Evaluating the Predictive Value of Standards-Based Formative Assessments to Inform Online Instruction and Academic Program Management

Using data to inform instruction is an essential component of effective educational practice. The ability to gather and utilize data in multiple ways enables educators to gain significant insight into student learning, curriculum effectiveness, and instructional strategies. To obtain this valuable information, educators have at their disposal many different forms of assessment, each with a specific purpose. For example, to guide instructional practices, teachers often rely on formative assessments, a wide variety of methods utilized in evaluating student learning and comprehension during the progression of a lesson or course. Formative assessments help teachers identify student progress in acquiring skills, understanding concepts, and learning standards. This type of evaluation enables teachers to determine gaps in student learning, which allows them to differentiate instruction and improve student achievement.

Teachers and administrators alike strive to promote high expectations for student learning and ensure students make progress toward demonstrating proficiency on state academic standards. Given these goals, the ability to predict student end of year performance is of significant importance. When formative and summative assessments are deliberately aligned, the formative assessment results provide educators with what Ainsworth (2007, p. 80) calls "predictive value" as to how students will likely perform on subsequent assessments, allowing educators to evaluate learning needs and make instructional modifications. When proven to predict future performance, formative assessments can serve as efficient screening tools for identifying students who are not on track to perform at grade level proficiency on end-of-year assessments. Utilizing the predictive value of formative assessments can be useful for administrators as well—knowing how many and which students may be at-risk can inform an administrator's decisions related to academic program management, such as how best to allocate resources, implement school or system wide initiatives, or take other strategic actions.

UTILIZING VALID ASSESSMENTS TO DRIVE STUDENT SUPPORT INITIATIVES

Data-driven frameworks that promote positive student academic and behavioral outcomes, such as Multi-tiered Systems of Support (MTSS), have become embedded in many schools. MTSS is a whole-school framework for improving learning outcomes for all students delivered through a continuum of evidence-based practices (Ziomek-Diagle, Goodman-Scott, &Donohue, 2016). MTSS provides a proactive approach through which struggling students can be identified and receive targeted intervention. Recommendations for best practices in the planning and implementation of an MTSS include emphasis on data-driven decision making, timely interventions, and consistent monitoring of student progress to improve academic and behavioral outcomes. Adherence to these recommendations hinges on the ability to integrate data into the day-to-day operations of the school.

