

# Stealth Assessment Development: Final Report

Rebecca Tucker

With funding from MITACS Unit and PIVP Vouchers

2/14/20

## Assessment Creation

- A review of available reading comprehension assessments was completed, with two main formats most common: 1 – children read passages and answer multiple choice questions and 2 – children read passages and answer questions orally. Given the format of Dreamscapes content and the context the assessment would be given in, the first format was deemed the only appropriate option.
- The assessment was created by combining aspects of currently available assessments and current Dreamscapes content in order to have a new assessment that follows common procedures while being closely matched to game content.
- Some aspects of currently available reading comprehension assessments taken from the completed review included creating grade levels that include a set number of passages and questions, a rule that moves children through grade level to place them at a level that is neither too easy nor too hard.
- Additional considerations when creating the assessment included balancing the length of the assessment with including enough passages/questions to be accurate, varying types of passages within each grade level, including passages meant for students at the beginning and end of each school grade (e.g., early Grade 3 and late Grade 3 passage within the Grade 3 set), and varying the difficulty and type of questions used for each passage.
- The original assessment contained 6 multiple choice questions per passage; however, upon piloting, completion time was too long and the number of questions per passage was reduced to 4.

\*\* See provided assessment for final assessment structure, passages/items used, and instructions for moving students through the grade sets.

## Participant Info:

All students were in public school in Nova Scotia.

Grade	N	TOSREC Index Score: M (SD)	TOWRE Standard Score: M (SD)	TOWRE Grade Equivalent: M (SD)
2	14	85.50 (13.44)	87.21 (16.37)	1.31 (0.77)
3	16	88.38 (14.20)	90.75 (16.99)	2.43 (1.01)
4	12	85.42 (16.96)	88.83 (14.88)	3.45 (2.29)
5	12	90.17 (16.94)	95.08 (22.13)	4.95 (3.72)
6	12	96.83 (14.17)	99.83 (15.09)	6.23 (1.98)
TOTAL	66	89.09 (15.14)	92.09 (17.30)	--

## Variables in Study

The primary variables for the created assessment include:

- Number of grade levels moved: the difference between where they started and where they were placed.
- Grade level placed at: after completing full assessment as per instructions, the highest passage level for which the student answered at least 3 of 4 question correctly is the level they were placed at.

Below are four examples of students in Grade 4 may have moved through the assessment and what these two variables would be for that student.

Student A:

- Story 4.1: 3 of 4 correct
  - Story 4.2: 1 of 4 correct
- Grade placed at: 4.1**  
**Number of Grades moved: 0**

Student B:

- Story 4.1: 3 of 4 correct
  - Story 4.2: 3 of 4 correct
  - Story 5.1: 3 of 4 correct
  - Story 5.2: 2 of 4 correct
- Grade placed at: 5.1**  
**Number of Grades moved: 1**

Student C:

- Story 4.1: 2 of 4 correct
  - Story 3.1: 3 of 4 correct
  - Story 3.2: 3 of 4 correct
- Grade placed at: 3.2**  
**Number of Grades moved: -0.5**

Student D:

- Story 4.1: 3 of 4 correct
  - Story 4.2: 3 of 4 correct
  - Story 5.1: 2 of 4 correct
  - Story 5.2: 2 of 4 correct
- Grade placed at: 4.2**  
**Number of Grades moved: 0.5**

***Other Measures Included:***

**TOSREC** = Test of Silent Reading Efficiency and Comprehension. This is a brief test of both the speed and accuracy with which students can read sentences for comprehension. The students are provided 3 minutes to read as many short sentences as they can while answering whether each sentence was true or false.

**TOWRE** = Test of Word Reading Efficiency – Sight Word Efficiency subtest. This is a brief test of both the speed and accuracy with which students can read a list of sight words. The students are provided 45 seconds to read as many of the words in the list as they can.

Index and standard scores are standardized scores that place performance in relation to other students at the same age (TOWRE) or grade who took the test at the time of the school year (TOSREC). These scores can be used to get percentiles and place students into descriptive categories, which are used in some results below. The categories line up as per below:

<b>Descriptive Category</b>	<b>Index Scores</b>	<b>Percentiles</b>
Above Average	111 – 145+	77 – 99+
Average	90 – 110	25 – 75
Below Average	80 – 89	9 – 23
Poor	71 – 79	3 – 8
Very Poor	<55 – 70	<1 – 2

**Results Summary**

*Assessment Timing/Length*

- The assessment was not officially timed; however, it took anywhere from 20 to 40 minutes depending on student age and reading ability.
- While the time to completion is somewhat longer than originally discussed, it is in line with other assessments that have children answer multiple choice questions (as compared to those with orally provided answers). It may be that the extra reading time needed for answering the questions, especially for young or struggling readers, is what necessitates the extra time.
  - Possible changes to the questions/structure were considered to shorten the assessment, such as reducing the number of questions, a less stringent rule that would place student's

more quickly, and not having grades separated into two levels (e.g., early Grade X and late Grade X). However, each option ran the risk of reducing the assessment's likelihood of accurately placing a child at an appropriate grade level due to a) having too few questions to form an accurate picture of their capability at a specific grade level and/or b) not adequately accounting for the amount of growth in reading skill within a school year and placing them too early or too late within that range. As such, the assessment was kept as is.

