Comparative Outcomes Study of DRC BEACON Efficacy:

English Language Arts and Mathematics in Grades 4–8

Michelle Boyer, Ph.D.

October 16, 2023

ABSTRACT

This paper describes a comparative outcomes efficacy study in which achievement growth on DRC BEACON (BEACON), an interim assessment, is used as a factor explaining learning outcomes as measured by a large-scale state summative assessment. Findings show that within-year student gains on BEACON are associated with higher state summative assessment outcomes.

BACKGROUND

BEACON is an interim assessment that measures student achievement in English language arts (ELA) and mathematics (math) in grades 3–8. The assessment was developed by Data Recognition Corporation (DRC), includes multiple item types that measure college- and career-readiness standards, and is delivered primarily via the DRC INSIGHT test engine in an item-level computer-adaptive test (CAT) format that adjusts to student ability levels throughout test administration (DRC, 2020).

BEACON is typically administered three times a year, providing feedback that can help educators evaluate student learning and monitor student progress throughout the year. In addition to providing status scores and performance-level information, BEACON reports include measures of growth, Lexile and Quantile scores (refer to Appendix A for details of the concordance study), information about learning progressions, and links to instructional support materials and college- and career-ready standards that are tailored to student performance results. The interactive reporting system also offers the opportunity to disaggregate, categorize, and sort data as needed. Samples of individual student reports are provided in Appendix B.

Instructional strategies are provided through the BEACON Educator Instructional Strategies (BEIS), which include system-embedded resources and instructional supports to assist educators with coherently incorporating BEACON results and instructional guidance in classroom applications. Development of BEIS materials was informed by the work of Black and Wiliam (2004), Alonzo and Gearhart (2006), Gong (2008), Hess (2008a, 2008b), and McTighe and Wiggins (2005), and was created to align with every tested standard.

INTENDED USES OF BEACON

BEACON assessment results are intended for use in informing instruction and monitoring student progress. Specifically, student assessment results are linked to learning progressions and BEIS resources, and they provide a direct connection between student performance over time and targeted instructional supports and strategies. Educators will first review student results in the context of the defined learning progressions and will then link to BEIS resources. For example, when using BEACON reports to determine which students need support, are near target, or are prepared in terms of grade-level expectations, educators may use student score and performance-level results alongside the learning progression reports and BEIS (refer to Appendix C for examples of how this information is reported and linked) to support student and group differentiation in instruction in a way that focuses specific attention on the precisely identified learning needs.

Educators may also use BEACON results to identify areas where additional supports and enrichment activities may advantage students and their continued learning progress. Additionally, the BEIS are available by content and grade level to support educators in supplementing local programs and in planning for remediation. This approach to supporting teaching and learning is expected to positively impact student learning outcomes as it incorporates detailed evidence of the knowledge and skills that students have acquired at specific points throughout the year and provides detailed information about how to incorporate this evidence into addressing individual and group learning needs.

VALIDITY AND EFFICACY

The *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 2014) define validity as, "…the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of a test." Therefore, a strong validity argument will include:

- Clear definition(s) of test score uses and intended score interpretations.
- Identification and collection of sources of evidence that demonstrate scores are valid (or not) for those uses.
- Formal evaluation of the efficacy of the defined test uses.

"BEACON assessment content is aligned with state standards and the BEACON blueprints; is developed and reviewed through rigorous processes; and is accessible to students through adherence to universal design principles, universal tools, and assigned accommodations."

This means that the intended test score uses should not only be clearly defined, but scores must also be evaluated through an ongoing collection of evidence supporting their validity for those intended uses. The BEACON Technical Report (DRC, 2020) provides a collection of reliability and validity evidence that has been gathered in support of the use of the BEACON assessments. It is summarized below.

- **Content:** BEACON assessment content is aligned with state standards and the BEACON blueprints; is developed and reviewed through rigorous processes; and is accessible to students through adherence to universal design principles, universal tools, and assigned accommodations.
- **Relationship with other Variables:** BEACON scores show a strong relationship with the state summative assessments, indicating that assessments measure similar things in a reliable manner.
- **Response Processes:** Student responses to the BEACON items follow expected patterns: item discrimination, model fit, and omit patterns as proxies for unexpected response behaviors.
- **Internal Structure:** BEACON has strong score reliability, score scale properties, and adaptivity in terms of student experience and score accuracy.
- **Consequences:** BEACON items have been subject to rigorous bias, fairness, and sensitivity reviews, and differential item functioning has been examined.

Given the intended localized uses for BEACON to support teaching and learning and progress monitoring, it is important to collect evidence that BEACON is useful and efficacious for those purposes. In the context of consequential validity arguments, efficacy is defined as the ability to produce the desired or intended result, and it may be considered as a component of validity that focuses primarily on the consequences of test use.

RESEARCH QUESTIONS

In order to examine the efficacy of using BEACON for its intended purposes to inform instruction and monitor student learning progress, this study focuses on the relationship between within-year growth on BEACON during the 2022/2023 school year and outcomes on a state summative assessment. This is done by categorizing student growth on BEACON into quartiles and evaluating outcome differences between students who take the BEACON assessments and students who do not take the BEACON assessments.

