

Study of the Predictive Relationship of the Smarter Balanced Interim Comprehensive Assessment with the Smarter Balanced Adjusted Summative Assessment

Karla L. Egan, Ph.D.

Elizabeth A. Stone, Ph.D.

Joseph A. Martineau, Ph.D.

March 2026



# Table of Contents

Table of Contents .....	1
List of Tables .....	2
Executive Summary .....	3
E.1 Assessment Overview .....	3
E.2 Student Sample .....	3
E.3 Test Score Relationships .....	4
E.3 Accuracy of ICA Classifications .....	4
E.3 Partial Correlations .....	5
Section 1: Introduction.....	6
1.1 Background.....	6
1.2 Purpose of the Study .....	6
1.3 Assessment Overview.....	7
Section 2: Methods .....	7
2.1 Data Preparation.....	7
2.2 Post-stratification Weighting .....	8
2.3 Linear Regression .....	8
2.3 Classification Accuracy .....	8
2.4 Partial Correlations .....	8
Section 3: Results.....	9
3.1 Study Sample .....	9
3.1.1 Data Cleaning.....	9
3.1.2 Data Weighting .....	10
3.2 Weighted Descriptive Statistics .....	13
3.3 Regression Analyses .....	14
3.4 Classification Accuracy .....	14
3.5 Results by Season .....	15
Section 4. Partial Correlation with Time .....	17
Section 5. Conclusion .....	26
References.....	27

## List of Tables

Table 1. Number of California Students Taking the Interim Comprehensive Assessment and the Summative Assessment in the 2023-24 School Year.	3
Table 2. Attenuated and Disattenuated Correlations between Winter ICA and Smarter Balanced Summative Assessment Scores.	4
Table 3. Classification Accuracy of ICA Scores (for the Winter ICA Administration).	4
Table 4. Results of Data Filtering.	9
Table 5. Target, Unweighted, and Weighted Distribution for Demographic Variables in ELA by Grade.	11
Table 6. Target vs. Unweighted and Target vs. Weighted Distributions.	12
Table 7. Weighted Descriptive Statistics for Weighted Sample.	13
Table 8. Regression Coefficients and R <sup>2</sup> Values.	14
Table 9. Classification Accuracy Statistics.	15
Table 10. Sample Size by Season.	16
Table 11. Corrected Correlation Coefficients by Season.	16
Table 12. Classification Accuracy Statistics by Season.	17
Table 13. Weighted Partial Correlations of Days between ICA and Summative Assessment Completion and Summative Assessment Scale Score (Controlling for ICA Scale Score) by Grade and Subject.	17
Table 14. Results of Regression Analysis of Summative Assessment Scale Score as a Function of ICA Scale Score and Days between ICA and Summative Assessment Administration.	18

## Executive Summary

This report examines the predictive relationship between scores on the Smarter Balanced Interim Comprehensive Assessment (ICA) and the Smarter Balanced Summative Assessment (Summative Assessment) across English language arts (ELA) and mathematics for California students. Conducted with data from the 2023-24 school year, this study evaluates whether the ICA, an optional interim assessment, can serve as a reliable indicator of end-of-year performance on the Summative Assessment, an assessment used for accountability purposes.

The study sample included students in grades 3-8 and 11 who participated in both the ICA and the Summative Assessment. To ensure the representativeness of results, post-stratification weighting was applied, adjusting for demographic variables, including ethnicity, gender, and proficiency level. Descriptive statistics and regression analyses were conducted to explore the relationship between ICA and Summative Assessment scores, with particular focus on classification accuracy metrics to assess how well the ICA predicts Summative Assessment proficiency levels.

### E.1 Assessment Overview

California administers the Smarter Balanced Summative Assessment to students in grades 3-8 and 11 as part of its California Assessment of Student Performance and Progress (CAASPP) System (California Department of Education, 2024). The ICA is not required by the California Department of Education; however, California districts may administer the ICA as part of their balanced assessment systems. The potential for selection bias was addressed by post-stratification weighting of the observed sample to closely represent the demographic distributions on the summative assessment. The ICA is based on the full Smarter Balanced blueprint while the CAASPP version of the Summative Assessment is based on the Smarter Balanced adjusted blueprint. The ICA is reported on the same scale as the Summative Assessment.

### E.2 Student Sample

Students who took the ICA during the 2023-24 school year and the Summative Assessment in spring 2024 are included in this study. Table 1 reports the number of students included in the study. This sample is based only on those schools that participated in the ICA and that made their data available to Smarter Balanced. Because of the potential for selection bias in administering the ICA, the data were weighted in post-stratification to reflect the overall student population in terms of gender, race/ethnicity, IEP status, LEP status, economic disadvantage status, migrant status, and performance level on the Summative Assessment.

*Table 1. Number of California Students Taking the Interim Comprehensive Assessment and the Summative Assessment in the 2023-24 School Year.*

Content Area	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
ELA	10,681	11,646	11,757	8,403	10,503	11,244	10,182

Content Area	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Math	14,860	14,796	15,589	14,225	12,301	12,866	12,938

### E.3 Test Score Relationships

Table 2 shows the attenuated and disattenuated correlations between the ICA and the Summative Assessment scores for the winter administration of the ICA. These disattenuated correlations range from 0.82 to 0.89. Correlations for the early fall, late fall, and spring ICA administrations were similar. These findings suggest that ICA scores can serve as robust indicators of student proficiency as measured by the Summative Assessment.

*Table 2. Attenuated and Disattenuated Correlations between Winter ICA and Smarter Balanced Summative Assessment Scores.*

Content Area	ICA/Summative Correlation	Disattenuated Correlation
ELA	0.78	0.85
ELA	0.80	0.87
ELA	0.81	0.88
ELA	0.78	0.85
ELA	0.80	0.87
ELA	0.81	0.87
ELA	0.82	0.88
Math	0.78	0.82
Math	0.84	0.89
Math	0.79	0.86
Math	0.79	0.86
Math	0.81	0.89
Math	0.80	0.88
Math	0.78	0.86

### E.3 Accuracy of ICA Classifications

Table 3 shows the classification accuracy statistics for the proportion of students accurately classified as proficient on the Summative Assessment given their winter ICA score. The accuracy rates range from 0.82 to 0.86 across content areas and grades indicating that the ICA is particularly effective in identifying students likely to achieve proficiency. Accuracy rates for early fall, late fall, and spring ICA administrations were similar.

*Table 3. Classification Accuracy of ICA Scores (for the Winter ICA Administration).*

Grade	ELA		Math	
	ELA ICA Threshold	Classification Accuracy	Math ICA Threshold	Classification Accuracy
3	2393	0.815	2413	0.816
4	2435	0.834	2460	0.847
5	2474	0.842	2489	0.835

<b>Grade</b>	<b>ELA ICA Threshold</b>	<b>ELA Classification Accuracy</b>	<b>Math ICA Threshold</b>	<b>Math Classification Accuracy</b>
6	2516	0.823	2508	0.829
7	2545	0.817	2538	0.862
8	2566	0.835	2547	0.843
11	2588	0.835	2607	0.844

### E.3 Partial Correlations

In addition, a question of interest was whether there is a relationship between Summative Assessment scale scores and the days between ICA and Summative Assessment administration after controlling for ICA scale scores. Positive partial correlations were found between Summative Assessment scale scores and days between ICA and Summative Assessment administration in all grades except grade 11 (where partial correlations were near zero). This indicates that, at least in grades 3-8, as more time elapses between ICA and Summative Assessment administration, more learning tends to occur between ICA and Summative Assessment administration, as would be expected.

# Section 1: Introduction

## 1.1 Background

The study examined the predictive relationship between scores on the Smarter Balanced Interim Comprehensive Assessment (ICA) and scores on the Smarter Balanced Summative Assessment (Summative Assessment) across English language arts (ELA) and mathematics for California students. Conducted with data from the 2023-24 school year, the study sample included students in grades 3-8 and 11 who participated in both the ICA and the Summative Assessments.

A predictive study examines the claim that scores from an earlier administration predict the scores on a subsequent administration. The *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA) et al., 2014) stress the importance of conducting predictive studies when assessment scores are used for consequential purposes. Also, the Council of Chief State School Officers and the Association of Test Publishers (Carlson, 2013), in their discussion of operational best practices for test development, call for evaluation studies whenever claims are made about test scores that classify students into groups. When educational assessments are used for predictive purposes, the accuracy of this classification of test takers into achievement levels is of particular importance, especially when used for school accountability purposes.

Predictive inferences made about test scores must be supported by evidence in order to satisfy a validity argument for the intended uses and interpretations of those test scores (AERA et al., 2014). This evidence must address the accuracy of the predictions and must examine any propositions made about intended score inferences, including generalizability. Such studies should examine the score relationships using correlation and linear regression analyses, as well as evaluate classification accuracy and misclassification rates.

By evaluating the ICA scores' prediction of Summative Assessment scores, this study examined the strength of the predictive relationship between two different, but related, assessments. We would expect a strong, but not perfect, correlation between scores on the two assessments for at least two reasons. First, the two assessments were developed using related, but not identical, test specifications. Second, the interim nature of the ICA assumes that teachers will use the assessment scores to gauge progress and adjust instruction with the goal of influencing student performance on the later Summative Assessment. When the instructional interventions based on ICA results improve the performance of initially low-scoring students over the course of the school year, "the predictive relationships will decrease to the extent that student learning is improved" (AERA et al., 2014, p. 184).