ASSESSMENT OF LEARNING OBJECTIVES FROM MULTIPLE ATTEMPT SCORES

Research has shown that students benefit from feedback and flexibility in assignments. A common teaching practice is to allow students more than one attempt on certain assignments, with the average or highest score taken when there are multiple attempts. This allows early attempts to serve as feedback for students and increases the amount of practice afforded with the learning material (Doorn, Janseen, & O'Brien, 2010). Within the context of online learning environments, the allowance of multiple attempts to answer assessment items, and even some online exams, has become prevalent and results in a series of observed scores for assessments (Bergner, Colvin, & Pritchard, 2015). This series of scores can serve as a valuable source of data. First attempt scores can inform educators in providing scaffolded feedback while subsequent attempts allow teachers to make inferences about student perseverance and learning gains. This information could enable teachers to encourage students and celebrate progress.
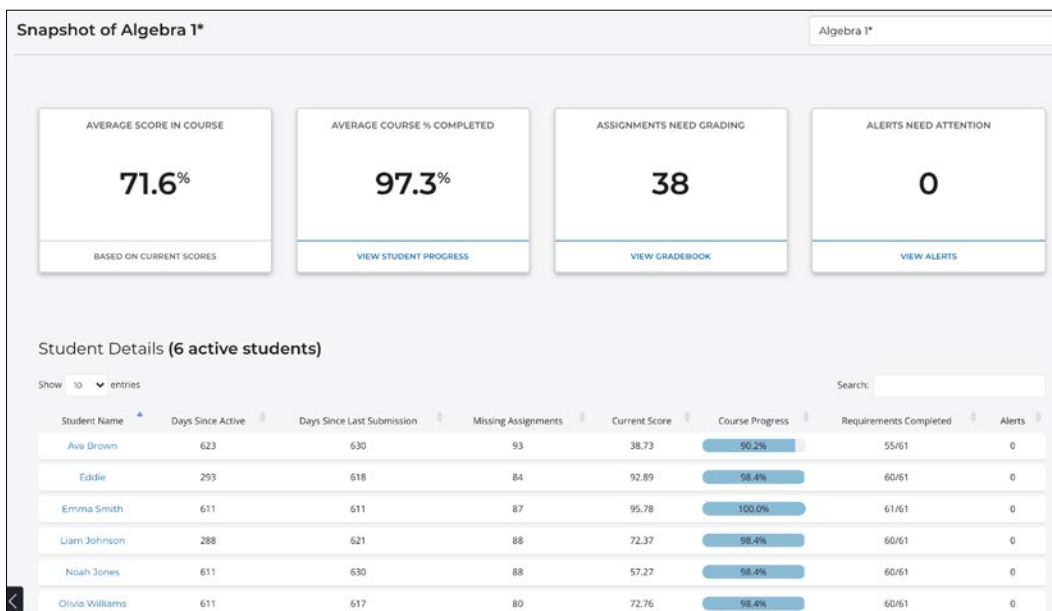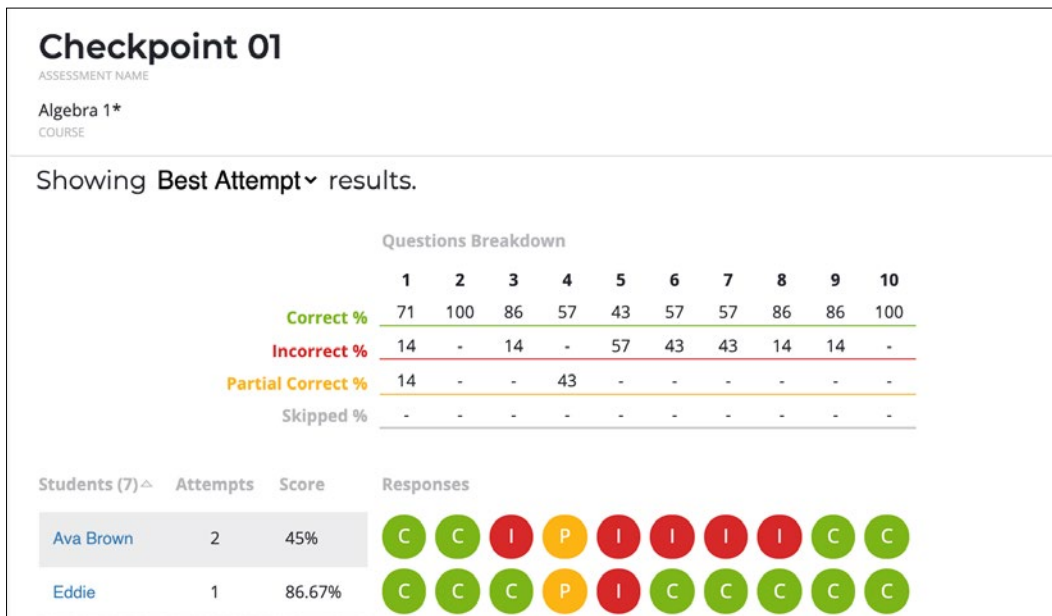
THE CURRENT STUDY

Given the importance of establishing the predictive value of formative assessments, we attempted to answer the following research questions:

1. To what extent do in-course formative assessment results predict end-of-year-performance on high-stakes state assessments?

2. To what extent do multiple attempts on in-course formative assessments impact the ability to predict end-of-year state testing performance?

METHODS

👤 PARTICIPANTS

This study included 895 high-school students attending an accredited public charter school offering online courses to Arizona students in grades 6 – 12.