- The average number of passages read within this time were as follows:
  - Grade 2:  $M = 3.07$ ,  $SD = 0.96$
  - Grade 3:  $M = 3.13$ ,  $SD = 0.50$
  - Grade 4:  $M = 3.08$ ,  $SD = 1.16$
  - Grade 5:  $M = 2.69$ ,  $SD = 1.97$
  - Grade 6:  $M = 6.00$ ,  $SD = 1.71$

### *Engagement*

- During piloting, students provided some feedback on the assessment process. This is summarized below:
  - Children were well engaged overall, reporting the task was good and not too hard
  - Children appeared to enjoy the activity, with some variance depending on the passage (e.g., one child reported enjoy a particular passage more than others)
  - Overall, the children appeared to understand the task and become more engaged as the task continued.
  - The children noticed some features of the assessment, such as varying level of difficulty across passages and questions
- During testing, students were not asked for specific feedback. Some notes are summarized below:
  - In this context, some kids found the task a little long and had more difficulty with staying engaged and attentive, particularly those who were experiencing difficulties with it. This difference in relation to piloting is not unexpected when working with a wider number and variety of children within their school day.
  - No specific engagement or behavioural differences were noted in relation to what would be expected based on tester experience, with most appearing typical for child age and context (at school, new people/situation)
  - No specific differences in engagement or behaviour were noted in relation to other similarly lengthy tasks completed as part of the study
- It should be noted that the assessment was provided as a pen and paper task, without graphics or other interactive aspects that would typically be part of the video-game context. As such, the level of engagement noted in this study is likely an underestimation of what would be typical once implemented within the game.

### **Correlations:**

#### *All Participants*

Variable	1.	2.	3.	4.
<b>1. Grade Placed At</b>	-			
<b>2. Number of Grades Moved</b>	.843**	-		
<b>3. TOSREC Index Score</b>	.518**	.626**	-	
<b>4. TOWRE Standard Score</b>	.512**	.628**	.791**	-
<b>5. TOWRE Grade Equivalent</b>	.755**	.653**	.682**	.777**

\*\*  $p < .01$

*Separated by Grade*

Grade 2				
Variable	1.	2.	3.	4.
1. Grade Placed At	-			
2. Number of Grades Moved	.913**	-		
3. TOSREC Index Score	.811**	.756**	-	
4. TOWRE Standard Score	.634*	.604*	.781**	-
5. TOWRE Grade Equivalent	.852**	.786**	.895**	.881**
Grade 3				
Variable	1.	2.	3.	4.
1. Grade Placed At	-			
2. Number of Grades Moved	.925**	-		
3. TOSREC Index Score	.347	.348	-	
4. TOWRE Standard Score	.381	.477	.726**	-
5. TOWRE Grade Equivalent	.374	.436	.737**	.967**
Grade 4				
Variable	1.	2.	3.	4.
1. Grade Placed At	-			
2. Number of Grades Moved	.992**	-		
3. TOSREC Index Score	.450	.430	-	
4. TOWRE Standard Score	.548	.579*	.759**	-
5. TOWRE Grade Equivalent	.315	.317	.901**	.848**
Grade 5				
Variable	1.	2.	3.	4.
1. Grade Placed At	-			
2. Number of Grades Moved	.999**	-		
3. TOSREC Index Score	.633*	.634*	-	
4. TOWRE Standard Score	.610*	.609*	.779**	-
5. TOWRE Grade Equivalent	.612*	.618*	.737**	.956**
Grade 6				
Variable	1.	2.	3.	4.
1. Grade Placed At	-			
2. Number of Grades Moved	1.00**	-		
3. TOSREC Index Score	.898**	.898**	-	
4. TOWRE Standard Score	.912**	.912**	.922**	-
5. TOWRE Grade Equivalent	.885**	.885**	.818**	.936**