## 1.2 Purpose of the Study

The purpose of the study was to examine the predictive relationship between scores on the ICA and scores on the Summative Assessment across ELA and mathematics for California students in grades 3-8 and 11. The research question guiding the study was,

*For the student population, can the ICA serve as a reliable indicator of end-of-year performance on the Summative Assessment?*

### 1.3 Assessment Overview

California administers the Smarter Balanced Summative Assessment to students in grades 3-8 and 11 as part of its California Assessment of Student Performance and Progress (CAASPP) System (California Department of Education, 2024). The ICA is not required by the California Department of Education; however, California districts may administer the ICA as part of their balanced assessment systems. The ICA is based on the full Smarter Balanced blueprint while the CAASPP Summative Assessment is based on the Smarter Balanced adjusted blueprint. The ICA is reported on the same scale as the Summative Assessment.

The ICA and the Summative Assessment are two components of the Smarter Balanced Assessment System. Both assessments use items from the same item bank, are built to similar content specifications, measure the same depth of knowledge and cognitive complexity, and are reported on the same scale. Despite these similarities, the assessments differ in several important ways.

First, the ICA is based on the full Smarter Balanced test blueprint, while the Summative Assessment follows an adjusted blueprint for CAASPP, making the ICA the longer test of the two. Additionally, the Summative Assessment is administered as a computer-adaptive test (CAT), adapting to a student's ability level as they progress, whereas the ICA is a fixed-form assessment. Because there is only one version of the ICA, students taking the ICA multiple times during a school year will encounter the same form each time.

Furthermore, while the Summative Assessment is mandatory for all California students and is administered under standardized conditions for state and federal accountability purposes, the ICA is optional and may be administered at the discretion of districts. The ICA is often used by classroom teachers as a practice tool for the Summative Assessment and as a means of gathering information that may guide instructional decisions.

## Section 2: Methods

### 2.1 Data Preparation

Data were provided by the Smarter Balanced Assessment Consortium and are from the 2023-24 school year. All students who had at least one ICA scale score and a Summative Assessment scale score from Smarter Balanced were initially included in analyses unless the Summative Assessment took place prior to the ICA. These students' scores were eliminated from the analyses because the study focused on the predictive value of the ICA. When students had more than one ICA score in the same subject area and school year, only the first score was retained.

**Identification and Removal of Outliers.** Because extreme values can affect means, correlations, and other statistics, guidelines for removing outliers were followed and were applied to both ICA and Summative scale scores. This procedure involved calculating the interquartile range (IQR), representing the difference between the 75th percentile (Q3) and the

25th percentile (Q1) of the score range. Students with scores greater than  $Q3 + 1.5 * IQR$  or less than  $Q1 - 1.5 * IQR$  were removed to create this dataset.

In addition to removal of outliers, two other conditions resulted in removal of cases from the dataset: 1) students identifying as gender nonbinary, and 2) students for whom the race/ethnicity value was null. These cases were removed because the population distributions for the Summative Assessment (the target against which post-stratification weighting was conducted) did not contain a nonbinary gender category or a null race/ethnicity category.

## 2.2 Post-stratification Weighting

Because of the non-negligible likelihood of selection bias posed by schools/LEAs self-selecting into the ICA, post-stratification weighting was used to adjust the sample to be more representative of the students in California in terms of gender, race/ethnicity, IEP status, LEP status, economic disadvantage status, migrant status, and Summative Assessment performance level. Marginal population proportions of these variables were derived from demographic summary tables in Chapter 7 (Appendices) of the *CAASPP Smarter Balanced Summative Assessment 2022–23 Technical Report* (California Department of Education, 2024) and reflect the most recently published data. The *anesrake* package in R achieves weighting through iterative “raking” across the same variables in the sample dataset until convergence is reached and the marginal proportions of the sample data match those of the targets as closely as possible. Weights from this procedure were capped at a maximum of 3, following standard guidelines. Note that the sample sizes remain the same after weighting, as opposed to other methods that reduce sampling of some groups to achieve representative proportions. Instead, students are assigned weights that come into play in subsequent analyses.

## 2.3 Linear Regression

After weighted regressions were run to determine the intercept and slope of the line relating Summative Assessment scale score to ICA scale score, a predicted Summative Assessment scale score was derived for each student using this linear equation. Actual and predicted scale scores were classified into the four achievement levels using the published cut scores for Smarter Balanced assessments. These levels were then binned into “Proficient” (Standard Exceeded, Standard Met) and “Not Proficient” (Standard Nearly Met, Standard Not Met).

## 2.3 Classification Accuracy

To evaluate the ICA’s accuracy in predicting students’ performance on the Summative Assessment, several performance metrics were calculated: accuracy, false-positive rate, false-negative rate, specificity, precision, and area under the receiver operating characteristic curve (AUC). These analyses utilized the *WeightedROC* package in R. Specifically, the ICA’s accuracy in predicting student Summative Assessment scores at or above proficiency was evaluated.

## 2.4 Partial Correlations

In addition, a key question was whether there is a partial correlation of Summative Assessment scale scores and days between ICA and Summative Assessment administration after controlling for ICA scale score (where a positive partial correlation would indicate that the more days there are between interim and summative scale scores, the more learning can occur before the Summative Assessment). To evaluate this possibility, weighted partial correlations were

calculated and a weighted linear regression of Summative Assessment scale scores on ICA scale scores and days between ICA and Summative Assessment administration was conducted.

## Section 3: Results

### 3.1 Study Sample

The sample is based on the 2023-24 administration of the ICA and the spring 2024 administration of the Summative Assessment. Sample demographic characteristics for unweighted and weighted distributions for ELA and mathematics were examined by subgroup and by achievement levels.

#### 3.1.1 Data Cleaning

In data cleaning, the data were filtered in three ways. First, cases with null race/ethnicity values were removed as null race/ethnicity is not a possible category in the reference distributions from Smarter Balanced technical reports. Second, cases with a gender value of nonbinary were removed as there were no nonbinary gender values in the technical report reference distributions. Third, outliers were removed. Outliers were defined for each grade and subject as a scale score value less than the 25<sup>th</sup> percentile minus 1.5 times the interquartile range or a scale score value greater than the 75<sup>th</sup> percentile plus 1.5 times the interquartile range. Both Summative Assessment and ICA scale scores were evaluated in identifying outliers. Table 4 summarizes the results of data filtering, with counts of cases removed by each type of filtering, the total number of filtered cases, and the percent of cases that were filtered. As this table shows, between 0.8% and 2.8% of cases were filtered out, depending on content area and grade.

*Table 4. Results of Data Filtering.*

Content Area	Grade	N Before Filtering	N outliers filtered	N nonbinary filtered	N null ethnicity filtered	Total N filtered	N After Filtering	Percent of Cases Filtered
ELA	3	10,858	90	0	87	177	10,681	1.6%
ELA	4	11,847	101	2	98	201	11,646	1.7%
ELA	5	11,852	1	1	93	95	11,757	0.8%
ELA	6	8,507	25	3	76	104	8,403	1.2%
ELA	7	10,646	0	5	138	143	10,503	1.3%
ELA	8	11,413	14	3	152	169	11,244	1.5%
ELA	11	10,363	97	11	73	181	10,182	1.7%
Math	3	15,192	210	0	122	332	14,860	2.2%
Math	4	15,217	299	3	119	421	14,796	2.8%
Math	5	15,905	199	1	116	316	15,589	2.0%
Math	6	14,465	46	4	190	240	14,225	1.7%
Math	7	12,543	104	6	132	242	12,301	1.9%
Math	8	13,100	88	2	144	234	12,866	1.8%
Math	11	13,151	100	14	99	213	12,938	1.6%

### 3.1.2 Data Weighting

Table 5 shows, for ELA and mathematics, the target, unweighted, and weighted distributions of all demographic variables in the dataset. Table 6 shows the discrepancies between the target and unweighted distributions and the target and weighted distributions using a heat map with darker shades indicating larger discrepancies. These tables show that Asian, Black, multiracial, Filipino, White, and LEP students are underrepresented in the unweighted distributions, and that Hispanic students, students with IEPs, and economically disadvantaged students are overrepresented in the unweighted distributions. These tables also show that the post-stratification weighting procedure performed exceptionally well in all cases except for Grade 6 ELA where White students are underrepresented in the weighted sample by 6.5% and Hispanic or Latino students are overrepresented by 5.2% in the weighted sample. The Grade 6 ELA weighted sample was used in analyses in spite of these modest under- and over-representations because the resulting distributions are as close to the target distributions as was possible to achieve.

Table 5. Target, Unweighted, and Weighted Distribution for Demographic Variables in ELA by Grade.