RACE	N	%
White/Caucasian	529	59
Black	47	5
Hispanic or Latino	214	24
Asian	14	2
Multiracial	60	6
American Indian/Alaskan Native	26	3
Native Hawaiian/Pacific Islander	5	1

F/RP Lunch	N	%
Does not qualify	485	54
Qualifies	410	46

AGE	N	%
15	46	5
16	193	22
17	280	31
18+	376	42

⚲ STRONG**MIND**™ STANDARDS-BASED COURSE ASSESSMENTS

StrongMind courses include assessments that are intentionally crafted to align to state academic standards for English Language Arts (ELA) and Mathematics by breaking down each standard into learning objectives for both instruction and assessment purposes. The assessments, embedded at multiple points within the online course, are designed to assess the instruction of discrete learning objectives in multiple ways throughout the online course. Providing both low-stakes practice and high-stakes testing evaluation, the assessments facilitate learning, practice, critical thinking and assess student proficiency.

The assessments are technology enhanced, utilizing both teacher and system grading options. System scored assessments include workbook activities, checkpoints, unit exams, and final exams:

- **WORKBOOK ASSIGNMENTS** - use targeted questions for practice and to assess students' mastery of an activity's objectives. Workbook assignments are utilized at multiple points throughout each course.

- **CHECKPOINT ASSESSMENTS** – designed to assess student mastery of lesson objectives, are lower-stakes assessments that can help students and teachers identify gaps in knowledge that need to be filled before unit and final exams. There is one checkpoint assessment per lesson, totaling five for each unit.

- **UNIT EXAMS** – assess mastery of unit objectives. Students complete one of these exams at the end of each unit.

- **FINAL EXAMS** – designed to assess mastery of standards and key outcomes for the course. There is one final exam per course.

Teacher graded assessments include discussion board posts as well as writing and project activities.

- **DISCUSSION BOARD POSTS** – give students the opportunity to critically evaluate and share thoughts and ideas about the topics covered in the lesson. They promote student engagement and interaction by prompting students with questions/topics that are open-ended. Each unit typically requires students to submit one or two discussion board posts.

- **WRITING AND PROJECT ACTIVITIES** – necessitate a greater level of thinking and reflection on course content. These assignments require students to apply their knowledge and writing capabilities to demonstrate mastery of the content presented in the lesson and across lessons. These assessments result in posters, infographics, presentations, essays, experiments, and other similar deliverables. Students complete one or two projects for each course.

ARIZONA'S STATEWIDE ACHIEVEMENT ASSESSMENT FOR ENGLISH LANGUAGE ARTS AND MATHEMATICS (AZMERIT)

This study also utilized the annual statewide achievement test for Arizona students, AzMERIT, administered in the Spring of 2019. For high school students, AzMERIT assessment follows the completion of coursework in ELA grades 9-11, as well as Algebra I, Algebra II, and Geometry. Standardized tests, such as state achievement tests, present the same or very similar assessment items to all students while closely adhering to prescribed procedures for test administration and scoring.
This facilitates the comparison of scores across individuals (APA, 2014). Student achievement on the AzMERIT is classified into one of four performance levels: Minimally Proficient, Partially Proficient, Proficient, and Highly Proficient. For the purposes of this study, we examined student's AzMERIT performance relative to the cut score required for proficiency in each tested subject. The Proficient level cut score was chosen because it indicates that students are meeting grade-level expectations for achievement on state standards. Because AzMERIT utilizes a vertical scale across grades and within each subject, a residual score was created by subtracting the student's AzMERIT score from the Proficient cut score for the grade and subject area in order to compare student performance across subject areas on a common proficiency scale.

TESTED SUBJECT	N	%
Algebra I	142	14
Algebra II	120	12
ELA Grade 9	222	21
ELA Grade 10	232	22
ELA Grade 11	190	18
Geometry	134	13

TABLE 2

DATA ANALYSIS

Validity refers to the degree to which test score interpretations are supported by evidence and theory, especially regarding the legitimate uses of test scores. This study utilized course assignment and exam grades as a measure of formative assessment performance. Assignments with a grade of 0 were not considered to be a true attempt at assignment completion and were excluded from analysis. Each student's course assessment results were matched to a corresponding AzMERIT score in the appropriate tested subject in order to conduct the analysis.

