Study of the Relationship Between the Early Assessment Program and the Smarter Balanced Field Tests

Contract #5417

Report on the study of the relationship between the Early Assessment Program and the Smarter Balanced Field Test

Prepared for the California Department of Education by Educational Testing Service

Final Submitted March 13, 2015

Listening. Learning. Leading.®

Table of Contents

Section 1: Executive Summary **1**
1.A. Introduction 1
1.B. Background 1
1.C. Statement of Purpose 1
1.D. Report Structure 2
1.E. Summary of Results 2
Section 2: Methods **3**
2.A. Data 3
2.B. Analyses 3
Section 3: Results **5**
3.A. Descriptive Statistics 5
3.B. Prediction Linking 9
3.C. Issues and Risks Associated with Smarter Balanced Field-Test Data 11
Section 4: Conclusion **13**
4.A. Relationship 13
References **14**
Appendix A: Demographic Information for the FT and EAP Samples **15**
Appendix B: EAP Scale Scores and FT Scale Scores Relationships **19**
Appendix C: Conditional Probabilities **22**

List of Tables

Table 2.1 Vertical Scaling Constants on the Reporting Metric 3
Table 2.2 Cut Scores of EAP Exemption Status 3
Table 3.1 Smarter Balanced FT Scale Scores Summary Statistics for California Grade 11 Students 5
Table 3.2 California Grade 11 Student Participation in the EAP and Matched Samples 5
Table 3.3 Smarter Balanced FT Scale Scores Summary Statistics of the Matched Samples 5
Table 3.4 Summary of Exemption Status for 2014 EAP Test Takers 6
Table 3.5 Summary of Exemption Status for the EAP Tests of the Matched Samples 6
Table 3.6 Percentages of Students Receiving Exemption Status on the EAP-ELA by the Smarter Balanced FT Performance Level 7
Table 3.7 Percentages of Students Receiving Exemption Status on the EAP-ALG II by the Smarter Balanced FT Performance Level 7
Table 3.8 Percentages of Students Receiving Exemption Status on the EAP-HSM by the Smarter Balanced FT Performance Level 8
Table 3.9 Correlation of EAP Scale Scores and the Smarter Balanced FT Scale Scores 8
Table 3.10 MLR Results to Predict Exemption Status Given the Performance on EAP-ELA Using the Standardized Smarter Balanced FT Scale Score as Predictor 10
Table 3.11 MLR Results to Predict Exemption Status Given the Performance on EAP-ALG II Using the Standardized Smarter Balanced FT Scale Score as Predictor 10
Table 3.12 MLR Results to Predict Exemption Status Given the Performance on EAP-HSM Using the Standardized Smarter Balanced FT Scale Score as Predictor 10
Table 3.13 Estimated Smarter Balanced FT Scale Score Cutoffs for Achieving Conditional Exemption Status or Greater Based on EAP Performance with 0.70 and 0.80 Probability 11
Table A.1 Demographic Information for the Smarter Balanced Field Test Sample, Grade Eleven (2300 ⩽ SS ⩽ 2800) and the SB FT sample removed due to the restriction of the SB scale score range applied to this study 15
Table A.2 Demographic Information for the EAP-ELA Overall Sample and Matched Sample, Grade Eleven (2300 ⩽ SS ⩽ 2800) 16
Table A.3 Demographic Information for the EAP-ALG II Overall Sample and Matched Sample, Grade Eleven (2300 ⩽ SS ⩽ 2800) 17
Table A.4 Demographic Information for the EAP-HSM Overall Sample and Matched Sample, Grade Eleven (2300 ⩽ SS ⩽ 2800) 18

Table of Figures

Figure B.1 Scatterplot of the EAP-ELA Scale Scores with the Smarter Balanced FT Scale Scores 19
Figure B.2 Scatterplot of the EAP-ALG II Scale Scores with the Smarter Balanced FT Scale Scores 20
Figure B.3 Scatterplot of the EAP-HSM Scale Scores with the Smarter Balanced FT Scale Scores.....21
Figure C.1 Conditional Probability of EAP-ELA Exemption Predicted by Smarter Balanced FT Scale Score 22
Figure C.2 Conditional Probability of EAP-ALG II Exemption Predicted by Smarter Balanced FT Scale Score 23
Figure C.3 Conditional Probability of EAP-HSM Exemption Predicted by Smarter Balanced FT Scale Score.....24

Section 1: Executive Summary

1.A. Introduction

In his "Recommendations for Transitioning California to a Future Assessment System," State Superintendent of Public Instruction Tom Torlakson suggested the use of the grade eleven Smarter Balanced Assessment Consortium summative English language arts/literacy (ELA) and mathematics assessments to serve as the indicator of college readiness for entry into college credit-bearing courses. This is a task that has been fulfilled through the California Standards Test (CST)/Early Assessment Program (EAP) since 2004.