Content Area	Group*	Target Proportion							Unweighted Proportion							Weighted Proportion						
		3	4	5	6	7	8	11	3	4	5	6	7	8	11	3	4	5	6	7	8	11
ELA	female	.491	.490	.490	.488	.488	.485	.489	.490	.496	.493	.485	.488	.481	.486	.491	.490	.490	.488	.488	.485	.489
ELA	amind	.004	.004	.004	.004	.004	.004	.004	.002	.002	.002	.003	.002	.002	.002	.004	.004	.004	.005	.004	.004	.004
ELA	asian	.100	.104	.102	.096	.097	.097	.098	.059	.053	.056	.037	.034	.039	.036	.100	.104	.102	.112	.097	.097	.098
ELA	black	.049	.050	.050	.051	.051	.050	.048	.059	.060	.059	.055	.067	.070	.061	.049	.050	.050	.064	.051	.050	.048
ELA	multi	.061	.058	.056	.054	.052	.050	.044	.025	.024	.023	.010	.018	.017	.021	.061	.058	.056	.030	.052	.050	.044
ELA	filip	.021	.022	.022	.023	.023	.024	.027	.012	.016	.013	.010	.011	.013	.012	.021	.022	.022	.029	.023	.024	.027
ELA	hisp	.556	.556	.560	.566	.566	.568	.567	.752	.752	.751	.837	.777	.764	.728	.556	.555	.560	.618	.566	.568	.567
ELA	hipac	.004	.004	.004	.004	.004	.004	.004	.003	.003	.003	.003	.003	.003	.002	.004	.004	.004	.005	.004	.004	.004
ELA	white	.205	.202	.202	.202	.203	.203	.208	.087	.090	.093	.046	.086	.091	.139	.205	.202	.202	.137	.203	.203	.208
ELA	iep	.126	.129	.128	.126	.122	.121	.107	.117	.118	.119	.111	.109	.109	.102	.126	.129	.128	.122	.122	.121	.107
ELA	lep	.228	.214	.197	.166	.145	.133	.105	.312	.287	.249	.305	.207	.178	.109	.228	.214	.197	.178	.145	.133	.105
ELA	ecd	.628	.632	.633	.631	.630	.628	.608	.810	.804	.797	.874	.812	.814	.773	.628	.632	.633	.634	.630	.628	.608
ELA	mig	.008	.007	.008	.008	.007	.008	.007	.017	.015	.014	.017	.012	.013	.003	.008	.007	.008	.009	.007	.008	.007
ELA	PL1	.350	.370	.330	.310	.300	.300	.230	.355	.396	.350	.360	.319	.290	.156	.350	.370	.330	.313	.294	.294	.235
ELA	PL2	.220	.190	.200	.250	.230	.250	.210	.243	.202	.205	.270	.235	.272	.204	.220	.190	.200	.248	.227	.247	.213
ELA	PL3	.200	.200	.250	.270	.310	.300	.290	.191	.201	.253	.264	.307	.309	.336	.200	.200	.250	.268	.309	.299	.292
ELA	PL4	.230	.240	.220	.170	.170	.160	.260	.211	.202	.191	.107	.139	.129	.304	.230	.240	.220	.171	.170	.160	.260
Math	female	.491	.490	.490	.488	.488	.485	.489	.489	.498	.494	.488	.486	.482	.496	.491	.490	.490	.488	.488	.484	.489
Math	amind	.004	.004	.004	.004	.004	.004	.004	.002	.002	.002	.003	.002	.002	.003	.004	.004	.004	.004	.004	.004	.004
Math	asian	.101	.105	.103	.097	.098	.098	.098	.050	.046	.045	.033	.033	.042	.033	.101	.105	.103	.097	.098	.098	.098
Math	black	.048	.050	.050	.051	.050	.050	.048	.044	.046	.048	.062	.061	.064	.056	.048	.050	.050	.051	.050	.050	.048
Math	multi	.061	.058	.056	.053	.051	.050	.044	.024	.022	.021	.018	.017	.018	.019	.061	.058	.056	.053	.051	.050	.044
Math	filip	.021	.022	.022	.022	.023	.024	.027	.011	.015	.011	.011	.011	.015	.012	.021	.022	.022	.022	.023	.024	.027
Math	hisp	.556	.555	.560	.567	.567	.568	.568	.783	.784	.789	.790	.785	.754	.773	.556	.555	.560	.567	.567	.568	.568
Math	hipac	.004	.004	.004	.004	.004	.004	.004	.002	.003	.002	.003	.003	.003	.002	.004	.004	.004	.004	.004	.004	.004
Math	white	.205	.202	.202	.202	.202	.202	.207	.084	.082	.083	.081	.086	.101	.103	.205	.202	.202	.202	.202	.202	.207
Math	iep	.125	.128	.128	.125	.121	.120	.106	.117	.115	.115	.113	.116	.109	.099	.124	.128	.127	.125	.121	.120	.106
Math	lep	.235	.220	.203	.172	.150	.137	.107	.341	.313	.272	.267	.222	.180	.109	.235	.220	.203	.172	.150	.137	.107
Math	ecd	.627	.631	.632	.631	.629	.628	.608	.813	.815	.814	.831	.823	.805	.790	.627	.631	.632	.631	.629	.628	.608
Math	mig	.008	.007	.008	.008	.007	.008	.007	.022	.021	.021	.020	.014	.014	.004	.008	.007	.008	.008	.007	.008	.007
Math	PL1	.330	.310	.410	.410	.420	.480	.510	.321	.315	.419	.453	.442	.453	.464	.330	.310	.417	.410	.420	.480	.518
Math	PL2	.220	.280	.250	.260	.250	.220	.210	.226	.314	.268	.268	.261	.236	.261	.220	.280	.252	.260	.250	.220	.212
Math	PL3	.250	.220	.150	.160	.160	.130	.150	.257	.223	.162	.156	.169	.150	.184	.250	.220	.151	.160	.160	.130	.151
Math	PL4	.200	.190	.180	.170	.170	.170	.120	.195	.148	.151	.123	.129	.161	.091	.200	.190	.180	.170	.170	.170	.120

\* amind = American Indian or Alaska Native, black = Black or African American, multi = Two or more races, filip = Filipino, hisp = Hispanic or Latino, hipac = Native Hawaiian or Pacific Islander, iep = has an individualized education plan, lep = limited English proficiency, ecd = economically disadvantaged, mig = migrant status, PL1 = summative score in performance level 1, PL2 = summative score in performance level 2, PL3 = summative score in performance level 3, PL4 = summative score in performance level 4.

Table 6. Target vs. Unweighted and Target vs. Weighted Distributions.

Content Area	Group*	Unweighted - Target   Proportions							Weighted - Target   Proportions						
		3	4	5	6	7	8	11	3	4	5	6	7	8	11
ELA	female	.001	.006	.003	.004	.000	.004	.003	.000	.000	.000	.000	.000	.000	.000
ELA	amind	.002	.002	.002	.001	.002	.002	.002	.000	.000	.000	.001	.000	.000	.000
ELA	asian	.041	.051	.046	.059	.063	.058	.062	.000	.000	.000	.016	.000	.000	.000
ELA	black	.011	.010	.009	.004	.016	.019	.013	.000	.000	.000	.013	.000	.000	.000
ELA	multi	.036	.035	.032	.044	.033	.033	.023	.000	.000	.000	.024	.000	.000	.000
ELA	filip	.009	.006	.010	.013	.012	.011	.016	.000	.000	.000	.006	.000	.000	.000
ELA	hisp	.197	.196	.191	.271	.211	.197	.161	.000	.000	.000	.052	.000	.000	.000
ELA	hipac	.001	.001	.002	.001	.001	.002	.003	.000	.000	.000	.000	.000	.000	.000
ELA	white	.118	.112	.109	.157	.116	.112	.069	.000	.000	.000	.065	.000	.000	.000
ELA	iep	.009	.012	.010	.015	.013	.011	.005	.000	.000	.000	.004	.000	.000	.000
ELA	lep	.084	.073	.052	.139	.062	.044	.004	.000	.000	.000	.012	.000	.000	.000
ELA	ecd	.182	.172	.164	.242	.182	.186	.166	.000	.000	.000	.003	.000	.000	.000
ELA	mig	.010	.008	.006	.009	.005	.005	.004	.000	.000	.000	.001	.000	.000	.000
ELA	PL1	.005	.026	.020	.050	.019	.010	.074	.000	.000	.000	.003	.006	.006	.005
ELA	PL2	.023	.012	.005	.020	.005	.022	.006	.000	.000	.000	.002	.003	.003	.003
ELA	PL3	.009	.001	.003	.006	.003	.009	.046	.000	.000	.000	.002	.001	.001	.002
ELA	PL4	.019	.038	.029	.063	.031	.031	.044	.000	.000	.000	.001	.000	.000	.000
Math	female	.002	.008	.004	.001	.002	.003	.007	.000	.000	.000	.000	.000	.000	.000
Math	amind	.002	.002	.002	.002	.002	.002	.002	.000	.000	.000	.000	.000	.000	.000
Math	asian	.052	.059	.058	.064	.064	.055	.065	.000	.000	.000	.000	.000	.000	.000
Math	black	.004	.004	.002	.011	.010	.014	.009	.000	.000	.000	.000	.000	.000	.000
Math	multi	.037	.036	.034	.035	.034	.031	.025	.000	.000	.000	.000	.000	.000	.000
Math	filip	.010	.007	.011	.012	.012	.009	.016	.000	.000	.000	.000	.000	.000	.000
Math	hisp	.228	.229	.229	.223	.219	.186	.205	.000	.000	.000	.000	.000	.000	.000
Math	hipac	.002	.001	.002	.001	.001	.001	.002	.000	.000	.000	.000	.000	.000	.000
Math	white	.121	.120	.119	.120	.116	.101	.104	.000	.000	.000	.000	.000	.000	.000
Math	iep	.007	.013	.013	.012	.005	.011	.007	.000	.000	.000	.000	.000	.000	.000
Math	lep	.106	.092	.069	.096	.072	.042	.002	.000	.000	.000	.000	.000	.000	.000
Math	ecd	.186	.184	.182	.200	.193	.177	.182	.000	.000	.000	.000	.000	.000	.000
Math	mig	.014	.013	.013	.012	.006	.007	.003	.000	.000	.000	.000	.000	.000	.000
Math	PL1	.009	.005	.009	.043	.022	.027	.046	.000	.000	.007	.000	.000	.000	.008
Math	PL2	.006	.034	.018	.008	.011	.016	.051	.000	.000	.002	.000	.000	.000	.002
Math	PL3	.007	.003	.012	.004	.009	.020	.034	.000	.000	.001	.000	.000	.000	.001
Math	PL4	.005	.042	.029	.047	.041	.009	.029	.000	.000	.000	.000	.000	.000	.000