Logically, larger within-year growth on BEACON would suggest the likelihood that more learning has occurred and that slower growth (or a decline in scores) would suggest the likelihood that either (a) learning is not progressing as well as students who achieved higher growth, or (b) students are not well-motivated to do their best on the assessment. If these assumptions are true, then it is likely that larger differences in summative scores between students who show more growth on BEACON (as distinguished from students who show less growth or who do not take the BEACON assessments) will be observed.

Therefore, the primary research question addressed in this study is: Does larger within-year growth over Fall, Winter, and Spring administrations of BEACON (as distinguished from students who show slower growth ((or a decline in score)) or who do not take the BEACON assessments) result in higher summative assessment score outcomes? The secondary research question is: How much BEACON growth is associated with significantly higher summative test scores for students?

"....the primary research question addressed in this study is: Does larger within-year growth over Fall, Winter, and Spring administrations of BEACON...?"

METHOD

This study is a comparative outcomes study in which summative test scale scores in ELA and math are compared between students grouped into quartiles (Q) of average growth on BEACON over three interim administrations during the 2022/2023 school year and students who do not take BEACON. A single factor with 5 levels is used, where the factors are: Q1 growth, Q2 growth, Q3 growth, Q4 growth, and BEACON not administered.

DATA

The data used in this study comprises student-level matched assessment results drawn from valid census data from a large state on three assessments: BEACON interim assessments (Fall, Winter, Spring administrations), the Spring 2022 state summative assessments in ELA and Math in grades 3–8, and the Spring 2023 state summative assessment in ELA and math in grades 4–8. The administration windows for BEACON were August 1, 2022–November 30, 2022; December 1, 2022–March 31, 2023; and April 1, 2022–June 30, 2023. The summative assessment administrations occurred between April 11 and May 20, 2022, and April 10 and May 19, 2023.

The number of matched students who completed all three BEACON interim assessments in grades 4–8 during the 2022/2023 school year was 17,458 in ELA and 18,659 in math. The number of students in the matched sample taking the summative assessments was 533,879 in ELA and 540,915 in math. Detailed N-counts by grade, subject, and test after matching are provided under the *Results* section of this document.

Note that full technical details regarding the measurement properties of the BEACON assessments, including score reliability, validity, and the computer-adaptive item assignment process for the BEACON assessments can be found in the BEACON Technical Report (DRC, 2020). For reference in interpreting the study results, the score ranges for the BEACON ELA and math assessments are provided in Table 1. The score ranges for the summative ELA and math assessments are noted in Table 2.

Table 1. BEACON Scale Score Ranges

Test	Grade	Minimum Scale Score	Maximum Scale Score
ELA	3	160	800
ELA	4	180	820
ELA	5	200	840
ELA	6	220	860
ELA	7	240	880
ELA	8	260	900
MATH	3	160	800
MATH	4	180	820
MATH	5	200	840
MATH	6	220	860
MATH	7	240	880
MATH	8	260	900

Table 2. Summative Assessment Scale Score Ranges

Test	Grade	Minimum Scale Score	Maximum Scale Score
ELA	3	180	830
ELA	4	210	775
ELA	5	210	760
ELA	6	140	820
ELA	7	165	785
ELA	8	225	730
MATH	3	290	705
MATH	4	270	715
MATH	5	265	725
MATH	6	285	700
MATH	7	265	740
MATH	8	275	755

ANALYSES

As it would not be appropriate to randomly assign students to experimental groups, the Spring 2022 summative test scores are used in an analysis of covariance (ANCOVA) to control for differences in the five student groups defined through a regression adjustment (Kleinbaum, Kupper, Nizam, & Muller, 2008; Kuehl, R., 2000; Pituch & Stevens, 2016). In a general linear model, the type III sums of squares for the treatment levels are corrected (adjusted) for the regression relationship. This has the effect of allowing a comparison of outcome means over the defined student groups at the mean value of the students' prior year summative score. This process effectively removes the variation due to the covariate that may otherwise be attributed to group differences.

This model is specified as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 Z + E$$

where Y represents the Spring 2023 summative assessment scores; β_0 is the model intercept and provides the point of reference from which to compare BEACON non-user outcomes on Y outcomes to each of the 4 groups of students represented by X_1 through X_4 (i.e. BEACON average growth in quartiles 1–4 respectively); Z represents the Spring 2022 summative assessment scores, which is the covariate used to adjust for group differences on the outcome and β_5 is its slope; β_1 through β_4 represent the slopes of the regression lines for each group, when Y is regressed on X, and E is the model error.

The following assumptions apply when interpreting the ANCOVA results:

- The dependent variable and covariate variables are measured on a continuous scale.
- The independent variable consists of two or more categorical, independent groups.
- There is independence of observations.
- The dependent variable is approximately normally distributed within each subpopulation.
- There is homogeneity of variances.
- The covariate is linearly related to the dependent variable at each level of the independent variable.
- There is homogeneity of regression slopes.

