and 75-minute test forms. Furthermore, the Lu & Sireci (2007) indices for evaluating test speededness in a single administration indicate that neither seat time resulted in undue speededness. This evidence suggests that any seat time between 60 and 75 minutes would be appropriate for the revised MET reading and grammar section.

References

Bejar, I. I. (1985), *Test Speededness Under Number-Right Scoring: An Analysis of the Test of English as a Foreign Language*. ETS Research Report Series, 1985: i-57. doi:10.1002/j.2330-8516.1985.tb00096.x

Bolt, D. M., Cohen, A. S. and Wollack, J. A. (2002), *Item Parameter Estimation Under Conditions of Test Speededness: Application of a Mixture Rasch Model With Ordinal Constraints*. Journal of Educational Measurement, 39: 331-348. doi:10.1111/j.1745-3984.2002.tb01146.x

Lu, Y. and Sireci, S. G. (2007), *Validity Issues in Test Speededness*. Educational Measurement: Issues and Practice, 26: 29-37. doi:10.1111/j.1745-3992.2007.00106.x

Oshima, T. C. (1994), *The Effect of Speededness on Parameter Estimation in Item Response Theory*. Journal of Educational Measurement, 31: 200-219. doi:10.1111/j.1745-3984.1994.tb00443.x

Van der Linden, W. J. (2011), *Setting time Limits on Tests*. Applied Psychological Measurement, 35: 183-199. doi: 10.1177/0146621610391648.x



MICHIGAN LANGUAGE ASSESSMENT

Contact Information

Patrick McLain
Assessment Statistician
Michigan Language Assessment
Email: mclain.p@michiganassessment.org

MichiganAssessment.org