\* amind = American Indian or Alaska Native, black = Black or African American, multi = Two or more races, filip = Filipino, hisp = Hispanic or Latino, hipac = Native Hawaiian or Pacific Islander, iep = has an individualized education plan, lep = limited English proficiency, ecd = economically disadvantaged, mig = migrant status, PL1 = summative score in performance level 1, PL2 = summative score in performance level 2, PL3 = summative score in performance level 3, PL4 = summative score in performance level 4.

### 3.2 Weighted Descriptive Statistics

This section provides weighted descriptive statistics for the study sample across grade levels and content areas. Table 7 shows the mean scale scores for the ICA and the Summative Assessment, the reliability for both assessments, the (weighted) correlation between the ICA and Summative Assessment scores, and the (weighted) correlation coefficient after correcting for attenuation. For ELA, the mean scores for the ICA and the Summative Assessment show a steady increase across grade levels, with average scores progressing from 2395 in grade 3 to 2597 in grade 11 on the ICA, and from 2412 to 2596 on the Summative Assessment. The similarity in progression between ICA and Summative Assessment scores reflects expected growth in student proficiency over time, and the consistency between ICA and Summative Assessment mean scores suggests that the ICA may provide a reasonably accurate snapshot of student achievement as measured by the Summative Assessment. The reliability of both tests is in the 0.90 range, indicating a reliable assessment. The correlation results show a strong relationship between the ICA and Summative Assessment scores.

In mathematics, mean ICA and Summative Assessment scores similarly increase with grade level, progressing from 2403 in grade 3 to 2559 in grade 11 for the ICA, and from 2424 in grade 3 to 2551 in grade 11 for the Summative Assessment. As with ELA, the close relationship of the mathematics ICA and Summative Assessment mean scores across grades indicates that the ICA captures meaningful patterns of achievement over time, tracking closely with the Summative Assessment. Weighted disattenuated correlations for mathematics scores are similarly strong, ranging from 0.82 to 0.89, confirming the close relationship between ICA and Summative Assessment measures. The reliability of both the mathematics ICA and Summative Assessment is also strong, with coefficients at or above 0.90, suggesting consistent and stable measurements. The correlation coefficients indicate a strong relationship between the ICA and Summative Assessment mathematics scores.

*Table 7. Weighted Descriptive Statistics for Weighted Sample.*

Content Area	Grade	N	ICA		ICA Reliability	Summative Reliability	ICA/Summative Correlation	Disattenuated Correlation
			Mean SS	Summative Mean SS				
ELA	3	10,681	2394.7	2411.6	0.92	0.92	0.78	0.85
ELA	4	11,646	2428.9	2453.6	0.92	0.92	0.80	0.87
ELA	5	11,757	2465.3	2492.2	0.92	0.93	0.81	0.88
ELA	6	8,403	2504.9	2510.4	0.91	0.93	0.78	0.85
ELA	7	10,503	2538.7	2540.4	0.91	0.93	0.80	0.87
ELA	8	11,244	2554.0	2551.4	0.91	0.94	0.81	0.87
ELA	11	10,182	2596.6	2596.4	0.91	0.94	0.82	0.88
Math	3	14,860	2403.4	2424.4	0.95	0.94	0.78	0.82
Math	4	14,796	2446.4	2463.5	0.95	0.93	0.84	0.89
Math	5	15,589	2469.3	2481.6	0.93	0.92	0.79	0.86
Math	6	14,225	2482.6	2497.8	0.92	0.92	0.79	0.86
Math	7	12,301	2508.2	2513.7	0.91	0.92	0.81	0.89
Math	8	12,866	2516.4	2521.2	0.90	0.92	0.80	0.88
Math	11	12,938	2559.1	2551.0	0.91	0.92	0.78	0.86

### 3.3 Regression Analyses

Table 8 shows the regression coefficients including the intercept, slope, and  $R^2$  values for the regression models by grade and subject. In ELA, the slope coefficients range from approximately 0.79 to 0.92, and the  $R^2$  values range from 0.61 to 0.67, suggesting that ICA scores account for a substantial proportion of the variance in Summative Assessment scores. For mathematics, the slope coefficients range from 0.82 to 0.98, with  $R^2$  values between 0.60 and 0.70. All models were statistically significant ( $p$ -value  $<.05$ ).

Table 8. Regression Coefficients and  $R^2$  Values.

Content				
Area	Grade	Intercept	Slope	R2
ELA	3	227.2	0.912	0.605
ELA	4	209.7	0.924	0.644
ELA	5	268.0	0.902	0.662
ELA	6	522.1	0.794	0.605
ELA	7	495.1	0.806	0.645
ELA	8	519.9	0.795	0.652
ELA	11	411.3	0.842	0.665
Math	3	64.7	0.982	0.602
Math	4	175.9	0.935	0.698
Math	5	86.4	0.970	0.630
Math	6	321.8	0.877	0.627
Math	7	139.3	0.947	0.660
Math	8	235.6	0.908	0.639
Math	11	442.7	0.824	0.613

### 3.4 Classification Accuracy

Table 9 shows the classification accuracy statistics for ELA and mathematics along with the ICA scale score threshold predicting proficiency or beyond on the Summative Assessment. The classification accuracy metric represents the percentage of students whose predicted Summative Assessment achievement level (proficient vs. not proficient) based on ICA scores matched their actual Summative Assessment achievement level. In both ELA and mathematics, classification accuracy is high across all grades, with ELA accuracy ranging from 82% to 84% and mathematics accuracy ranging from 82% to 86%.

Sensitivity, or the ability of the ICA to correctly identify students who achieve a “Proficient” level score on the Summative Assessment, shows variation across grades, with ELA sensitivity values ranging from 79% to 84% and mathematics from 79% to 86%.

Specificity, which measures the ability to correctly classify students who do not meet proficiency, remains consistently high for both content areas, ranging from 80% to 85% in ELA

and from 82% to 87% in mathematics. These high specificity values indicate that the ICA effectively classifies non-proficient students as non-proficient.

Precision, or the proportion of students correctly classified into the “Proficient” level among those predicted to achieve a proficient score is strong in ELA with values near 80% across grades. In mathematics, precision ranges from 66% to 80%.

Additionally, the AUC values for both content areas range from 0.89 (Grade 3 mathematics) to 0.94 (Grade 7 mathematics). An AUC closer to 1.0 indicates a strong ability of the ICA to differentiate between students who score in the “Proficient” level and those who score in the “Not Proficient” level.

*Table 9. Classification Accuracy Statistics.*

Content Area	Grade	ICA					
		Threshold	Accuracy	Sensitivity	Specificity	Precision	AUC
ELA	3	2393	0.815	0.836	0.800	0.759	0.900
ELA	4	2435	0.834	0.829	0.838	0.801	0.913
ELA	5	2474	0.842	0.832	0.851	0.832	0.918
ELA	6	2516	0.823	0.817	0.828	0.788	0.897
ELA	7	2545	0.817	0.791	0.841	0.821	0.902
ELA	8	2566	0.835	0.832	0.838	0.813	0.911
ELA	11	2588	0.835	0.822	0.852	0.872	0.914
Math	3	2413	0.816	0.790	0.837	0.799	0.890
Math	4	2460	0.847	0.849	0.846	0.793	0.925
Math	5	2489	0.835	0.852	0.827	0.709	0.918
Math	6	2508	0.829	0.855	0.816	0.696	0.912
Math	7	2538	0.862	0.853	0.866	0.759	0.937
Math	8	2547	0.843	0.862	0.835	0.692	0.925
Math	11	2607	0.844	0.862	0.838	0.664	0.926

### 3.5 Results by Season

The data were also examined by season. The data were disaggregated as follows:

- Early fall: August, September, October
- Late fall: November, December
- Winter: January, February, March
- Spring: April, May, June

Table 10 shows the sample sizes for each disaggregation. In grades 3-8 in both content areas and Grade 11 Mathematics, most students took the ICA in winter. The exception to this trend was in Grade 11 ELA where most students took the ICA in late fall.

Table 10. Sample Size by Season.