The homogeneity of slopes is a restrictive assumption that the regression slopes (β_1 through β_4) for each group (X_1, through X_4), when Y is regressed on Z for that group. Where this is not true, the outcome difference for that group cannot be considered equal for all values of Z, so interpretations should be adjusted accordingly.

This design is used to characterize how student learning progressed through the Fall, Winter, and Spring BEACON administrations during the 2022/2023 school year and to use this characterization to group students for the purpose of comparing their outcomes in response to the research questions posed. Note that grade 3 is not included in the ANCOVA analyses as grade 3 students do not have 2022 summative assessment scores.

RESULTS

Generally, results indicate that there are consistently larger, positive outcome differences for students who achieved more growth on BEACON during the 2022/2023 school year. Student's whose average BEACON growth in math was observed to be in the 3rd and 4th quartiles is associated with summative assessment scores that are statistically significantly higher than students who do not use BEACON. In ELA, student's whose average BEACON growth was observed to be in the 4th quartiles is associated with summative assessment scores that are statistically significantly higher than students who do not use BEACON.

Table 3 provides the scale score associated with lower bound of each growth quartile range as well as the minimum and maximum growth observed in the data.

Table 3. BEACON Observed Scale Score Growth by Quartile, 2022/2023 School Year

Content	Grade	BEACON Observed Scale Score Growth by Quartile Range (QR)					
		QR-1	QR-2	QR-3	QR-4	Min	Max
ELA	4	-69.0	0.5	16.0	32.5	-69.0	113.5
	5	-106.0	-1.5	14.0	31.0	-106.0	146.5
	6	-96.0	-0.6	10.5	27.5	-96.0	141.0
	7	-118.0	-0.9	10.5	31.0	-118.0	157.0
	8	-131.0	-8.5	14.0	38.0	-131.0	185.0
Math	4	-84.5	9.0	23.0	40.0	-84.5	181.0
	5	-84.0	9.5	25.5	44.0	-84.0	199.0
	6	-94.5	4.5	23.0	42.0	-94.5	170.0
	7	-101.5	4.0	17.5	40.0	-101.5	175.5
	8	-153.0	4.0	31.0	44.0	-153.0	220.0

Note: Shaded cells show where statistically significant higher summative scores were observed for BEACON users in Spring 2023.

Although the use of quartile ranges is a convenient and sample dependent grouping factor, the actual score ranges are meaningful for quantifying and understanding the level of BEACON growth that may yield higher scores on the outcome. As stated, the top 2 quartiles for math and the top quartile for ELA are associated with higher outcomes, but the corresponding growth ranges that are associated with significantly higher outcomes is similar between ELA and math. The growth values that result in statistically significantly higher outcomes are shaded.

This approach may be especially useful in that it supports a clear understanding of the observed BEACON scale score growth in 2022/2023 that is associated with statistically significant outcome differences that favor BEACON users, which is ultimately more meaningful than applying a simple 2-level factor of BEACON users and non-users. Table 4 provides the details of the analysis of covariance results, showing where student growth on BEACON results in statistically significant differences in outcomes.

Table 4. ANCOVA Results

Content	Gr.	Outcome Differences in Scale Score Units between BEACON Users and non-Users by BEACON Growth Quartile Range (QR)				N-Count for BEACON Users (by QR) and non-Users					η^2	R^2	Sig.
		QR-1	QR-2	QR-3	QR-4	QR-1	QR-2	QR-3	QR-4	non-Users			
ELA	4	-8.09	-4.38	-2.60	4.49	824	823	810	841	101728	0.01	0.73	<0.00
	5	-8.40	-7.37	-3.65	1.14	846	823	849	836	103585	0.01	0.74	<0.00
	6	-10.03	-4.43	-1.21	2.63	1155	677	893	909	102041	0.01	0.72	<0.00
	7	-8.31	-5.21	-4.72	2.75	868	860	874	869	102736	0.01	0.75	<0.00
	8	-2.51	-1.69	0.73	7.86	924	935	916	926	106331	0.01	0.73	<0.00
Math	4	-9.70	-4.12	0.02	9.70	875	869	843	873	102221	0.02	0.75	<0.00
	5	-6.00	-2.00	3.00	11.74	914	902	910	921	103907	0.02	0.75	<0.00
	6	-4.61	2.35	8.00	16.01	932	948	939	964	103162	0.01	0.72	<0.00
	7	-8.07	-4.67	-2.93	5.25	1299	611	956	973	104605	0.02	0.76	<0.00
	8	-1.69	3.67	8.16	14.82	993	968	979	990	108361	0.01	0.72	<0.00

Note: Shaded cells denote an outcome difference between BEACON users and non-users that is not statistically significant.

As shown in the ANCOVA results, a statistically significant amount of variance is explained by BEACON average score growth, grouped by quartiles, after controlling for other background variables that are captured by students' prior year summative assessment scores. The range of observed scale score growth on BEACON is quite wide (up to 373 points for grade 8 math) and generally wider in upper grades. However, students growing by an average of at least 27.5 scale scores in ELA and 23.0 scale scores in math tend to, on average, do statistically significantly better than non-BEACON users on the summative assessment by an average of 2.6 to 7.9 scale scores for ELA and 3.0 to 16.0 scale scores for math.

