

Comparison of Progress Monitoring with Computer Adaptive Tests and Curriculum Based Measures

A study conducted by
Edward S. Shapiro, Ph.D.
Denise P. Gibbs, Ed.D.

Bethlehem, PA: Center for Promoting Research to Practice, Lehigh University
2014

Abstract

The purpose of this study was to compare both rates of reading achievement growth and predictive power of two widely-used assessments representing two different approaches to measurement – a computer adaptive assessment called STAR Reading and a curriculum based measurement called AIMSweb. A total of 117 students from a school district in Tennessee were included in the sample. Data collection spanned two school years, and included students who were progress monitored (taking a minimum of 4 tests per year) in grades 1 through 4 in one year, and in grades 2 through 5 the subsequent year. Across the two years, interventions for both groups of students were consistent. The results of this study indicate that both measures were able to detect incremental change, and provide further support that both computer adaptive measures such as STAR Reading and CBMs such as AIMSweb R-CBM are acceptable for progress monitoring. Of the two measures, only STAR Reading achieved a significant correlation with the state reading assessment.

Recommended citation for paper

Shapiro, E. S., & Gibbs, D. P. (2014). *Comparison of progress monitoring with computer adaptive tests and curriculum based measures*. Bethlehem, PA: Center for Promoting Research to Practice, Lehigh University. Available online from <http://coe.lehigh.edu/cprp/research/current>

Introduction

Measures that provide educators with accurate indicators of student growth are central to the Response to Intervention (RTI) framework. When a district changes the universal screening or progress monitoring measures, their RTI framework undergoes a significant change. This type of change occurred in the Tennessee district during the 2012-2013 school year. This district agreed to share data for analysis. The district is a mid-sized public district in Tennessee with five elementary schools. Roughly half the students in the district qualify for Title I services.

Student Demographic Information for District (District Size: ~3600 students, 5 schools)

Demographic Category	Students
African American	7%
Asian / Pacific Islander	0%
Hispanic	3%
Native American / Alaskan	0%
White	90%
Limited English Proficient	1%
Students with Disabilities	13%
Economically Disadvantaged	65%
Title 1	49%

Source: <http://edu.reportcard.state.tn.us/>

The district began to develop and implement an RTI framework for reading in grades K-5 during the 2007-2008 school year. A three tier model was developed with tier one consisting of 90 minutes of core reading instruction provided daily to all students in grades K-5, tier two consisting of 30 additional minutes of reading intervention provided daily to all students scoring below the 25th percentile on a universal screening measure, and tier three consisting of 60 additional minutes of intensive intervention provided to all students who did not evidence a positive response to the tier two reading intervention.

District-level and school-level staff worked to develop an effective RTI framework by analyzing student outcome data in addition to teacher reflections and input. Changes were implemented as needed regarding specific research-based reading instructional practices and interventions. Beginning with the 2009-2010 school year, tier two intervention in reading for grades 1-3 followed a standard treatment protocol format based upon implementation of the *Response Reading Instruction (RRI)* program.

The *Reading Triumphs* program was utilized for tier 2 intervention in grades 4 and 5. For students in need of tier 3 reading interventions, a determination was made regarding the student's specific areas of deficiency and an appropriate tier 3 intervention was implemented. Students with word-level reading deficits (accuracy or fluency) were given a placement test and were then placed in the appropriate level of intervention utilizing the *S.P.I.R.E.* program. Students who read with appropriate accuracy and fluency but who evidenced deficits in reading comprehension participated in the *Voyager Passport Reading* program at the tier 3 intervention level.

Curriculum based measures in reading (R-CBM) were utilized for both universal screening and for weekly progress monitoring for five years (from 2007-2008 to 2011-2012). The RTI teams needed to document growth or lack of growth in complex reading comprehension skills. Given that CBM measures are designed to assess outcomes of reading and not specific complex skills within reading processes, the RTI team examined an alternative measure to the existing R-CBM measure.

The district RTI team determined, at the end of the 2011-2012 school year, that a computer adaptive test could provide broader information regarding students' reading skills particularly in the area of reading comprehension. Beginning with the 2012-2013 school year, STAR Reading was selected by the district as the universal screening and progress monitoring tool.