Content Area	Grade	Early Fall	Late Fall	Winter	Spring
ELA	3	1,518	1,747	6,420	996
ELA	4	1,507	1,796	6,785	1,558
ELA	5	1,612	1,852	7,022	1,271
ELA	6	1,446	2,075	4,541	341
ELA	7	1,702	2,141	6,118	542
ELA	8	1,700	2,338	6,631	575
ELA	11	1,833	4,245	3,924	180
Math	3	1,531	3,981	7,646	1,702
Math	4	1,518	3,896	7,610	1,772
Math	5	1,654	3,980	8,375	1,580
Math	6	2,072	4,354	6,377	1,422
Math	7	2,158	3,584	5,012	1,547
Math	8	2,033	3,634	5,213	1,986
Math	11	1,339	4,120	7,018	461

Table 11 shows the corrected correlation coefficients by season of ICA administration. The correlation coefficients demonstrate strong relationships across seasons. The correlation coefficients tended to be lowest in early fall for grades 3-8 in both content areas. In grade 11, the correlation coefficient was lowest in spring for ELA and late fall for mathematics. In general, the correlation coefficients were highest in winter.

Table 11. Corrected Correlation Coefficients by Season.

Content Area	Grade	Early Fall	Late Fall	Winter	Spring
ELA	3	0.746	0.864	0.878	0.899
ELA	4	0.818	0.873	0.892	0.865
ELA	5	0.850	0.879	0.891	0.857
ELA	6	0.770	0.873	0.879	0.814
ELA	7	0.837	0.847	0.890	0.834
ELA	8	0.795	0.859	0.899	0.879
ELA	11	0.851	0.915	0.865	0.775
Math	3	0.732	0.825	0.871	0.889
Math	4	0.831	0.872	0.924	0.875
Math	5	0.806	0.808	0.894	0.861
Math	6	0.820	0.841	0.894	0.849
Math	7	0.840	0.842	0.891	0.921
Math	8	0.805	0.805	0.888	0.929
Math	11	0.882	0.776	0.876	0.791

Table 12 shows the accuracy statistics by season. Almost all of the results were in the 0.80 range, with only Grade 3 ELA, Grade 7 ELA, and Grade 3 mathematics in early fall and Grade 7 ELA in spring with results in the high 0.70 range.

*Table 12. Classification Accuracy Statistics by Season.*

Content Area	Grade	Early Fall	Late Fall	Winter	Spring
ELA	3	0.776	0.826	0.831	0.857
ELA	4	0.813	0.841	0.847	0.832
ELA	5	0.834	0.836	0.852	0.829
ELA	6	0.804	0.819	0.832	0.820
ELA	7	0.789	0.815	0.828	0.795
ELA	8	0.817	0.840	0.848	0.862
ELA	11	0.828	0.853	0.831	0.804
Math	3	0.769	0.820	0.840	0.847
Math	4	0.836	0.846	0.865	0.853
Math	5	0.807	0.830	0.858	0.817
Math	6	0.808	0.820	0.857	0.838
Math	7	0.857	0.847	0.862	0.871
Math	8	0.841	0.825	0.853	0.855
Math	11	0.886	0.837	0.846	0.857

## Section 4. Partial Correlation with Time

Another question of interest is whether there are significant partial correlations between days between ICA and Summative Assessment completion and Summative Assessment scale score controlling for ICA scale score. Those partial correlations are given in Table 13 along with their associated *p*-values.

*Table 13. Weighted Partial Correlations of Days between ICA and Summative Assessment Completion and Summative Assessment Scale Score (Controlling for ICA Scale Score) by Grade and Subject.*

Content Area	Grade	Weighted Partial Correlation	Partial Correlation P-Value
ELA	3	0.270	0.000
ELA	4	0.158	0.000
ELA	5	0.093	0.000
ELA	6	0.153	0.000
ELA	7	0.048	0.000
ELA	8	0.114	0.000
ELA	11	0.019	0.057
Math	3	0.330	0.000
Math	4	0.268	0.000
Math	5	0.236	0.000
Math	6	0.204	0.000

Content Area	Grade	Weighted Partial Correlation	Partial Correlation P-Value
Math	7	0.103	0.000
Math	8	0.060	0.000
Math	11	-0.038	0.000

These weighted partial correlations (except in high school) suggest that the longer the time period between ICA and Summative Assessment completion, the higher the Summative Assessment scale score after controlling for ICA scale score (with that partial relationship being stronger in lower grades than in higher grades). One hypothesis that could explain this finding is that ICA tests taken earlier in the year leave more room for learning before taking the Summative Assessment.

To further investigate the predictive relationship between ICA scale scores and days between ICA and Summative Assessment completion (as predictors) and Summative Assessment scale scores (as outcomes), a weighted predictive model was estimated for each grade and subject with the following form:

$$\widehat{SS}_{sum} = \hat{\beta}_0 + \hat{\beta}_1 SS_{ica} + \hat{\beta}_2 days$$

where

- $\widehat{SS}_{sum}$  is predicted Summative Assessment scale score,
- $SS_{ica}$  is ICA scale score,
- $days$  is the number of days between completion of ICA and Summative Assessment.

Results of the analysis are given in Table 14 where  $p_{\hat{\beta}_i}$  is the  $p$ -value of the regression coefficient  $\hat{\beta}_i$ . Table 14 shows that all regression coefficients were statistically significant (given  $\alpha_{critical} = 0.05$ ) except for the regression coefficient for  $days$  in Grade 11 ELA.

*Table 14. Results of Regression Analysis of Summative Assessment Scale Score as a Function of ICA Scale Score and Days between ICA and Summative Assessment Administration.*

Content Area	Grade	$\beta_0$	$\beta_1$	$\beta_2$	$p(\beta_0)$	$p(\beta_1)$	$p(\beta_2)$
ELA	3	36.87	0.973	0.297	0.038	0.000	0.000
ELA	4	134.6	0.948	0.157	0.000	0.000	0.000
ELA	5	230.6	0.913	0.095	0.000	0.000	0.000
ELA	6	469.8	0.806	0.164	0.000	0.000	0.000
ELA	7	475.7	0.811	0.055	0.000	0.000	0.000
ELA	8	471.5	0.808	0.129	0.000	0.000	0.000
ELA	11	405.9	0.842	0.027	0.000	0.000	0.057
Math	3	-210.0	1.081	0.334	0.000	0.000	0.000
Math	4	34.1	0.983	0.230	0.007	0.000	0.000
Math	5	-40.8	1.011	0.239	0.006	0.000	0.000
Math	6	190.2	0.919	0.215	0.000	0.000	0.000

Content Area	Grade	$\beta_0$	$\beta_1$	$\beta_2$	$p(\beta_0)$	$p(\beta_1)$	$p(\beta_2)$
Math	7	68.2	0.970	0.105	0.000	0.000	0.000
Math	8	192.1	0.923	0.069	0.000	0.000	0.000
Math	11	459.9	0.819	-0.053	0.000	0.000	0.000

Based on the results in Table 14, it is possible to predict Summative Assessment scale scores as a function of ICA scale scores and days between ICA and Summative Assessment administration. To better understand the partial relationships between *days* and Summative Assessment scale scores, predicted Summative Assessment scale scores are plotted as a dependent variable (on the y axis) with days between ICA and Summative Assessment administration plotted as a predictor (on the x axis) and different lines of prediction for ICA summative scale scores at intervals 50 points apart. Plots for ELA for grades 3, 4, 5, 6, 7, 8, and 11 are given in Figures 1, 2, 3, 4, 5, 6, and 7 respectively. Plots for mathematics for grades 3, 4, 5, 6, 7, 8, and 11 are given in Figures 8, 9, 10, 11, 12, 13, and 14, respectively.

These plots show similar results in ELA and mathematics for grades 3-8 with higher ICA scores predicting higher Summative Assessment scores and more days between ICA and Summative Assessment administration having a positive relationship with Summative Assessment scores. Grade 11 ELA has a similar set of relationships, though the relationship between Summative Assessment scores and days between ICA and Summative Assessment administration is less positive than in grades 3-8. Grade 11 mathematics has a slightly negative relationship between Summative Assessment scores and days between ICA and Summative Assessment administration.