When assessment scores are to be interpreted in multiple ways, such as describing student's current level of knowledge acquisition and to make predictions about future outcomes, each intended interpretation must be validated (AERA, 2014).

Regression analysis is a set of statistical processes for estimating the relationship between a dependent variable and one or more independent variables. Regression is widely used for estimation and prediction (Kutner et al., 2005). For this study, simple linear regression was utilized to examine how well scores on course assignments predicted students' AzMERIT scores.

We first examined how well students' best performance scores on the formative assessments predicted performance on the AzMERIT. Next, we examined how well students' assignment grades based on their first attempt predicted performance on the AzMERIT. Regression analysis was conducted for all assignments as well as each assignment group.

RESULTS – *Best course assessment scores*

Findings indicate that, when using students' best scores, performance on the standards-based course assessments was found to be a significant predictor of student performance on the statewide assessment $F(1, 100031) = 5024.34$, $p<.001$. Descriptive statistics for students' best performance scores, with means and standard deviations for all assignments and assignment groups are provided in Table 3.

The study found a positive and significant correlation (.217) between all course assessment scores and performance on AzMERIT with course assignments accounting for approximately

5% of the variance in the state assessment scores. When examining the relationship between course assignment categories and the state assessment, Unit Exams were found to have the highest correlation (.446) with AzMERIT performance, explaining approximately 20% of the variance in AzMERIT scores. Checkpoints (.345) and Final Exams (.341) also demonstrated moderate and significant correlations with AzMERIT scores, with each assignment group explaining approximately 12% of the variance in AzMERIT scores.

BEST COURSE ASSESSMENT SCORES

	N	MEAN	SD
All Assignments	100,024	89.93	22.81
Checkpoints	17,087	70.55	23.74
Discussions	23,024	89.63	18.03
Projects	3,801	74.92	25.35
Workbooks	50,783	88.47	21.22
Unit Exams	4,602	62.74	22.04
Final Exams	727	81.18	15.73

TABLE 3

CORRELATION BETWEEN ASSIGNMENT GROUPS AND AZMERIT SCORES BASED ON BEST ATTEMPT

	N ASSIGNMENTS	R	R SQUARED*
All Assignments	92,664	.268	.072
Checkpoints	16,763	.448	.200
Discussions	22,576	.152	.023
Projects	2,591	.285	.081
Workbooks	45,501	.278	.077
Unit Exams	4,507	.469	.220
Final Exams	726	.331	.109

TABLE 4

ASSIGNMENT GRADES BASED ON FIRST ATTEMPT

When utilizing students' first attempt scores, performance on the standards-based course assessments was again found to be a significant predictor of student performance on the statewide assessment $F_{(1, 92662)} = 7157.75$, $p<.001$. Descriptive statics for students' first attempt scores, with means and standard deviations for all assignments and assignment groups are provided in Table 5.

The study found a positive and significant correlation (.268) between all first attempt assessment scores and performance on AzMERIT with course assignments accounting for approximately

7% of the variance in the state assessment scores. When examining the relationship between course assignment categories and the state assessment, Unit Exams were found to have the highest correlation (.469) with AzMERIT performance, explaining 22% of the variance in AzMERIT scores. Checkpoints (.448) and Final Exams (.331) also demonstrated moderate and significant correlations with AzMERIT scores, with each Checkpoints explaining 20% of the variance in AzMERIT scores and Final Exams explaining approximately 11% of the variance.