Methodology

A comparison of the tier 2 and tier 3 progress monitoring data using STAR Reading in 2012-2013 with the tier 2 and tier 3 progress monitoring data using AIMSweb in 2011-2012 was completed as a component of the district's evaluation of this change in the assessment component of their RTI framework. Progress monitoring data for specific students receiving tier 2 or tier 3 interventions throughout both years allowed for a direct comparison of the R-CBM and CAT data in terms of indication of student growth.

The fact that the students took one progress monitoring measure in one year (AIMSweb R-CBM) and a different one (STAR Reading) in the following year is a limitation of the study. However, every other key variable was constant, including: the schools, students, RTI program, interventions, and interventionists, permitting us to address the following research questions:

- Are STAR Reading and AIMSweb (R-CBM) sensitive to incremental growth for progress monitoring?
- How do student actual rates of growth compare to the expected rates of growth in STAR Reading and AIMSweb?
- To what degree do student actual rates of growth compare to the expected rates of growth in STAR Reading and AIMSweb?
- How do STAR Reading and AIMSweb correlate to the Tennessee TCAP assessment?

Answers to these questions are presented below.

Analysis

Are STAR Reading and AIMSweb (R-CBM) sensitive to incremental growth for progress monitoring?

Based on the analysis described below, both AIMSweb and STAR-R are capable of demonstrating incremental growth for progress monitoring purposes. Both measures reflected growth in student performance at all grade levels.

Because progress monitoring is usually used for those students who are substantially below benchmark levels, an examination of growth rate for those below the 25th percentile at the starting point for each metric are displayed below. Table 1 reflects the outcomes for those assessed with AIMSweb in 2011-2012, and Table 2 reflects outcomes for those assessed with STAR in 2012-2013.

Table 1. Mean Rate of Improvement for AIMSweb (WCPM/Week) in 2011-2012 for those who began below and at or above the 25th percentile of the AIMSweb normative distribution.

	Start < 25 th Percentile				Start => 25 th Percentile			
Grade	N	Mean ROI	SD	Range	N	Mean ROI	SD	Range
1	17	1.22	0.60	0.10 to 2.22	17	0.62	0.50	-.23 to 1.50
2	20	1.20	0.47	0.58 to 2.76	16	0.66	1.00	-1.67 to 2.06
3	19	1.42	0.89	0.37 to 3.60	2	0.77	0.28	.57 to .77
4	15	1.15	0.63	0.31 to 2.22	6	1.16	2.14	-.11 to 1.16
OVERALL	71	1.25	0.85	0.10 to 3.60	41	0.72	1.04	-1.67 to 5.49

Table 2. Mean Rate of Improvement for STAR (SS pts/Week) in 2012-2013 for those who began below and at or above the 25th percentile of the STAR normative distribution.

	Start < 25 th Percentile				Start => 25 th Percentile			
Grade	N	Mean ROI	SD	Range	N	Mean ROI	SD	Range
2	32	2.46	2.31	-0.03 to 6.80	3	1.77	1.16	.58 to 2.90
3	28	3.96	2.59	-4.49 to 8.04	8	2.25	1.81	-.55 to 4.59
4	16	3.51	2.41	-0.11 to 8.19	6	1.68	1.16	0.29 to 3.40
5	13	2.47	2.38	-1.90 to 6.53	9	1.17	5.27	-3.40 to 11.27
OVERALL	89	3.12	2.49	-4.49 to 8.19	26	1.69	3.22	-3.40 to 11.27

Results for AIMSweb showed consistent growth with reasonable variability across grades for those below the 25th percentile. Average growth across grades was 1.25 wcpm/week. For those at or above the 25th percentile, growth rates were lower, averaging 0.72 wcpm/week across grades, with much greater variability compared to those below the 25th percentile. This same pattern was evident for STAR with somewhat higher levels of variability.

Outcomes of the basic examination of student growth rates using AIMSweb or STAR-R demonstrated consistently that student growth was evident under both metrics. Because it would be more common for progress monitoring to be initiated for students at the lower end of the performance in reading, the data were examined for those below and at or above the 25th percentile of the distributions of the measures on which they were assessed.

Outcomes of the basic examination of student growth rates using AIMSweb or STAR-R demonstrated consistently that student growth was evident under both metrics.