TESTING MODEL ASSUMPTIONS

The appropriateness of interpreting the differences in outcomes noted in Table 4 depends on the degree to which the slopes of their regression lines when regressing Y on Z may be considered equal. To test this assumed equality, the slopes were plotted for visual inspection and a version of the ANCOVA that included an interaction term between growth and prior summative scores was run. Although some of these interactions were statistically significant, an η^2 greater than 0.001 was not observed. Refer below to Figures 1–10 for plots of each slope. Also, the Spring 2023 summative assessment scores (outcome) and the Spring 2022 summative assessment scores (covariate) are approximately normally distributed within each group defined in the growth variable. Refer to Figures 11–20.

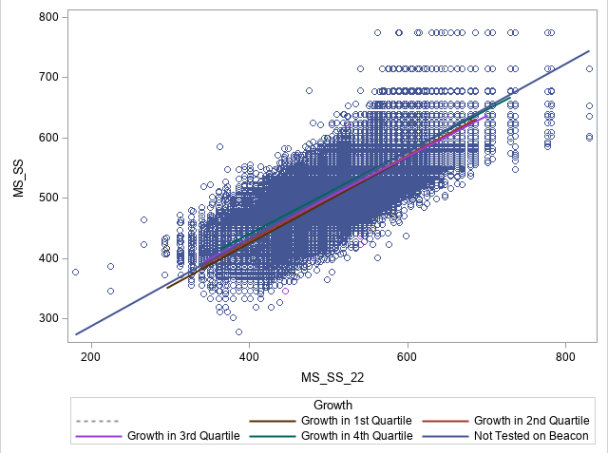

Figure 1. Test for Homogeneity of Regression Slopes – ELA Grade 4

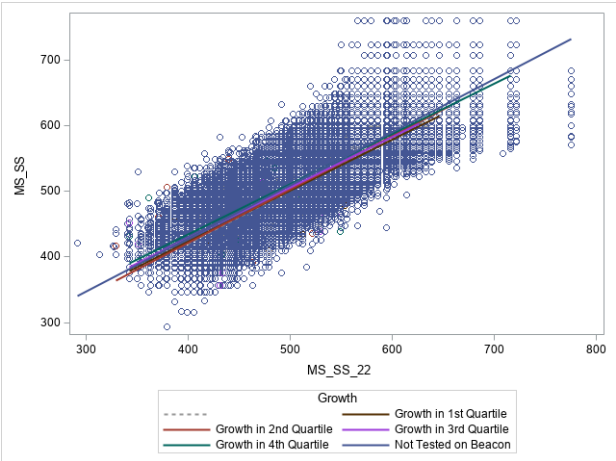

Figure 2. Test for Homogeneity of Regression Slopes – ELA Grade 5

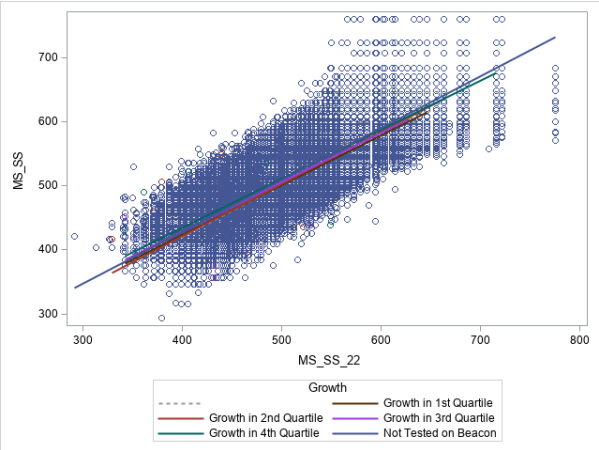

Figure 3. Test for Homogeneity of Regression Slopes – ELA Grade 6

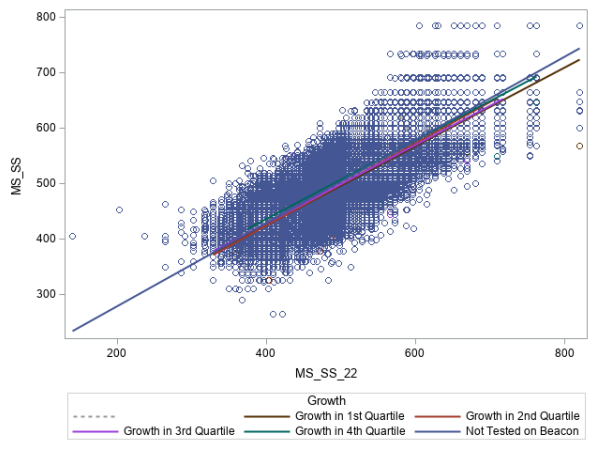

Figure 4. Test for Homogeneity of Regression Slopes – ELA Grade 7

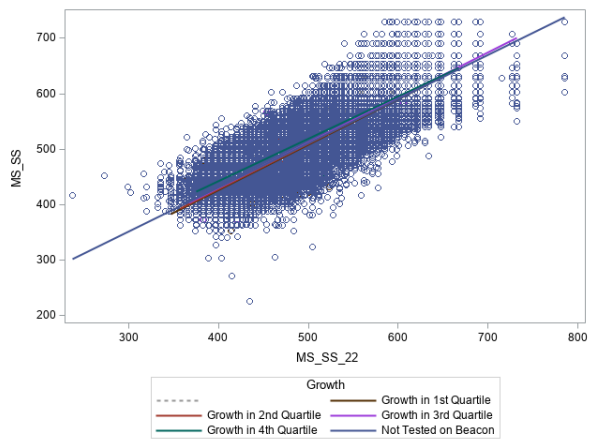

Figure 5. Test for Homogeneity of Regression Slopes – ELA Grade 8

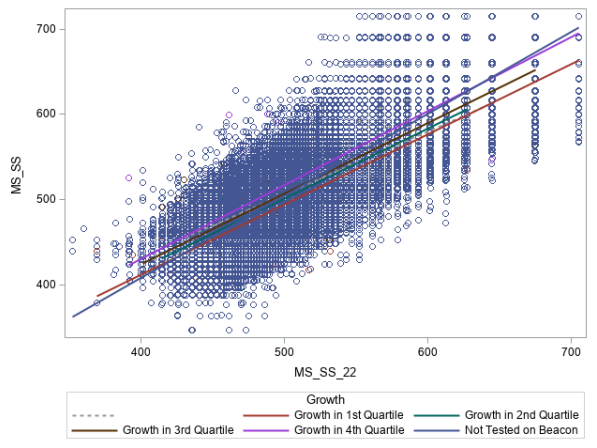

Figure 6. Test for Homogeneity of Regression Slopes – Math Grade 4

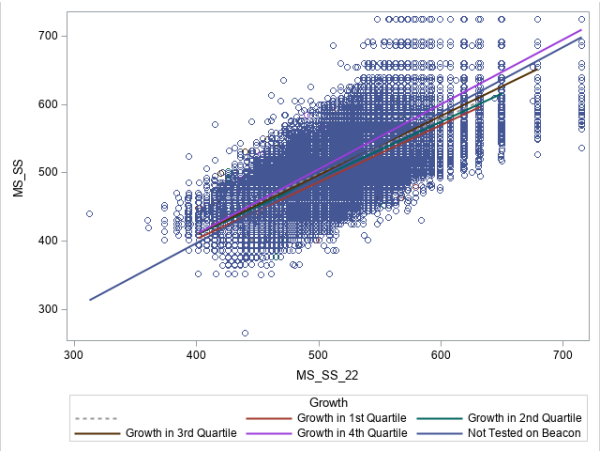

Figure 7. Test for Homogeneity of Regression Slopes – Math Grade 5

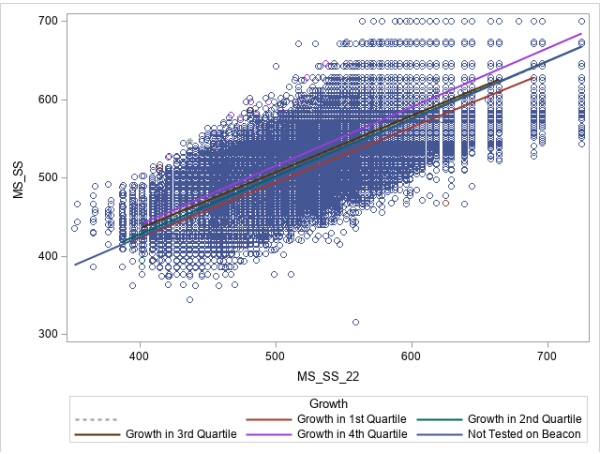