\*  $p < .05$ ; \*\*  $p < .01$

## Figures Showing Relationships with Standard Scores

Figure showing number of Grade Levels Moved by TOSREC Index Score

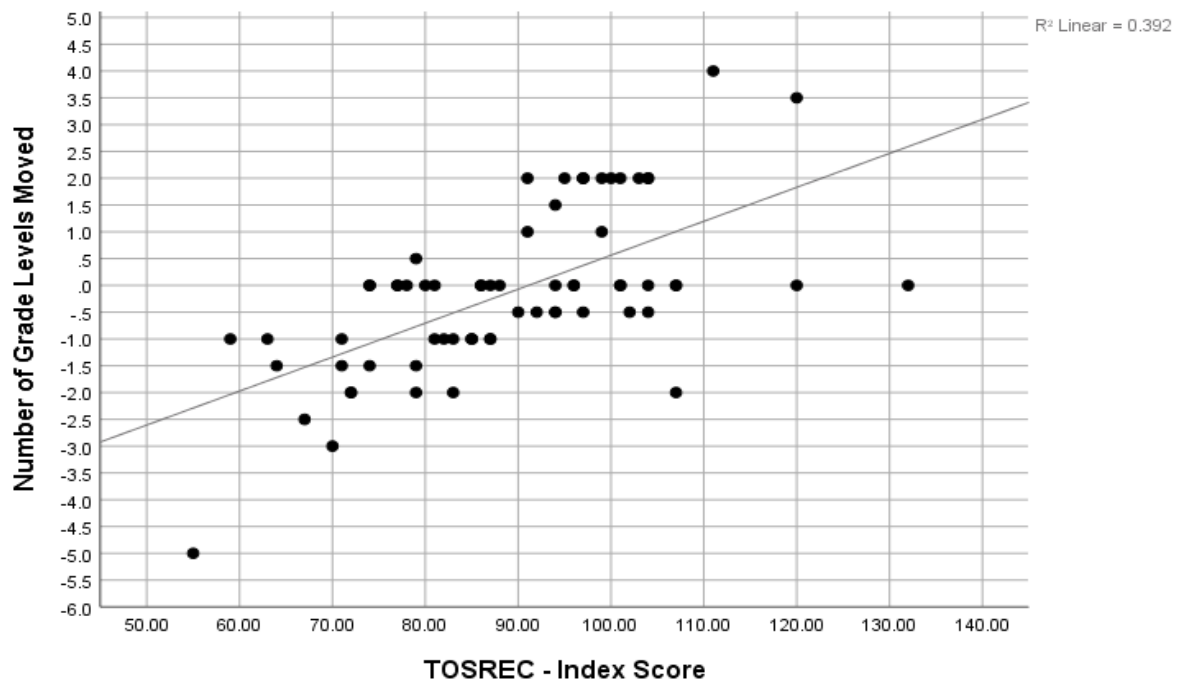