Based on *Education Code (EC)* Section 99300 and support from the California State University (CSU) and the California Community Colleges (CCC), starting with the 2014–15 school year, the CST/EAP will be replaced with the Smarter Balanced (SB) grade eleven ELA and mathematics summative assessments. This study was designed to assess the relationship between the EAP assessments (Summative High School Mathematics, Algebra II, and English–Language Arts) and the SB Field Test (FT) grade eleven assessment.

1.B. Background

The EAP is a cooperative effort between the California State University (CSU), California Department of Education (CDE), and the State Board of Education (SBE) to determine students' readiness for college credit-bearing courses. In 2014, the EAP consisted of three tests, each augmented with items developed specifically for the purpose of determining college readiness: the CST for Algebra II (EAP-ALG II), the CST for Summative High School Mathematics (EAP-HSM), and the CST for English–Language Arts (EAP-ELA). When a student in grade eleven completes the multiple-choice CST for ELA (Grade 11) and either the CST for Algebra II or CST for Summative High School Mathematics, he or she is given the opportunity to complete an additional set of multiple-choice items, as well as an essay for the ELA assessment. The completion of the augmentation items allows the student the opportunity to earn possible exemption from the CSU Entry Level Mathematics (ELM) test and/or English Placement Test (EPT).

1.C. Statement of Purpose

This document describes the data and procedures used to establish and evaluate the relationship between EAP scale scores and the SB FT scale scores. The CSU and CDE will be using the grade eleven Smarter Balanced ELA and mathematics tests to identify student exemption status in 2015. There are three possible results based on the EAP assessments (CSU, n.d. and 2012):

- **Unconditionally exempt (Ready for college-level CSU and participating California Community Colleges [CCC] English/Mathematics coursework)**—Students meet CSU and participating CCC placement standards for entry-level coursework and are exempt from the placement tests required upon admission.
- **Conditionally exempt (Ready for college-level CSU and participating CCC English/ Mathematics coursework–Conditional)**—Students are considered ready at that moment in time they take the test but will be encouraged to maintain their college-level proficiency in English and/or mathematics by participating in approved senior year coursework.

Students who successfully complete an approved senior year experience are exempt from the requirement to take the CSU's EPT and/or ELM test or the placement tests at participating CCCs and will be eligible to enroll in college-level English or mathematics courses upon admission.

- **Not exempt (Not yet demonstrating readiness for college-level CSU and participating CCC English/Mathematics coursework)**—Students are not ready for college-level coursework and are encouraged to enroll in a senior year activity to increase their English and/or mathematics skills and are required to take the EPT and/or ELM and placement tests at the CSU upon admission.

This study provides preliminary results to help the CSU and CDE understand the relationship between the EAP and the SB FT. While suggestive of the results that may be achieved with the operational SB assessments, the results of this study should not be considered definitive because the Field Test data only approximate performance on the operational SB assessments. The primary goals of this study are:

1. To identify the percentage of exempt students on EAP tests by scale score levels of the SB FTs;
2. To evaluate the correlation of EAP scale scores and scale scores from SB FTs; and
3. To predict the conditional probabilities of exemption on EAP tests, given scale scores from SB FTs as the predictor (logistic regression analysis).

1.D. Report Structure

This report examines the methods used to analyze the data (Section 2) and the results of these analyses (Section 3). It also includes three appendixes: Appendix A, which provides a demographic breakdown of the FT and EAP samples and the California enrollment data for students in grade eleven in 2013[1]; Appendix B, which provides scatterplots that show the relationships between the EAP scale scores and FT scale scores; and Appendix C, which shows the conditional probabilities of EAP exemptions predicted by FT scale scores.