Figure 8. Test for Homogeneity of Regression Slopes – Math Grade 6

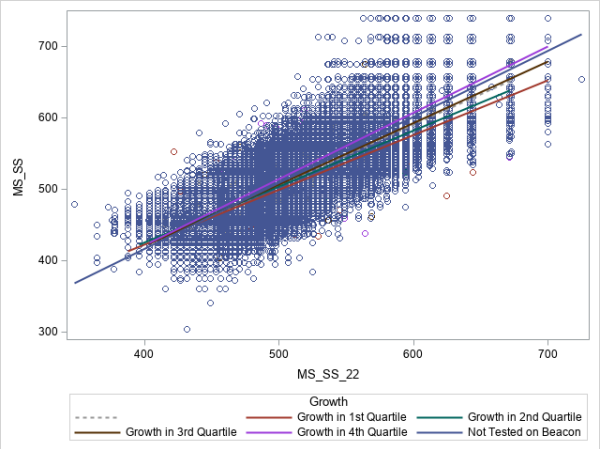

Figure 9. Test for Homogeneity of Regression Slopes – Math Grade 7

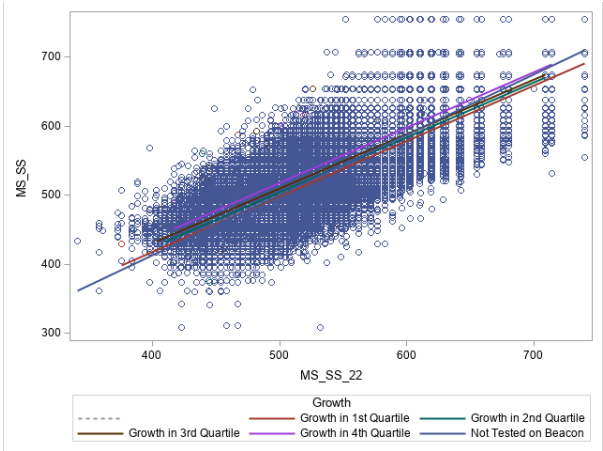

Figure 10. Test for Homogeneity of Regression Slopes – Math Grade 8

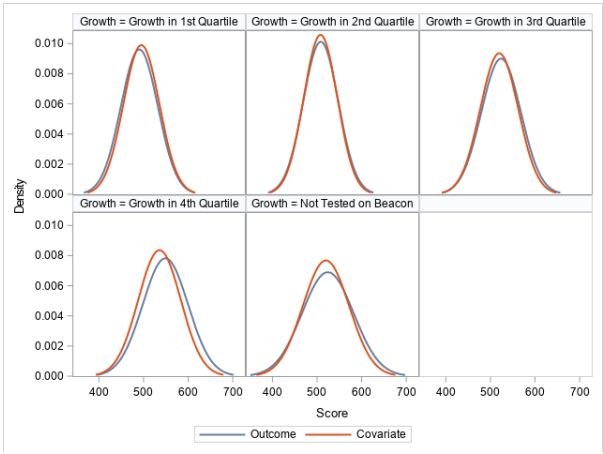

Figure 11. Density plots of Outcome and Covariate Scores by Student Growth Group, Math Grade 4

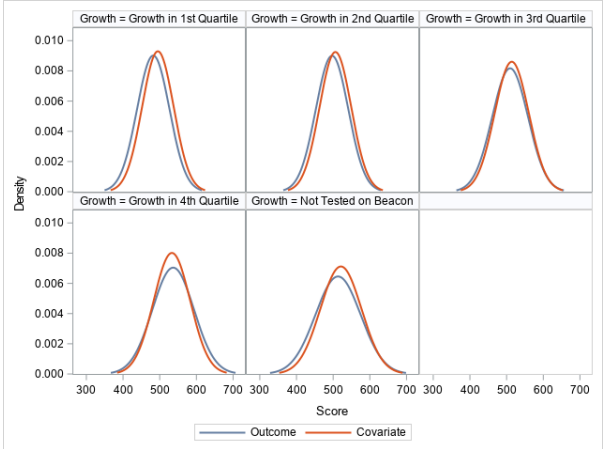

Figure 12. Density plots of Outcome and Covariate Scores by Student Growth Group, Math Grade 5

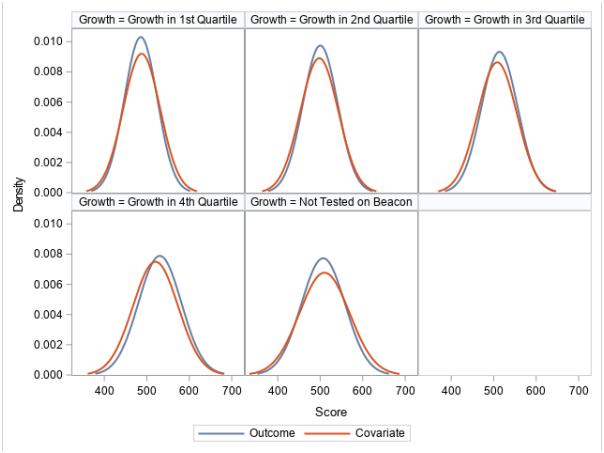

Figure 13. Density plots of Outcome and Covariate Scores by Student Growth Group, Math Grade 6

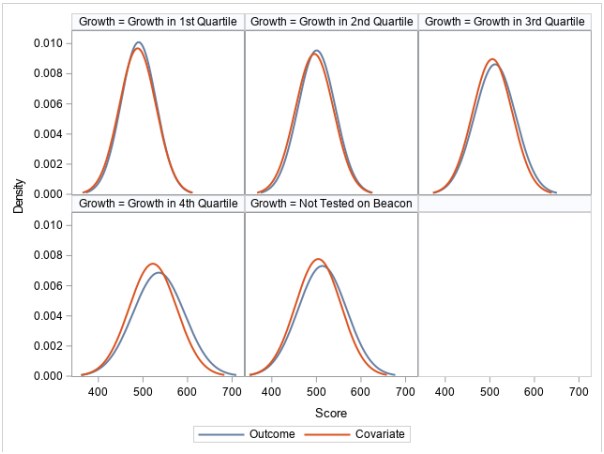

Figure 14. Density plots of Outcome and Covariate Scores by Student Growth Group, Math Grade 7