CORRELATION BETWEEN ASSIGNMENT GROUPS AND AZMERIT SCORES BASED ON FIRST ATTEMPT

	N	MEAN	SD
All Assignments	92,664	72.10	27.68
Checkpoints	16,763	58.71	24.46
Discussions	22,576	88.95	19.03
Projects	2,591	72.56	28.48
Workbooks	45,501	69.60	28.88
Unit Exams	4,507	61.25	22.37
Final Exams	726	80.45	16.90

TABLE 5

CORRELATION BETWEEN STUDENTS' FIRST ATTEMPT SCORES AND AZMERIT SCORES

FIRST ATTEMPT			
	N ASSIGNMENTS	R	R SQUARED*
All Assignments	92,664	.268	.072
Checkpoints	16,763	.448	.200
Discussions	22,576	.152	.023
Projects	2,591	.285	.081
Workbooks	45,501	.278	.077
Unit Exams	4,507	.469	.220
Final Exams	726	.331	.109

TABLE 6

PREDICTIVE VALUE OF BEST SCORES VS FIRST ATTEMPT SCORES

Findings from this study indicated that students' first attempt scores on course assessments were better predictors of performance on the statewide assessment than students' best attempt scores. We believe this is the result of testing or practice effect reflected in students' best scores. Practice effect is any change or improvement that results from repetition of tasks or activities. In the case of

practice effect, the variable of interest may improve simply from repeating the activity (APA, 2007). Results from the current study found a stronger correlation between first attempt scores (.268) as compared to best attempt scores (.217). Additionally, best attempt scores accounted for approximately 3% less of the variance in AzMERIT scores than first attempt scores.

COMPARISON OF CORRELATIONS BETWEEN FIRST AND BEST ASSESSMENT SCORES WITH AZMERIT PERFORMANCE

	HIGHEST ATTEMPT			FIRST ATTEMPT			
	N OF ASSIGNMENTS	R	R SQUARED*	N OF ASSIGNMENTS	R	R SQUARED*	DIFFERENCE IN R SQUARED
All Assignments	100,032	.217	.047	92,664	.268	.072	-.025
Checkpoints	17,087	.345	.119	16,763	.448	.2	-.081
Discussions	23,024	.153	.024	22,576	.152	.023	.001
Projects	3,801	.267	.071	2,591	.285	.081	-.010
Workbooks	50,792	.196	.039	45,501	.278	.077	-.038
Unit Exams	4,602	.446	.199	4,507	.469	.22	-.021
Final Exams	726	.341	.116	726	.331	.109	.007

TABLE 7

DISCUSSION

In this study, we examined how well scores on StrongMind course assignments predicted student performance on a high-stakes, statewide achievement test. Results indicated student performance on StrongMind standards-based course assignments was found to be a significant predictor of end-of-year state testing performance.

Students' first attempt scores were found to be better predictors of AzMERIT performance than best attempt. Our findings provide validity evidence for utilizing the predictive value of the course assessments to inform instruction and other academic decisions.

FACILITATING MTSS AND STUDENT SUPPORT STRATEGIES

Findings from this study confirm the value of StrongMind formative assessments in facilitating data driven frameworks, such as MTSS, that promote positive student academic and behavioral outcomes. By providing valid assessments which can identify students in need of academic support, StrongMind offers partner schools the

opportunity to implement MTSS and other data driven strategies. In utilizing StrongMind standards-based assessments, educators can identify students who may be struggling to learn and intervene early, so they can perform at or above grade level.

TRIANGULATION OF DATA TO UNDERSTAND STUDENT LEARNING

Because learning occurs over time and each student acquires knowledge at a different pace, it is crucial for teachers to determine each student's current level of academic performance and make adjustments to instruction based on that information. Heacox defines the process of differentiating instruction as, "changing the pace, level, or kind of instruction you provide in response to individual learners' needs, styles, or interests" (2012, p. 5). Differentiation is a way of thinking about instruction and learning that values the individual student and is dependent upon ongoing assessment. Sources of information that provide insight into how students are learning at a given point help teachers plan the next steps in instruction (Tomlinson, 2000).