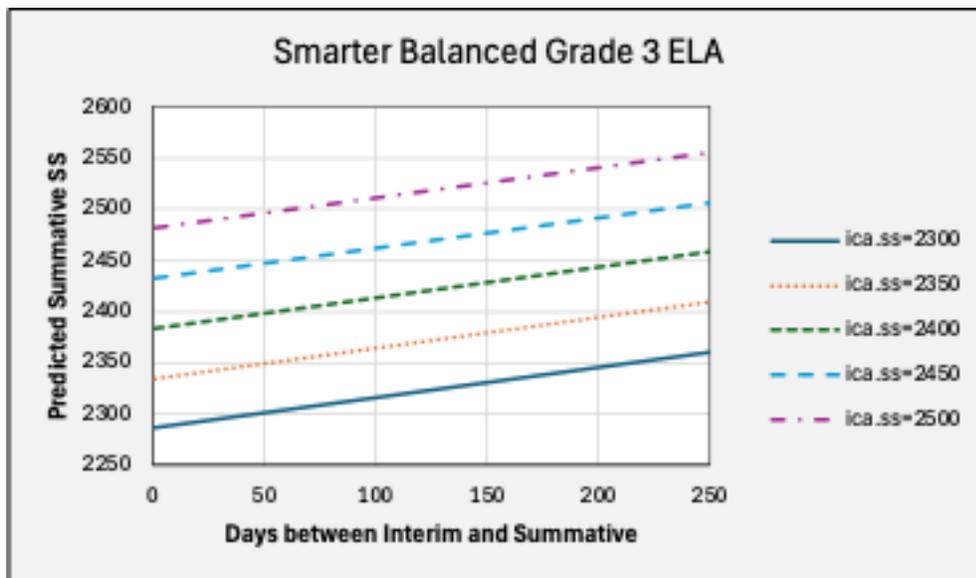


Figure 1. Predicted Grade 3 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

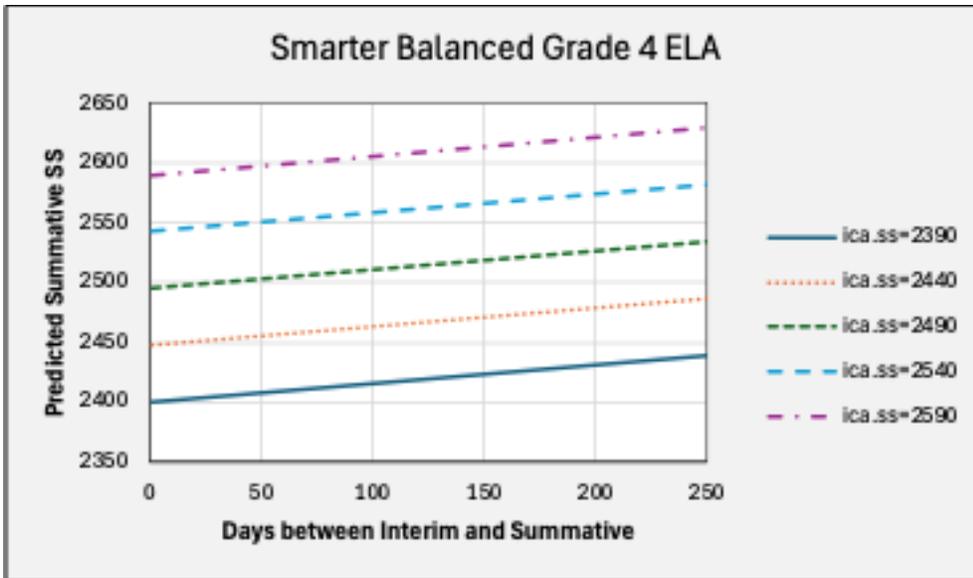


Figure 2. Predicted Grade 4 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

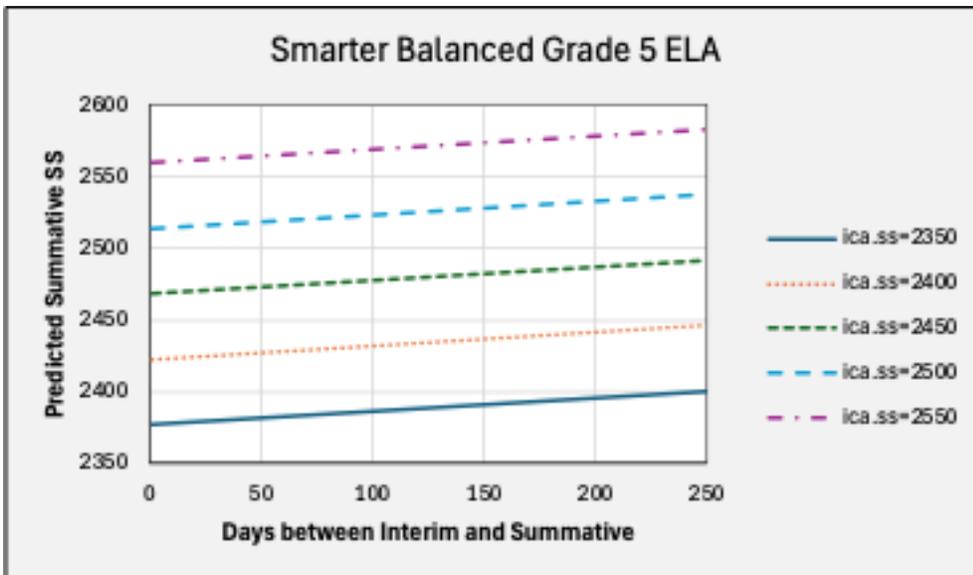


Figure 3. Predicted Grade 5 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

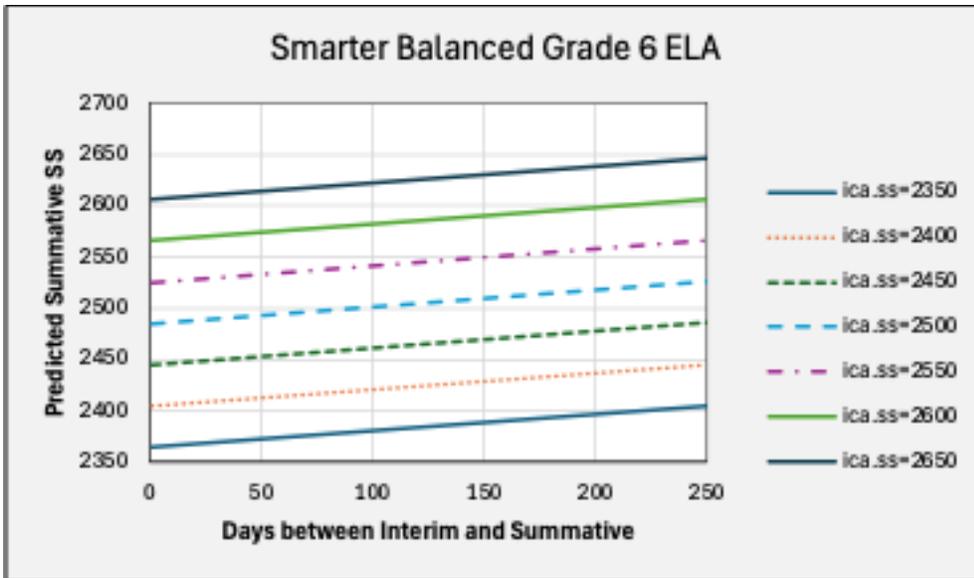


Figure 4. Predicted Grade 6 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

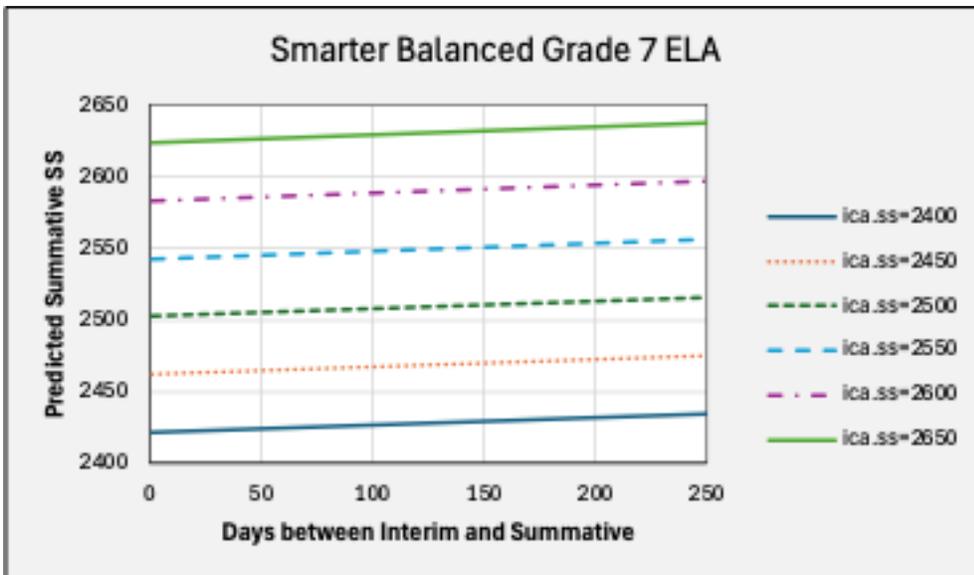


Figure 5. Predicted Grade 7 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

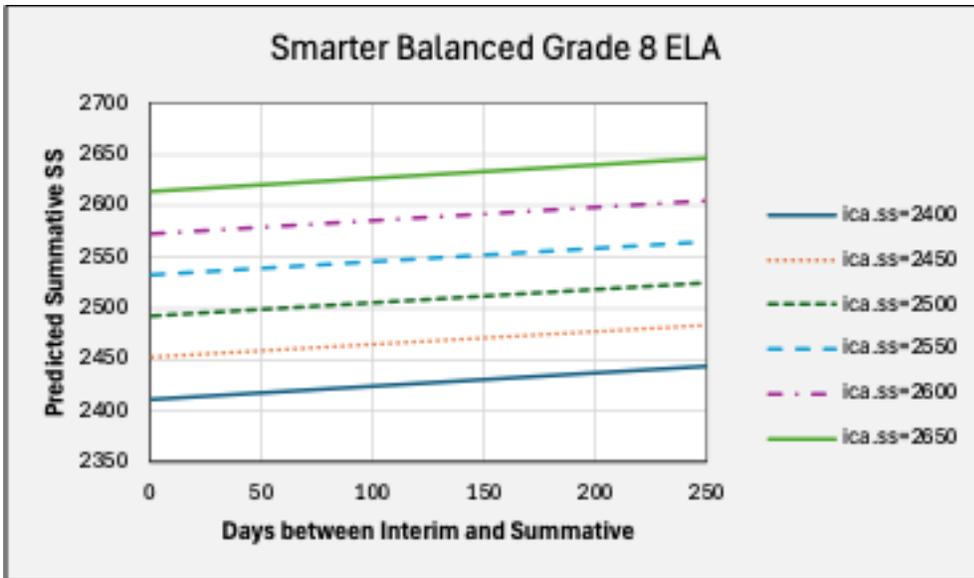


Figure 6. Predicted Grade 8 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