1.E. Summary of Results

The results show that the matched samples were generally representative of the overall sample for the corresponding EAP test. Correlations between the SB FT scale scores and the EAP test scale scores were between 0.49 and 0.68. The SB FT scale scores were shown to have a statistically significant effect on predicting the CSU ELM/EPT exemption status based on EAP performance for each of the three EAP tests.

[1] California Standardized Testing and Reporting (STAR) 2013 complete data obtained after demographic data correction completed by LEAs was used as the most up-to-date source for demographic profiles for eleventh grade test takers.

Section 2: Methods

2.A. Data

About six percent of grade eleven students in California were selected to be included in the "standard-setting sample" for the 2014 Smarter Balanced Field Test. Students in this sample were given either an ELA or mathematics assessment designed to resemble the Smarter Balanced operational test in length, difficulty, and content distribution. Among these students, some elected to take EAP assessments. This study compares results for students in the Smarter Balanced standard-setting samples with results for those same students on the EAP assessments.

Because of the schedule for the SB FT, some students took the EAP first and others took the SB FT first. Data corresponding to students who participated in both the EAP and SB FT were matched for analysis.

2.B. Analyses

The Smarter Balanced scale scores were obtained by applying a linear transformation on the SB proficiency estimates: ($SS = a * \theta + b$). The scale scores were rounded to an integer. The scaling constants a and b are provided by Smarter Balanced.

Table 2.1 lists the scaling constants for each subject for the theta-to-scaled score linear transformation.

Table 2.1 Vertical Scaling Constants on the Reporting Metric

Subject	Grade	Slope (a)	Intercept (b)
ELA	3–8, HS	85.8	2508.2
Math	3–8, HS	79.3	2514.9

Descriptive statistics including means, standard deviations, frequency distributions, and correlations were calculated across the test-taker groups that took the EAP, SB FT, and both EAP and the SB FT (referred to as matched samples hereafter). All descriptive statistics were disaggregated by EAP exemption classifications. (See Table 2.2 for cut scores associated with EAP exemption status.[2]) Demographic profiles of the EAP, SB FT, and matched samples were compared.

Table 2.2 Cut Scores of EAP Exemption Status

	Exemption Status		
Test	Unconditionally Exempt	Conditionally Exempt	Not Exempt
EAP-ELA	≥ 960	959–954	< 954
EAP-ALG II	≥ 856	855–845	< 845
EAP-HSM	≥ 943	942–920	< 920

[2] Pilot studies for the EAP tests were conducted in spring 2003. Results were used to set cut scores to determine examinee exemptions. EAP-ELA exemption status cut scores were implemented with the effective date of spring 2012 by the request of the CSU committee and the Chancellor's office, based on the findings from the EAP ELA conditional exemption cut score study conducted by the CSU in 2011.

The relationship between exemption on the CSU ELM/EPT assessments based on EAP performance and SB Field Tests was estimated by multinomial logistic regression (MLR) (Kolen & Brennan, 2004; Moran, Oranje, & Freund, 2009). Logistic regression is a type of probabilistic statistical classification model. It is used widely in many fields, including the social sciences. It measures the relationship between a categorical dependent variable, e.g., exemption status on CSU tests and one or more independent continuous variables, e.g., Smarter FT scale scores, by using probability scores as the predicted values of the dependent variable. Logistic regression can be binomial or multinomial.

Multinomial logistic regression is an extension of binary logistic regression which is used when modeling a categorical dependent variable with more than two levels, or categories ($J > 2$). The MLR generates J-1 sets of parameter estimates, comparing different levels of the dependent variable to a reference level. The model can be written as:

$$\mathrm{prob}(y_i = 1|x_i) = \frac{1}{1 + \sum_{j=2}^{J} \exp(x_i \beta_j)} \text{ for } m = 1$$

$$\mathrm{prob}(y_i = m|x_i) = \frac{\exp(\mathbf{x}_i \beta_m)}{1 + \sum_{j=2}^{J} \exp(\mathbf{x}_i \beta_j)} \text{ for } m > 1$$

Based on the MLR model, the projected probabilities of obtaining an ELA/ALG II/HSM score that qualifies as an unconditional exemption and conditional exemption status were estimated for various levels of the SB FT scale scores. Model fit of the MLR model to the data was also examined.

Section 3: Results

3.A. Descriptive Statistics

In this section, a description of the data used in the study is discussed.