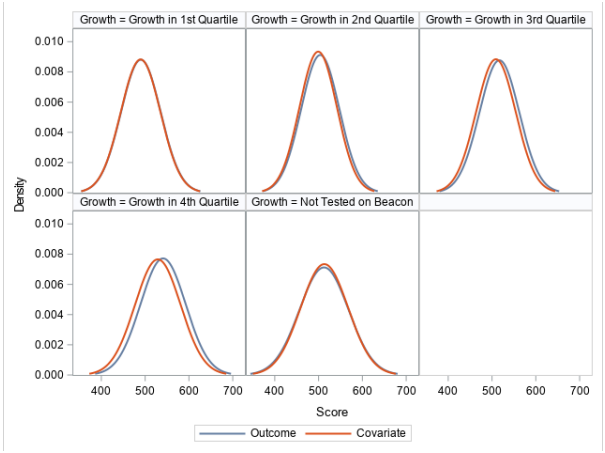

Figure 15. Density plots of Outcome and Covariate Scores by Student Growth Group, Math Grade 8

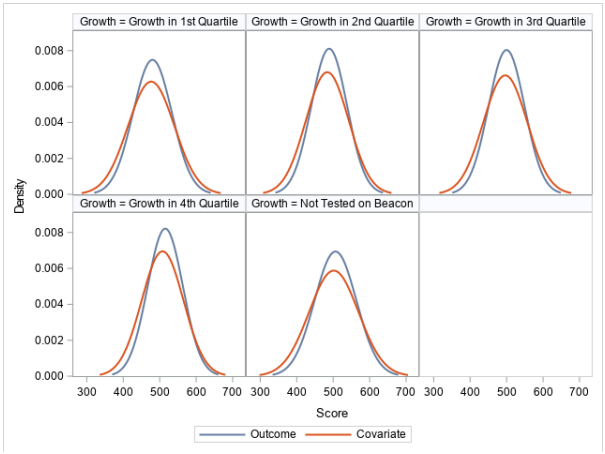

Figure 16. Density plots of Outcome and Covariate Scores by Student Growth Group, ELA Grade 4

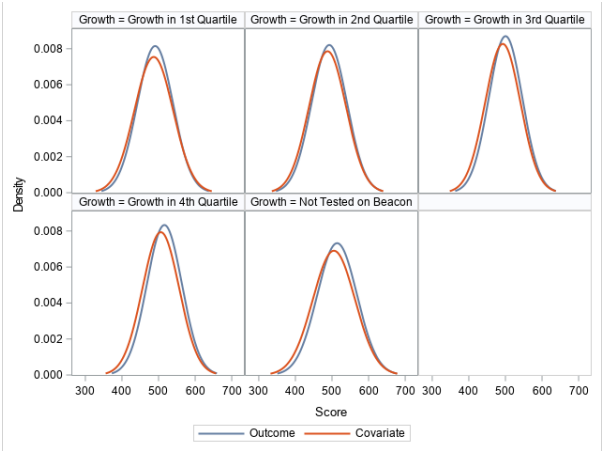

Figure 17. Density plots of Outcome and Covariate Scores by Student Growth Group, ELA Grade 5

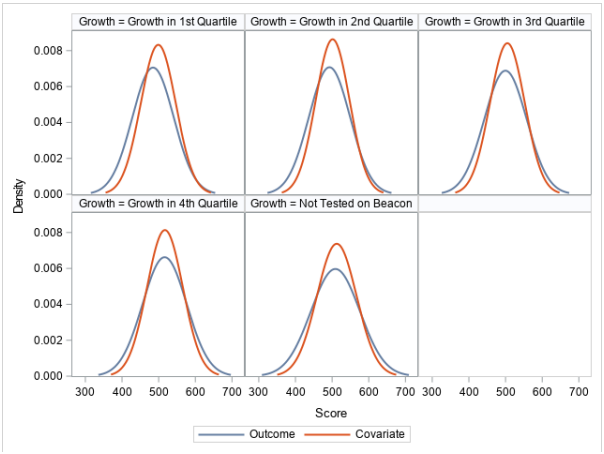

Figure 18. Density plots of Outcome and Covariate Scores by Student Growth Group, ELA Grade 6

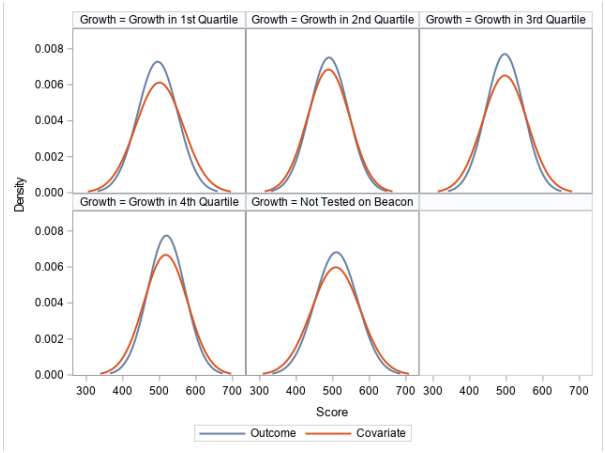

Figure 19. Density plots of Outcome and Covariate Scores by Student Growth Group, ELA Grade 7