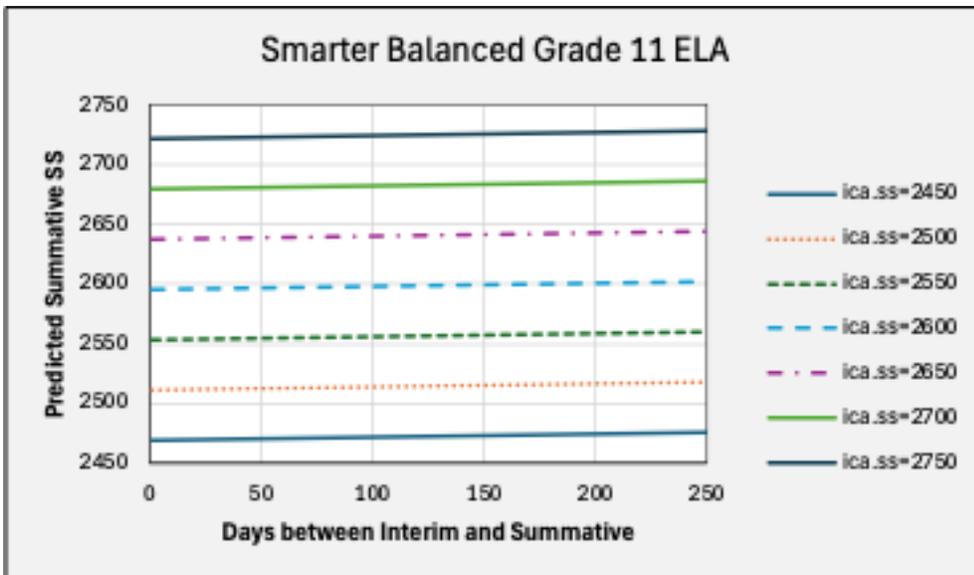


Figure 7. Predicted Grade 11 ELA Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

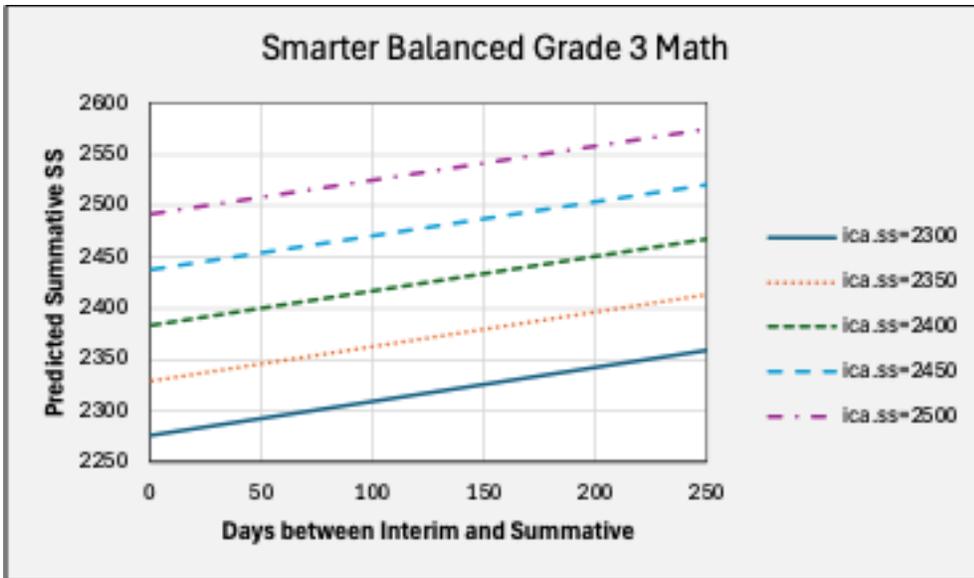


Figure 8. Predicted Grade 3 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

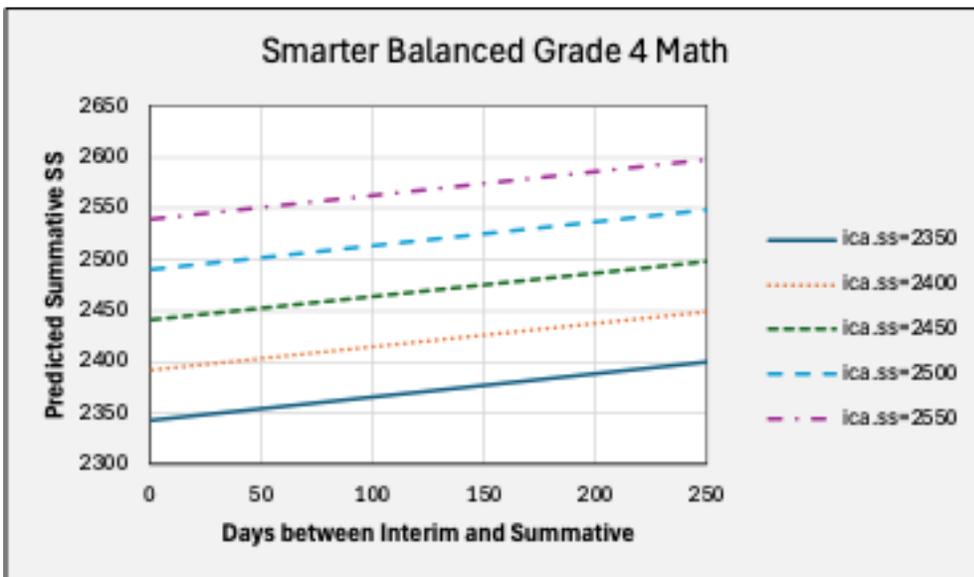


Figure 9. Predicted Grade 4 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

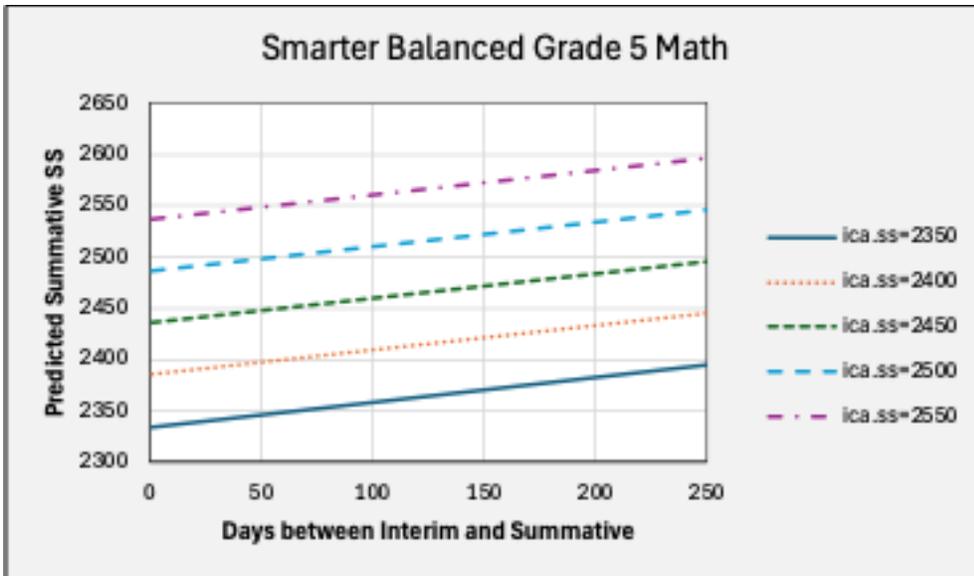


Figure 10. Predicted Grade 5 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