Growth among students below the 25th percentile was not surprisingly much higher (between 30-50% higher) than those at or above the 25th percentile, and the data were consistent across AIMSweb and STAR-R measures. The amount of variability (i.e., bounce in the data) was a little higher for the assessment using STAR-R than for AIMSweb.

How do student actual rates of growth compare to the expected rates of growth in STAR Reading and AIMSweb?

Although both measures showed growth, the key question is the degree to which each measure reflected growth against the amount of growth expected for students. Based on the analysis below, both assessments showed consistent to strong growth across grades.

Two approaches were used to establish the amount of growth expected for students:

- (1) An expected growth rate was set for each student by the data teams.
- (2) The expected growth rate used in the analysis was the normative growth rate for students who started at the initial percentile of the student. Expected growth was determined for both measures by their respective publishers. Both have identified growth norms indicating typical rates of weekly change, or growth, on their test scales, contingent upon the student's grade and initial score during the current academic year. For example, a second grade student whose AIMSweb initial score was at the 10th percentile had an expected growth rate based on AIMSweb normative data of 1.46 WCPM/Week. A second grade student starting at the 15th percentile had an expected growth rate of 1.29 WCPM/Week. Similarly, a second grade student on STAR-R starting at the 10th percentile, had an expected growth rate of 2.50 Scaled Score Points/week, whereas a student starting at the 15th percentile on STAR-R had an expected growth rate of 4.00 SS pts/week.

In an effort to standardize the process for comparing expected and actual student growth between the measures, it was decided to use the second approach described above to derive a metric reflective of the relationships between a student's actual and expected growth.

In order to provide a quantitative index of the relationship between a student's actual and expected growth, the GROWTH INDEX (GI) was calculated for each student by dividing the ACTUAL GROWTH/ EXPECTED GROWTH x 100. A GI equal to 100 indicated that the student had achieved a level of growth equal to their expected growth. A score greater than 100 indicated that the student's growth exceeded their expected growth. A score less than 100 indicated that the student's growth was less than their

expected growth. A negative GI score indicated that the student's trend across the year reflected negative growth, indicating the student's performance had declined across the year.

GI values were calculated for each student and examined for those below and at or above the 25th percentiles for both AIMSweb and STAR-R measures. The mean levels (and variance around the mean) reflect the degree to which each of the measures reflect performance that is representative of what was expected.

Because progress monitoring is usually used for those students who are substantially below benchmark levels, an examination of growth rates for those below the 25th percentile at the starting point for each metric is displayed below. Table 3 reflects the outcomes for those assessed with AIMSweb in 2011-2012, and Table 4 reflects outcomes for those assessed with STAR in 2012-2013.

Table 3. Mean Growth Index (GI) for AIMSweb (WCPM/Week) in 2011-2012 for those who began below and at or above the 25th percentile of the AIMSweb normative distribution.

Grade	Start < 25 th Percentile				Start => 25 th Percentile			
	N	Mean ROI	SD	Range	N	Mean ROI	SD	Range
1	17	213.29	140.14	16.16 to 449.74	17	47.64	39.65	-20.13 to 120.55
2	20	120.74	86.79	45.39 to 412.16	14*	58.13	60.50	-25.91 to 165.13
3	19	155.85	110.53	36.02 to 433.29	2	70.70	27.03	51.60 to 89.82
4	15	146.70	80.74	38.13 to 291.13	6	128.18	225.38	-14.08 to 583.67
OVERALL	71	157.77	110.21	16.16 to 449.74	39	64.98	96.96	-25.91 to 583.67

*Outlier case eliminated from analysis.

Table 4. Mean Growth Index (GI) for STAR-R (SS PTS/Week) in 2012-2013 for those who began below and at or above the 25th percentile of the STAR-R normative distribution.

Grade	Start < 25 th Percentile				Start => 25 th Percentile			
	N	Mean ROI	SD	Range	N	Mean ROI	SD	Range
2	29	92.01	82.73	-1.02 to 271.10	3	43.47	33.28	10.70 to 77.25
3	28	138.97	63.92	1.77 to 253.47	8	79.40	61.35	-18.91 to 155.23
4	16	131.30	90.45	-4.05 to 314.94	5	84.36	48.28	36.61 to 158.88
5	13	98.93	97.27	-89.57 to 259.28	8	63.41	276.11	-160.24 to 555.96
OVERALL	84	115.10	82.72	-89.57 to 314.94	24	70.61	158.22	-160.24 to 555.96

Results for AIMSweb below the 25th percentile showed consistent growth with reasonable variability across grades. Average GI across grades was 157.77. For those at or above the 25th percentile, the GI was lower, averaging 64.98 across grades, with much greater variability compared to those below the 25th percentile.