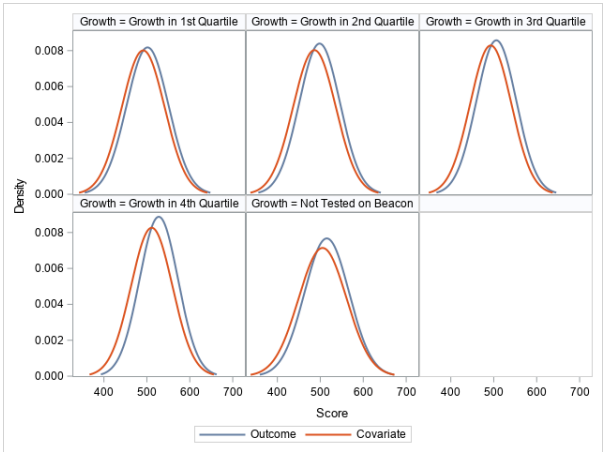

Figure 20. Density plots of Outcome and Covariate Scores by Student Growth Group, ELA Grade 8

DISCUSSION

This study offers evidence that when students are successful on BEACON as defined by their within-year growth over the three administrations (Fall, Winter, and Spring), they consistently achieve higher scores on the state summative assessment than non-users of BEACON and students who do not show growth (or show very small amounts of growth). In particular, student's whose average BEACON growth was observed to be in the 3rd and 4th quartiles in the sample used in this study had statistically significantly higher outcome scores than students who do not use BEACON. When students grew by at least 27.5 scale scores on the BEACON ELA assessment and 23.0 scale scores on the BEACON Math assessment they tended to achieve higher scores on the outcome measure that are roughly 3–16 scale score points higher than BEACON non-users.

These results provide one important piece of evidence in support of the efficacy of BEACON and its supporting materials when used to support student learning growth. An important consideration when interpreting these results is that this study cannot explain why students are not growing. There are many possible reasons for the variation in the observed growth rates, including, but not limited to, variation in the enacted curriculum differences in the instructional foci, student motivation during testing, or any other student factors that may advantage or disadvantage their performance during testing.

"This study offers evidence that when students are successful on BEACON as defined by their within-year growth over the three administrations, they consistently achieve higher scores on the state summative assessment than non-users of BEACON..."

Also, although BEACON interim assessment results provide important indications of how students are progressing in their learning, it is important, as always, to consider all sources of information that are available before drawing conclusions about the learning needs of individuals and groups of students. Also, although important and informative, summative assessment scores are generally not designed to capture information about everything that students may have learned in a subject. Summative assessments intentionally sample from the domain of what students are expected to be able to know and do, and results are most often used at aggregate levels, as is the case here, to draw more general conclusions about student learning and progress. In that way, summative assessments do provide a convenient, reliable, and standardized outcome measure to understand the effectiveness of BEACON as a tool to inform instruction and monitor progress over many students. Overall, these study results are positive and encouraging. They demonstrate that even at fairly modest levels of consistent growth on BEACON, student learning outcomes are, on average, higher for students who show these levels of growth on BEACON than for those who do not use BEACON.

REFERENCES

Alonzo, A.C. & Gearhart, M. (2006). Considering learning progressions from a classroom assessment perspective. Measurement: Interdisciplinary Research and Perspectives. Vol. 4(1&2) Mahwah, NJ: Lawrence Erlbaum. 99-108.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Black, P. & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (ed.), Towards Coherence between Classroom Assessment and Accountability. 103RF Yearbook of the National Society for the Study of Education, Part 2. Chicago, IL: National Society for the Study of Education. 20-50.

Data Recognition Corporation (2020). *BEACON Assessments Technical Report 2.0*. Research Report. Minneapolis, MN: Author.

Gong, B. (2008). Developing Learning Progressions to inform Formative Assessment: Five areas to develop. Presentation at the CCSSO FAST SCASS Meeting, February 6, 2008, Atlanta, GA. Center for Assessment

Hess, K. (2008a). Developing and Using Learning Progressions as a Schema for Measuring Progress National Center for the Improvement of Educational Assessment, Dover, New Hampshire

Hess, K. (2008b). Tools and Strategies for Developing and Using Learning Progressions. Five areas to develop. Presentation at the CCSSO FAST SCASS Meeting, February 6, 2008, Atlanta, GA. Center for Assessment

Kleinbaum, D., Kupper, L., Nizam, A., Muller, K. (2008). *Applied Regression Analysis and other Multivariable Methods, 4th Ed*. Belmont, CA: Thompson Brookes/Cole.

Kuehl, R. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis, 2nd Ed*. Pacific Grove, CA: Brookes/Cole.

McTighe, J. and Wiggins, G.P. (2005) Understanding by Design. Alexandria, VA: Association for Supervision and Curriculum and Development.

Pituch, K. & Stevens, J. (2016). *Applied Multivariate Statistics for the Social Sciences, 6th Ed*. Abingdon, Oxon: Routledge Taylor and Francis Group.

Copyright © 2024 Data Recognition Corporation