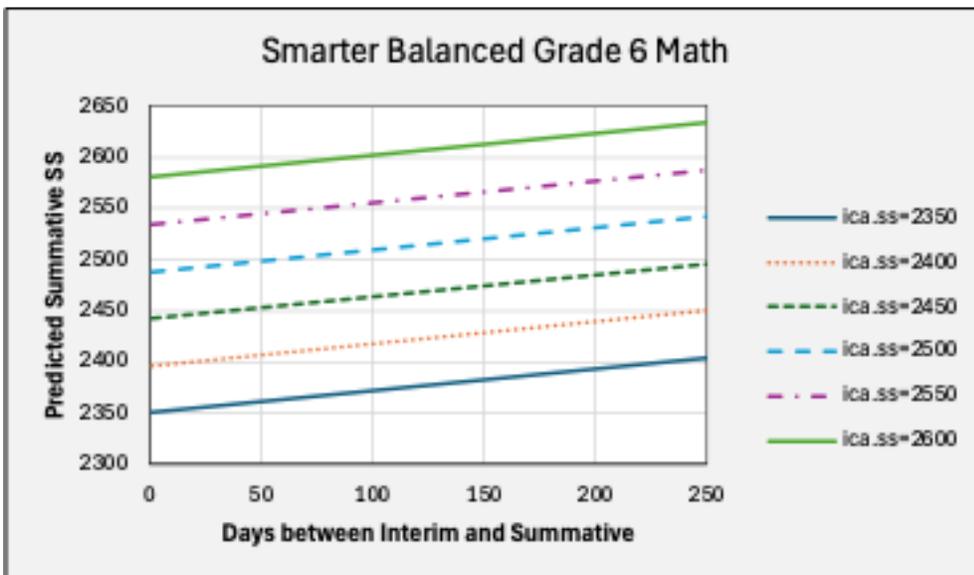


Figure 11. Predicted Grade 6 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

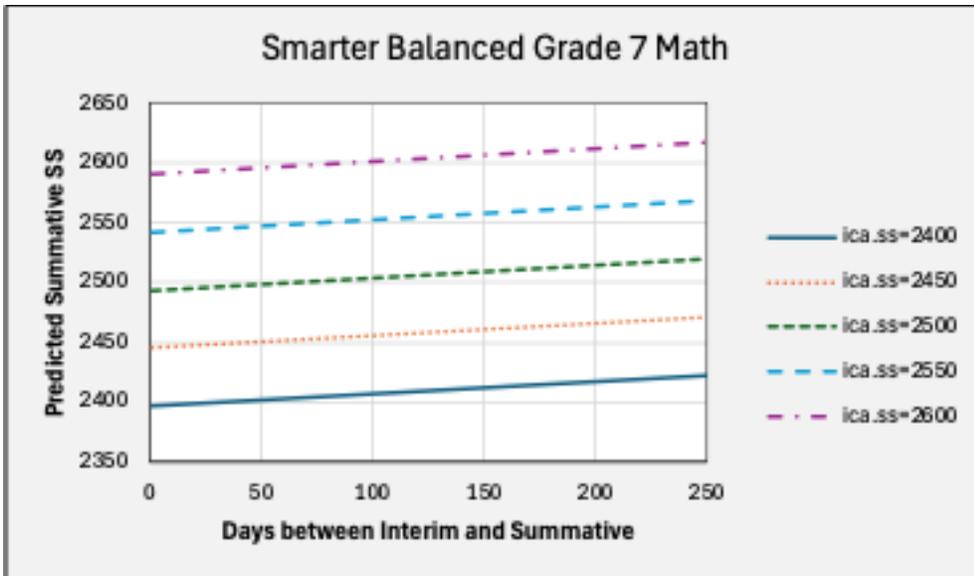


Figure 12. Predicted Grade 7 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

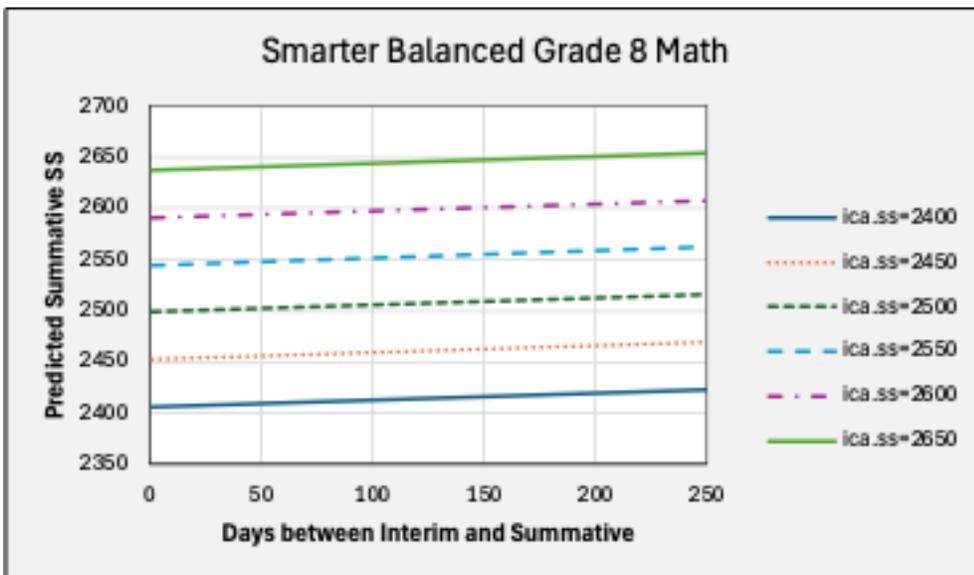


Figure 13. Predicted Grade 8 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

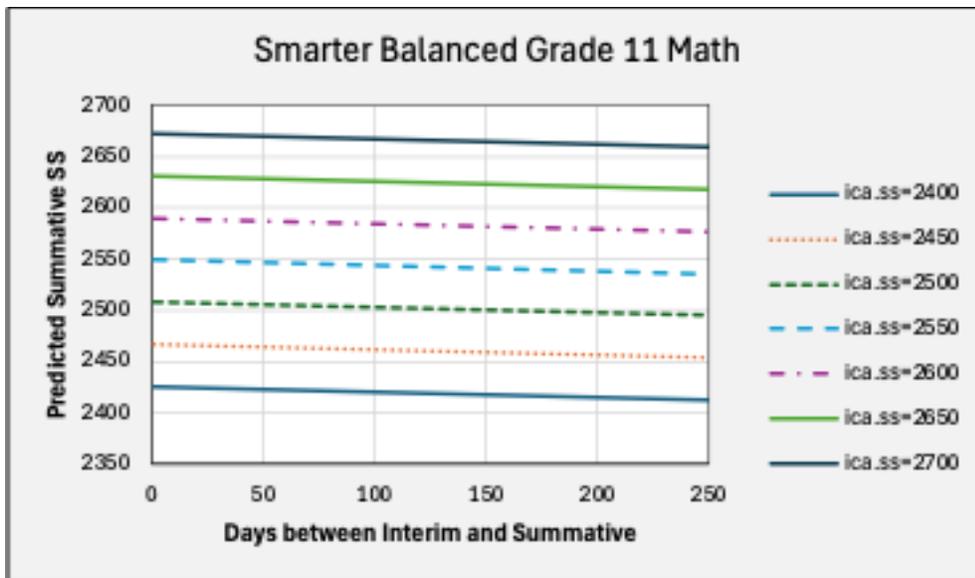


Figure 14. Predicted Grade 11 Mathematics Summative Assessment Scale Score as a Function of  $SS_{ica}$  and days.

## Section 5. Conclusion

This study provides substantial evidence of the predictive validity of the ICA for estimating student performance on the Smarter Balanced Summative Assessment in ELA and mathematics. Analyses of descriptive statistics, regression models, classification accuracy metrics, and partial correlations indicate that ICA scores are closely aligned with Summative Assessment scores.

The correlation and regression analyses show strong relationships between ICA and Summative Assessment scores, with consistently high correlations and reliable predictive strength across grades in both ELA and mathematics. These results underscore the ICA’s capability to offer a snapshot of student readiness throughout the school year that aligns with their end-of-year achievement, making it a useful tool for educators to gauge progress and adjust instruction as needed.

The classification accuracy metrics further validate the ICA’s effectiveness in predicting student performance categories on the Summative Assessment. High values in accuracy, specificity, and precision, along with robust AUC scores, affirm that the ICA can reliably identify students who are on track to meet proficiency standards, as well as those who may require additional support to reach these benchmarks. Notably, the sensitivity and precision metrics reflect that the ICA is particularly effective in identifying students who are likely to achieve or exceed standards, while maintaining a low rate of misclassification.

Finally, partial correlations between Summative Assessment scale scores and days between ICA and Summative Assessment administration (controlling for ICA scale scores) also suggest that longer times between ICA and Summative Assessment administration give more time for students to learn, resulting in higher Summative Assessment scores given the same ICA scores.

Future research beyond the current set of analyses would be useful. Studies should explore the predictive relationships for subgroups, examine these same predictive relationships with ICA data from other states, and explore how the usage of ICAs multiple times within a school year affects results.

In conclusion, this study supports the ICA's role as a valuable interim measure within the Smarter Balanced Assessment System. Its predictive alignment with Summative Assessment performance and its reliability in classifying student achievement levels suggest that the ICA serves as a critical tool for monitoring academic progress and guiding instructional decisions. By leveraging the ICA, educators can make data-informed interventions that support student achievement and readiness, ultimately enhancing educational outcomes and ensuring alignment with state standards.

## References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Carlson, J.E., (2013). *Operational best practices*. Washington, DC: Council of Chief State School Officers & Association of Test Publishers. Retrieved from <https://ccsso.org/sites/default/files/2023-09/ATP%20Best%20Practices%20Version%202-FINAL-082113-interior-300dpi.pdf>
- California Department of Education. (2024). *2022–23 CAASPP Smarter Balanced Technical Report*. Retrieved from <https://www.cde.ca.gov/ta/tg/ca/caaspprptstudies.asp>