For STAR, students below the 25th percentile showed strong growth against expected growth levels for grades 3 and 4 with growth just below expected levels for grades 2 and 5. Variability across students was much greater for students in grades 2 and 5 as well.

To what degree do student actual rates of growth compare to the expected rates of growth in STAR Reading and AIMSweb?

Another way to examine the relative relationships between the Growth Index generated by STAR and by AIMSweb, is to simply subtract the GI generated by STAR minus one generated by AIMSweb. Values that are positive would reflect that STAR displayed greater growth against the expected level of performance than AIMSweb. Values that are negative would indicate that AIMSweb showed greater growth against expected level of performance than STAR. This simple formula (Diff in GI = STAR GI – AIMSweb GI) was calculated using the GI based on the starting percentile of the student on the respective measures. Data were then examined for those beginning <25th percentile and those at or above the 25th percentile.

Table 5 reflects the percentage of students at each grade who met or exceeded their expected growth levels (i.e., GI ≥ 100) on STAR and AIMSweb, based on the starting percentile Standard Score for students on STAR. As evident from Table 5, AIMSweb reflected a higher percentage of students meeting or exceeding their goals with the exception of grade 2.

Table 5. Mean difference between STAR GI and AIMSweb GI for those starting with STAR Scaled Scores below and at or above the 25th percentile.

	Start < 25 th Percentile				Start => 25 th Percentile			
Grade	N	Mean Diff	SD	Range	N	Mean ROI	SD	Range
2	29	-12.84	-.90	-314.52 to 242.93	3	-170.73	187.41	-370.28 to 1.54
3	28	30.58	126.33	-389.81 to 199.60	8	6.63	90.23	-135.71 to 125.05
4	16	-33.34	164.47	-400.72 to 233.62	5	-9.31	75.01	-129.52 to 67.54
5	13	-59.33	135.81	-375.24 to 133.92	8	-50.62	281.89	-307.18 to 380.75
OVERALL	86	-9.54	138.57	-400.72 to 242.93	24	-37.94	181.17	-370.28 to 380.75

How do STAR Reading and AIMSweb correlate to the Tennessee TCAP assessment?

The final data point collected at the end of the year and its relationship to the Tennessee State Assessment was calculated. For AIMSweb, the correlations across all third and fourth grade students for 2011-2012 resulted in a correlation of .14 (n.s.) across 41 students. For STAR, correlations of ROI to the Tennessee State Test across all third, fourth, and fifth grade students for 2012-2013 was .42 ($p < .001$) across 67 students.

Conclusions and Findings

It is very important to recognize that the data set has a number of substantial confounds in the comparisons being made. Data are being compared across two years among the same students on two different measures. As such, both the measure and time are confounded and the impact of events occurring across the two years cannot be separated. Although the same intervention was used across most students across the entire time, both measures were never used in the same year. As such it is difficult to know differences in the impact of the measurement system to reflect outcomes against expected levels of performance.

Second, even looking at the same grade across years results in looking at two different cohorts of students. That is, grade 2 students under AIMSweb compared to grade 2 students under STAR are looking at two different groups of students, not the same students across these two measures.

Third, sample sizes are somewhat small, especially at the individual grade level. As such, conclusions based on this data set are certainly tenuous and need replication with larger, similar data sets.

Both AIMSweb and STAR reflect growth of student performance over time.

Because progress monitoring is used primarily for students whose performance is at the lower end of the performance distribution, it makes the most sense to focus the conclusions of the data set on these particular students. In general, both AIMSweb and STAR measures across grades reflected average levels of improvement across the school year among these students. Examination of the amount of variability in the data showed that STAR relative to AIMSweb showed somewhat more “bounce” in the data, particularly at grades 2 and 5. An important conclusion from the data is that BOTH measures were very capable of reflecting student change in performance over time.