Triangulation of data, or the process of obtaining data using more than one method to collect information on the same topic, is a methodological technique that leads to a broader and deeper understanding of the subject of interest (Flick, 2009). Utilizing multiple forms of assessment facilitates instructional differentiation by providing a rich description of student learning. Educators utilizing StrongMind course assessments have the ability to view students' first and best attempt scores on each item. With this additional information teachers can make inferences about student

engagement and learning. Access to first attempt scores allows teachers to evaluate student learning and provide scaffolded feedback, enabling learners to engage in more advanced thinking and problem solving than they could without help. Scores from subsequent attempts enable teachers to evaluate learning gains and make inferences about student engagement.

Additionally, the assessments included in this study include a variety of assessment types, each providing educators with a unique source of information and the opportunity to address student learning in different ways. Utilizing StrongMind's course assessments provides teachers the ability to assess the same learning objectives in multiple ways. In addition to data from student performance on assessments integrated throughout the online course, teachers must also measure student performance based on authentic assessment opportunities even in the online environment. Authentic assessment opportunities assess student ability to efficiently and effectively use a range of knowledge and skills to negotiate a complex task (Wiggins, 1998). Taken together, the multiple forms of assessment allow educators to adjust instructional practices and address learning needs in different ways.

LIMITATIONS AND FUTURE RESEARCH

A few limitations and recommendations for future research should be noted. Although the sample for the study was large, it was limited to students in a single school. Future research would benefit from a broader sample base and should be expanded to include students from several diverse geographical areas and multiple schools. Also, the current study utilized data for high

school students only. Future research would benefit from including data for middle school students as well. In order to assess the role of formative assessments on standards mastery, future research should also include analysis at a standards mastery level rather than at an assignment grade or a test score level.

CONCLUSIONS

Assessment results are often used to make important decisions about students and schools. The use of valid assessments is vital to the process of informing instruction and academic program management. Using a student's predicted performance, along with other sources of information, educators can make informed

decisions to help students master skills and close knowledge gaps. This data can also aid administrators in academic program management and support school-wide efforts to implement data-driven frameworks of student support.

AUTHORS

Tammy McKeown, Ph.D.

Yovhane Metcalfe, Ph.D

REFERENCES

Ainsworth, L. (2007). Common Formative Assessments: The Centerpiece of an Integrated Standards-Bases Assessment System. In D.B. Reeves (Ed.), Ahead of the curve: The power of assessment to transform teaching and learning. Bloomington, IN: Solution Tree.

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). Standards for educational and psychological testing.

American Psychological Association (2007). APA Dictionary of Psychology. Washington, DC.
Bergner, Y., Colvin, K., & Pritchard, D. (2015). Estimation of ability from homework items when there are missing and/or multiple attempts. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK '15). Association for Computing Machinery, New York, NY, USA, 118–125. DOI:https://doi.org/10.1145/2723576.2723582

Doorn, D., Janssen, S. & O'Brien, M. (2010). Student Attitudes and Approaches to Online Homework. International Journal for the Scholarship of Teaching and Learning: Vol. 4: No. 1, Article 5.

Flick, U. (2009). An introduction to qualitative research (4th ed.). Sage Publications Ltd.

Heacox, D. (2012). Differentiating instruction in the regular classroom: How to reach and teach all learners. Minneapolis, MN: Free Spirit Publishing.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied Linear Statistical Models (5th ed.). McGraw-Hill Irwin.

Tomlinson, C.A. (2000). Differentiation of Instruction in the Elementary Grades. Eric Digest. https://www.ericdigests.org/2001-2/elementary.html.

Wiggins, Grant. (1998). Ensuring authentic performance. Chapter 2 in Educative Assessment: Designing Assessments to Inform and Improve Student Performance. San Francisco: Jossey-Bass, pp. 21 – 42

Ziomek-Daigle, J., Goodman-Scott, E., Cavin, J., & Donohue, P. (2016). Integrating a multi-tiered system of supports with comprehensive school counseling programs. Professional Counselor, 6(3), 220-232.