Figure showing number of Grade Levels Moved by TOWRE Standard Score

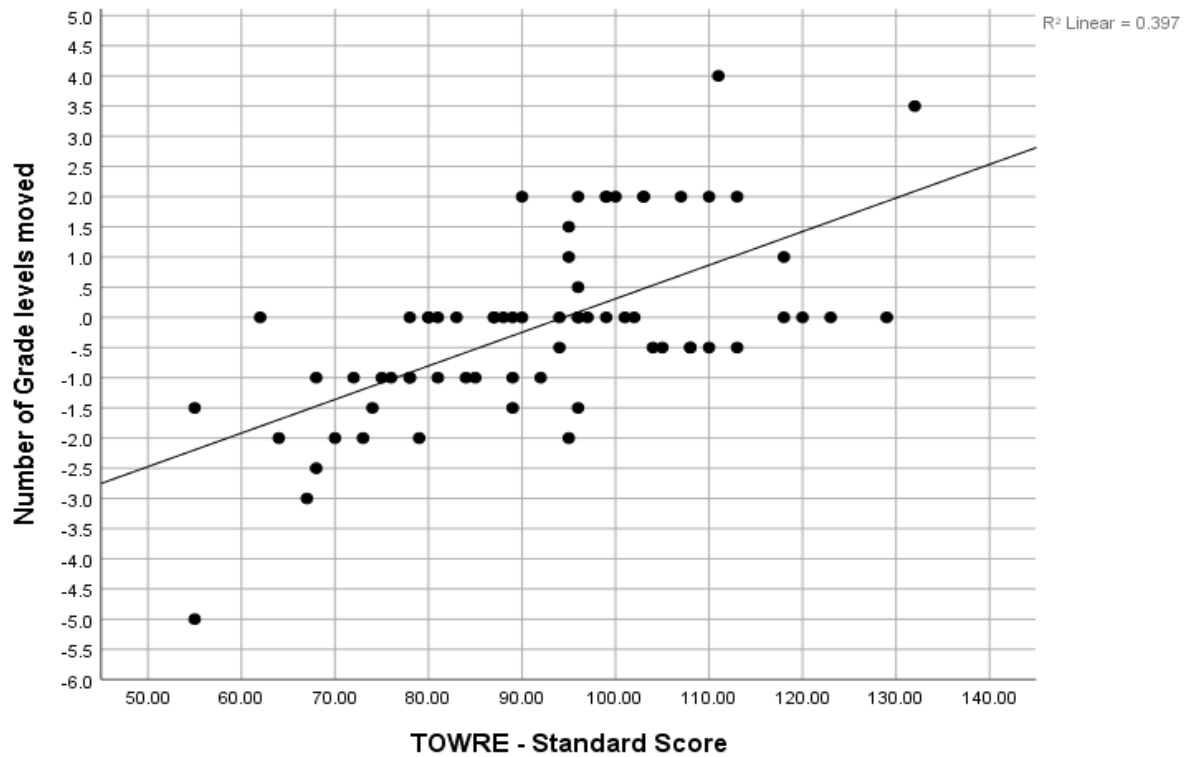


Figure showing number of Grade Levels Moved by Grade and TOSREC Index Score

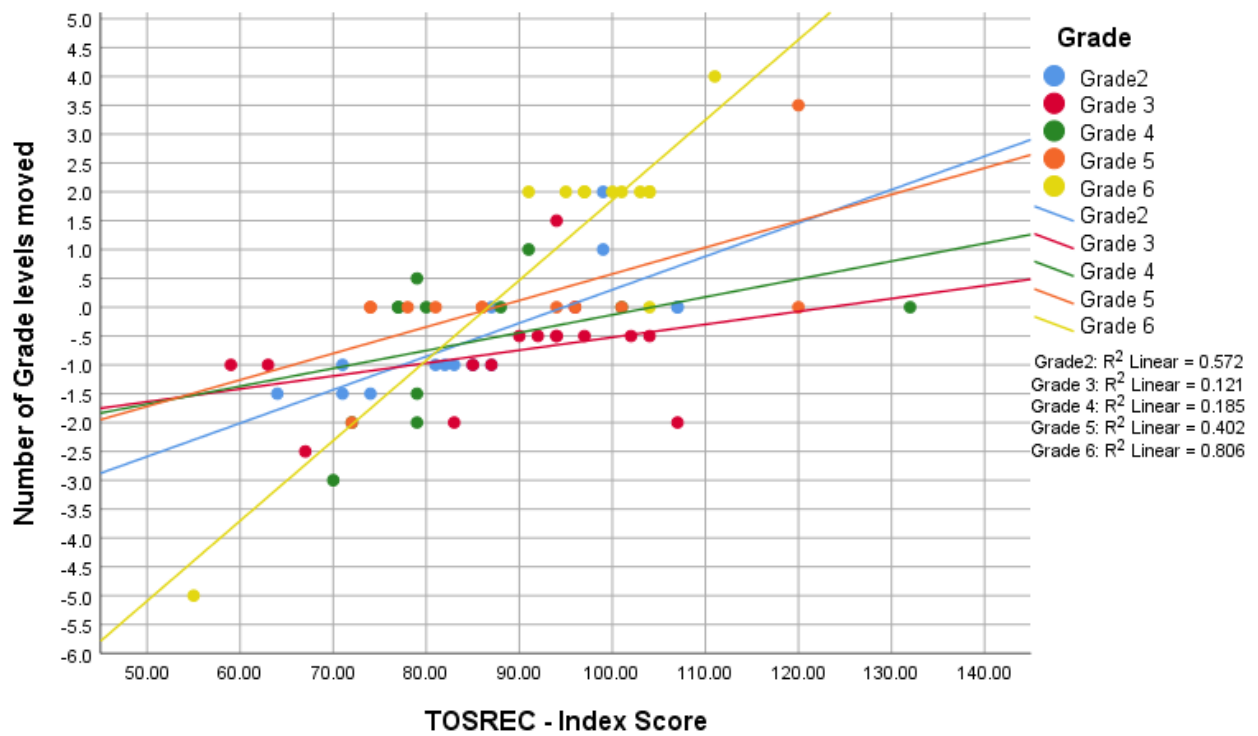