Both progress monitoring measures are designed to reflect relative outcomes of student performance against expected levels of growth. To establish a standardized level of growth, the normative rate of improvement expected for a student based on the percentile at which they started the year (i.e., their first data point) was used as the expected rate of improvement. The Growth Index (GI) was calculated for each student by dividing the actual attained ROI by the expected ROI and multiplying by 100. A score of 100 or better indicated a student had reached or exceeded the expected growth index, a score below 100 indicated the student had fallen short of the expected growth index, and a negative score indicated the student’s actual ROI had moved in a negative direction over the course of the year.

AIMSweb tended to reflect growth against expected levels of performance greater than STAR.

Examination of the Growth Index for AIMSweb and STAR across the two years showed that while both measures reflected substantial growth against expected levels of performance, AIMSweb consistently showed a level of growth against expected performance somewhat greater than STAR. This was true particularly for students who started the year below the 25th percentile of the distribution of the measure in which they were being measured. Similar findings were reflected in the examination of the percentage of students meeting or exceeding expected growth for each measure. The data reflect that only in grade 3 did the percentage of students meeting expected growth for STAR exceed the percentage for AIMSweb.

Only STAR Reading is significantly correlated to Tennessee TCAP state. The correlation between the ending data point of STAR and TCAP is moderate.

Examination of the correlation of the ending data point for STAR and AIMSweb to the state assessment outcome was found to be significant and moderate for STAR ($r = .42, p < .001$) and non-significant and low for AIMSweb ($r = .14, n.s.$). These data are important and suggest that STAR is the stronger predictor to state outcome among these students.

STAR Reading and AIMSweb are both sensitive to incremental growth for progress monitoring purposes.

An important question that has consistently been asked is whether STAR-R can be effectively used as a progress monitoring tool, especially compared to curriculum based measures like AIMSweb R-CBM. The data from Dyer County showed that among students for whom progress monitoring would be likely (i.e., those starting the year < 25th percentile), STAR-R data reflected student's incremental growth at all grades. Further, student growth in STAR matched or closely matched the expected level of growth. This was also evident for AIMSweb.

The data from this study indicate that STAR Reading is an equally acceptable method for progress monitoring as AIMSweb R-CBM.

Importantly, the amount of “bounce” in the STAR data relative to AIMSweb was greater, although it remained within the bounds that one would expect for a computer adaptive test. Because AIMSweb uses a single skill (i.e., oral reading fluency) as an index of other skills, bounce in the data occurs usually as a function only of the differences in passage difficulty as a student progresses across the year. In strong passage sets such as AIMSweb, the amount of bounce in the data is minimized.

In contrast, a computer adaptive test such as STAR assesses a broad range of reading skills resulting in the potential for more variability in student performance as students acquire skills differentially through the instructional process. Even with the greater amount of bounce in the STAR data, the level and range of the bounce is well within expected and accepted boundaries.

The data from this study indicate that STAR-R is an equally acceptable method for progress monitoring as AIMSweb R-CBM. The measure offers the added knowledge of skill development and specific direction for instructional modification, an element not evident for AIMSweb R-CBM measures.

About the Authors

Edward S. Shapiro, Ph.D., is professor of school psychology and director of the Center for Promoting Research to Practice in the College of Education at Lehigh University. He is the 2006 winner of the American Psychological Association's Division of School Psychology Senior Scientist Award. Professor Shapiro has authored 14 books and is best known for his work in curriculum-based assessment and interventions for academic skills problems. Among his many projects, Shapiro recently completed a federal project focused on the development of a multi-tiered, RTI model in two districts in Pennsylvania, and currently directs a U.S. Department of Education grant to train school psychologists as facilitators of RTI processes. He also collaborates with the Pennsylvania Department of Education in developing and facilitating the implementation of the state's RTI methodology.

Denise P. Gibbs, Ed.D., is director of the Alabama Scottish Rite Foundation Learning Centers and serves as a reading specialist and RTI consultant for school systems across the country. She has more than 30 years of teaching experience at all levels and is the author of three books: *RTI for Early Readers: Implementing Common Core Standards in Your K-5 RTI Model* (2012), *RTI in Middle and High School: Strategies and Structures for Literacy Success* (2009), and *Leading the Dyslexia Challenge: An Action Plan for Schoolwide Identification and Intervention* (2004).