Figure showing number of Grade Levels Moved by Grade and TOWRE Standard Score

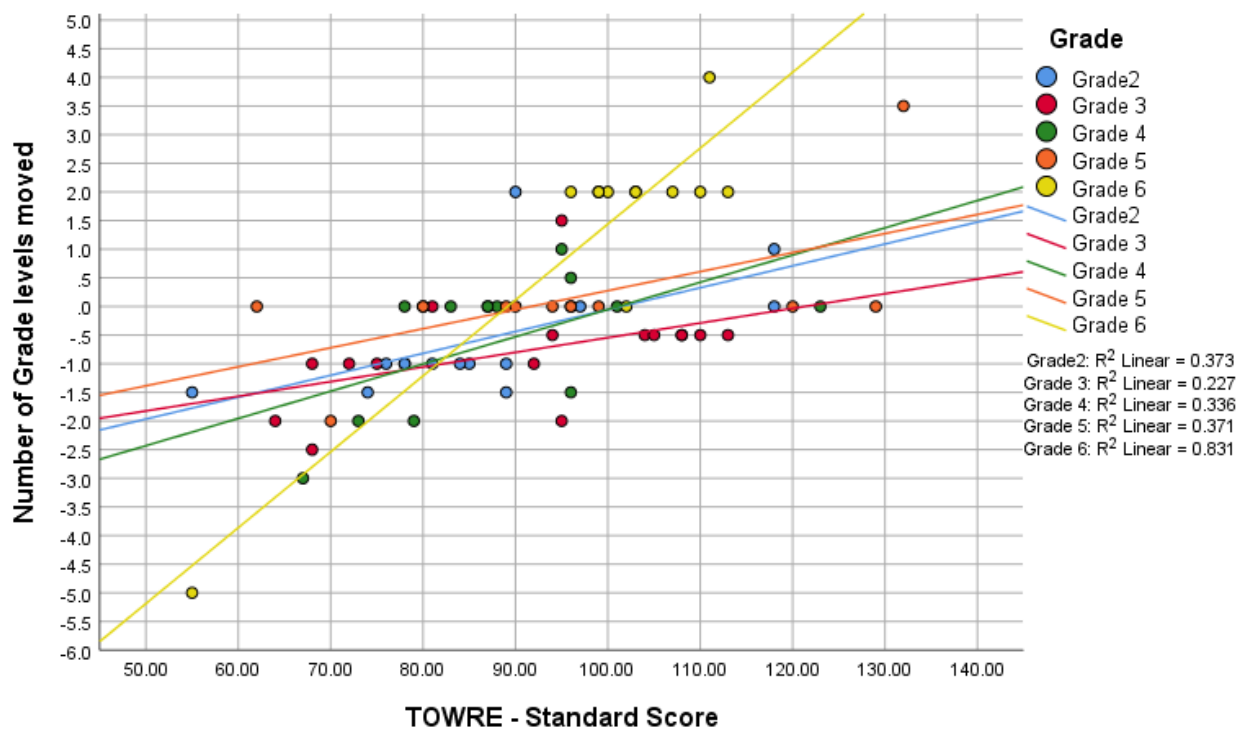


Figure showing grade level placed by grade and TOSREC index scores

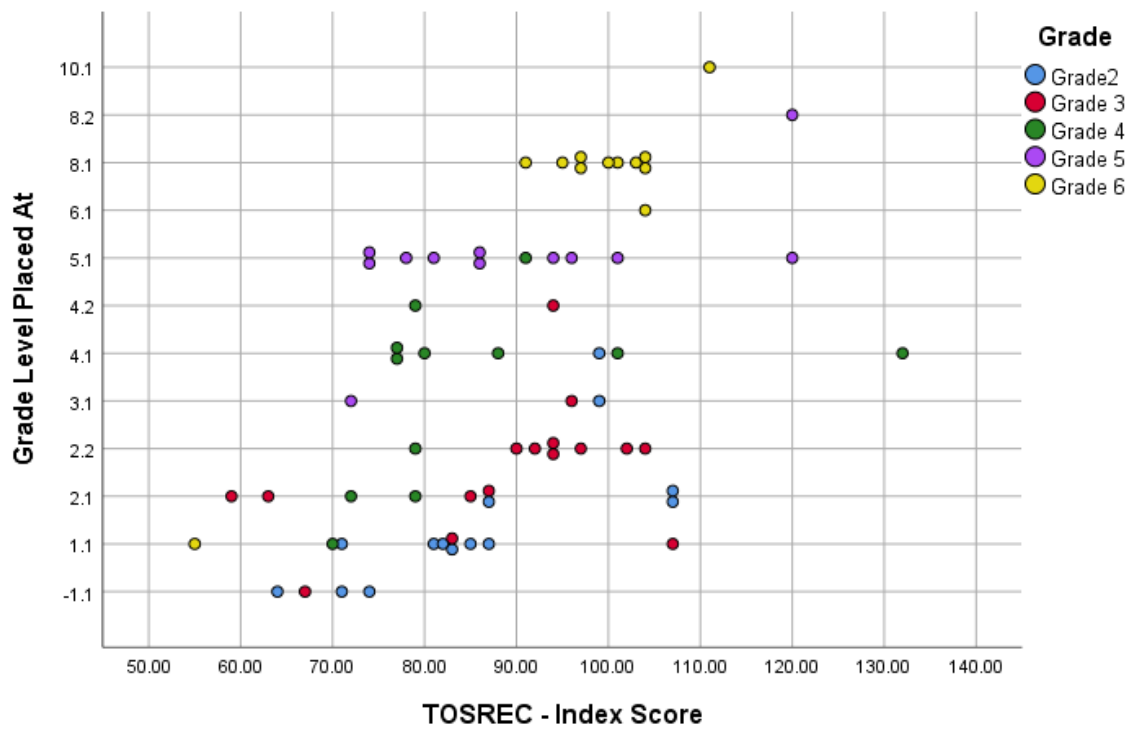
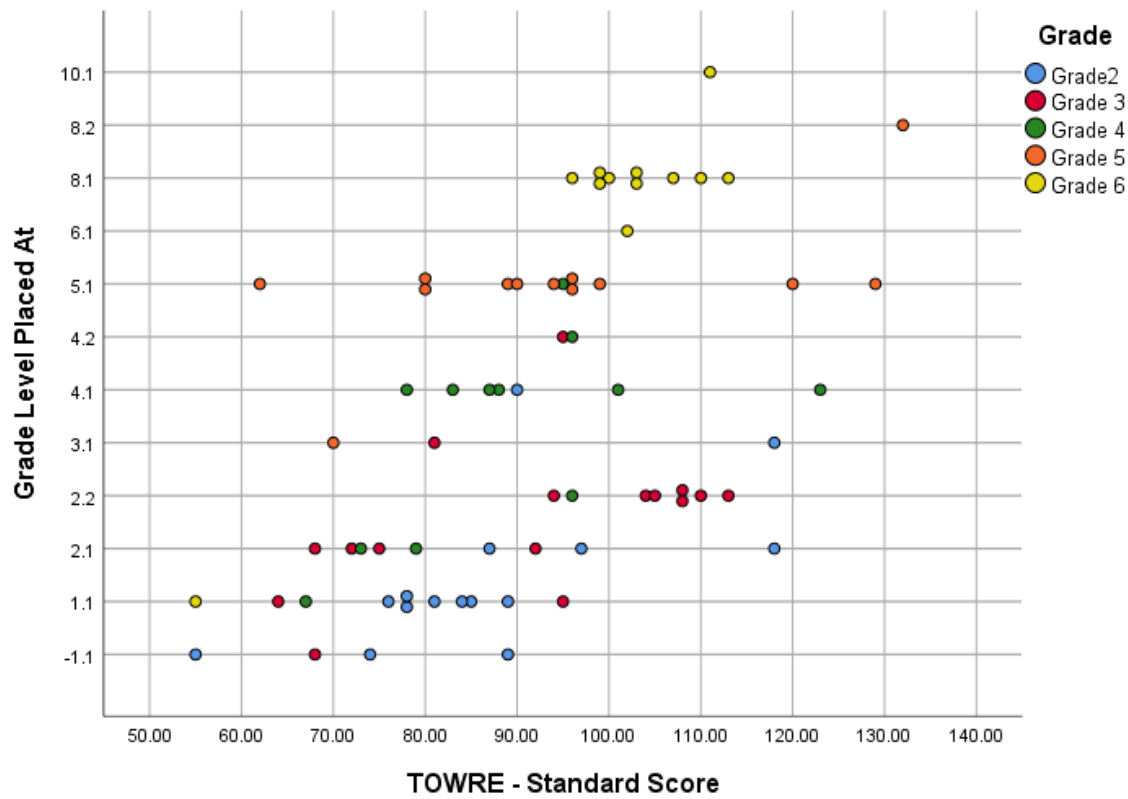


Figure showing grade level placed by grade and TOWRE standard scores





## Figures Showing Relationships with Ability Levels

Figure showing grade levels moved by grade and by TOSREC ability level:

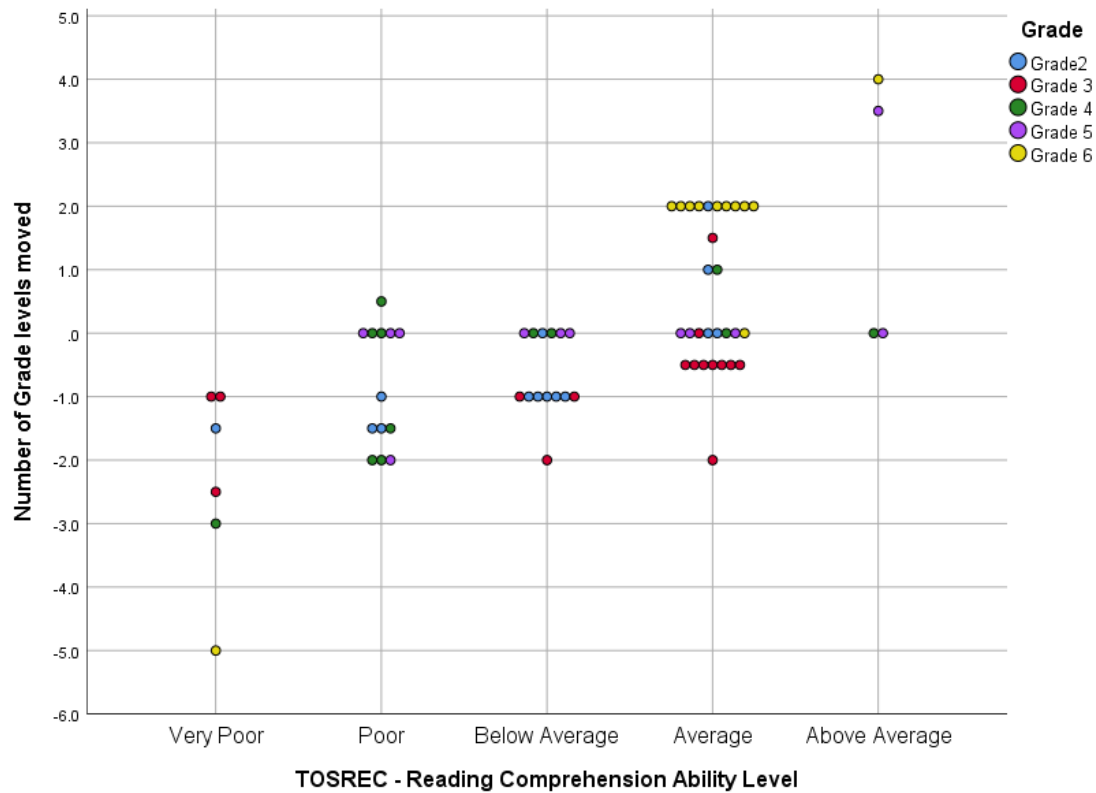
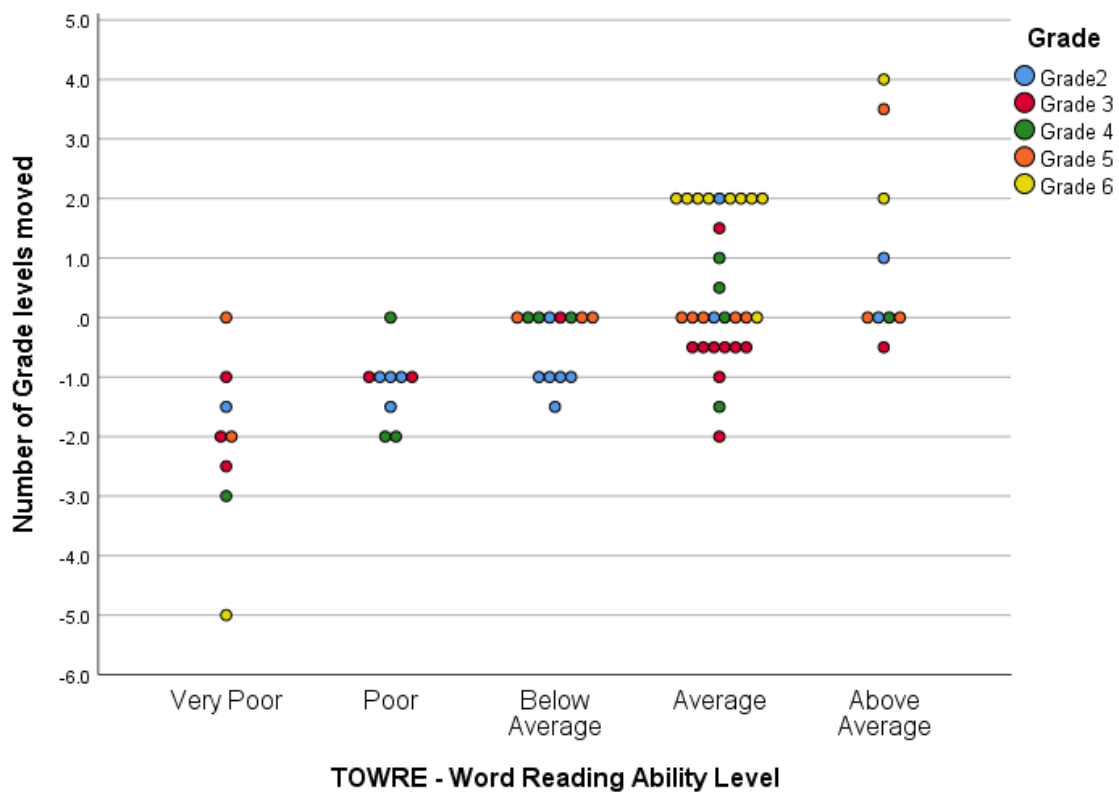


Figure showing grade levels moved by grade and by TOWRE ability level:





## Summary:

- Overall, the assessment results do tend to correlate with standardized scores and reading ability levels, with similar patterns found when using either TOSREC or TOWRE
  - *Grade Placed At:* As TOSREC and TOWRE scores increased, the grade level students were placed at was higher as well.
  - *Number of Grade Levels Moved:* As TOSREC and TOWRE scores increased, the number of grade levels students moved up by increased as well.
- There are some differences in these results across grade, again with similar patterns when using both TOSREC and TOWRE scores:
  - For Grade 2 students, this relationship was consistent and there were strong correlations. Students who scored in the Average range on the TOSREC and/or TOWRE tended to remain at grade level or move up a grade level and those who scored in the Below range tended to remain at grade level or move back a grade level. Finally, students who scored in the Poor or Very Poor grade levels tended to move back one or more grade levels.
  - For Grade 3 students, this relationship was not as consistent and there were no significant correlations. Students who scored within the Average range on the TOSREC and/or TOWRE moved back by 0.5 grade levels on average.
  - For Grade 4 students, there were also no significant correlations and the relationship was more inconsistent. This was particularly true for those with Poor abilities on the TOSREC, with similar numbers both staying at grade level 4.1 (or above) and moving back to Grade levels 2.1 and 2.2.
  - For Grade 5 students there were significant correlations; however, they typically did not move grade levels, regardless of reading ability. Across all reading ability levels, just 2 of 12 students moved up or down from the Grade 5.1 reading level. The two students who did move grades did so in the direction that would be expected.
  - For Grade 6 students, there were strong, significant correlations between the assessment results and reading ability. Notably, students who scored within the Average range on the TOSREC and/or TOWRE moved forward by 2 grade levels on average.
- Taken together, the assessment appears to be working as expected; however, some adjustments may be needed to ensure consistency across all grade levels it will be provided to.
- Note: Some grade 2 students received simplified questions (e.g., fewer and simpler words, no special instructions so they only ever had to choose one answer, etc.). There were no apparent differences between on any of the results presented when the Grade 2s were separated by question set they received. However, qualitatively, the students with the simplified questions tended to need less support and show fewer signs of frustration – especially in comparison items with special instructions (e.g., putting events of the story in order).

## Future Steps:

- Evaluate whether the assessment works within the game content. Although there are some inconsistencies with standardized measures of reading ability, with the assessment relating differently to standardized measures depending on grade, it may still line up well with in-game content across grades. This is because the assessment is drawn from the material that was designed for the game content and, as such, likely aligns more closely with it than standardized measures.

- If adjustments are needed once implemented, one starting place would be changing the cut-offs for moving kids through the assessment and determining their grade level. The rules of a) moving back if 2 of 4 questions for a passage were wrong and b) the grade placed at being the highest level in which at least 3 of 4 were right were created for this assessment. The goal of those cut-offs was to balance a student having enough of a challenge to maintain interest while reducing the chance of a student moving forward to content that is too hard based on guessing at earlier grade levels. However, once larger amounts of data are collected within context the assessment will be used (i.e., through game play), it may be that a more, or less, strict cut-off is more appropriate.
- If adjusting the assessment (and possibly game content) to have it align more closely with standardized reading ability measures across grades, I suggest starting with the questions rather than the passages. Based on these results, and the testing experience, it is possible that the differences across grades are related to the questions. In particular, the format and language used in the questions are similar across grades and they may need to be tailored more for each grade (e.g., a little simpler for younger grades and a little harder for older grades).
  - Qualitatively, younger kids needed more support with reading the questions themselves and understanding how to complete multiple choice questions. While examiners encouraged them to complete the task on their own, it was students in younger grades that asked for support more and that sometimes needed it provided in order to complete the task. Given that this support may not be available at home or school, it may be even harder for those students to understand the questions in practice than in the current testing context.
  - Qualitatively, younger kids appeared more likely to miss the special instructions included in some questions, such as questions requiring them to select more than once answer or put the events of the story in order. These may be other areas that adjustments could be made to questions moving forward.