As described previously, the SB FT eleventh grade dataset was based on the students from the SB standard-setting sample, which comprised 17,312 students in grade eleven taking the ELA test and 17,087 students in grade eleven taking the mathematics test. The SB scale score range of [2300, 2800] was chosen to truncate outlying student ability estimates, resulting in 17,094 eleventh graders taking the ELA test (1.3% of the students were excluded by the scale-score range restriction) and 16,341 eleventh graders taking the mathematics test (4.4% of the students were excluded by the scale-score range restriction). The summary statistics for the SB FT scale scores are presented in Table 3.1.

Table 3.1 Smarter Balanced FT Scale Scores Summary Statistics for California Grade 11 Students

	N	Mean	SD
ELA	17,094	2564	100.3
Mathematics	16,341	2560	103.8

Table 3.2 shows the number of students who participated in each of the EAP CSTs and the number in each matched sample. The names of the resulting matched samples are EAP-ELA, EAP-ALG II, and EAP-HSM.

Table 3.2 California Grade 11 Student Participation in the EAP and Matched Samples

Matched Sample Name	Number (N) EAP	Number (N), EAP + SB-ELA FT	Number (N), EAP + SB-Math FT
EAP-ELA	329,748	13,722	–
EAP-ALG II	96,937	–	4,488
EAP-HSM	112,369	–	5,293

The summary statistics for the SB FT scale scores for the matched samples are presented in Table 3.3.

Table 3.3 Smarter Balanced FT Scale Scores Summary Statistics of the Matched Samples

	EAP-ELA			EAP-ALG II			EAP-HSM		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
ELA	13,722	2569	97.6	–	–	–	–	–	–
Mathematics	–	–	–	4,488	2544	89.6	5,293	2624	87.9

As shown in Table 3.1 and Table 3.3, the average SB FT-ELA scale score for the EAP-ELA matched sample was higher than that of the overall SB FT sample. The average SB FT-mathematics scale score for the EAP-HSM matched sample was almost a standard deviation higher than that of the overall SB FT sample. However, the average SB FT-mathematics scale score for the EAP-ALG II matched sample was lower than that of the overall SB FT sample. Cohen's d was used to calculate the effect size[3] of the difference between the average SB FT

[3] Cohen's $d = (\bar{X}_1 - \bar{X}_2)/\sqrt{((n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2)/(n_1 + n_2 - 2)}$ was used for independent t-test, with the value of $d < 0.2$ as negligible, $d = 0.2$ as small, $d = 0.5$ as medium and $d = 0.8$ as large (Cohen, 1988).

scale score for each of the three EAP matched samples and the overall SB FT sample. The value of Cohen's d was 0.05 for comparing the average SB FT-ELA scale scores for the EAP-ELA matched dataset and the overall SB FT sample, which indicates a negligible effect size. The value of Cohen's d was 0.16 for comparing the average SB FT-mathematics scale scores for the EAP-ALG II matched sample and the overall SB FT sample, which indicates a negligible effect size; and 0.64 for comparing the average SB FT-mathematics scale score for the EAP-HSM matched sample and the overall SB FT sample, which indicates a medium effect size.

The summary of exemption status for the ELA, ALG II and HSM tests for the 2014 overall EAP population is presented in Table 3.4. The summary of exemption status for the three EAP tests of the matched samples is presented in Table 3.5.

Table 3.4 Summary of Exemption Status for 2014 EAP Test Takers

	EAP-ELA		EAP-ALG II		EAP-HSM	
	N	%	N	%	N	%
Unconditionally Exempt	82,270	24.95	2,882	2.97	19,091	16.99
Conditionally Exempt	47,883	14.52	12,578	12.98	72,989	64.95
Not Exempt	199,595	60.53	81,477	84.05	20,289	18.06
Total	**329,748**	**100.00**	**96,937**	**100.00**	**112,369**	**100.00**

Table 3.5 Summary of Exemption Status for the EAP Tests of the Matched Samples

	EAP-ELA		EAP-ALG II		EAP-HSM	
	N	%	N	%	N	%
Unconditionally Exempt	3,526	25.69	154	3.43	612	11.56
Conditionally Exempt	2,142	15.61	713	15.88	3,738	70.62
Not Exempt	8,054	58.69	3,621	80.68	943	17.82
Total	**13,722**	**100.00**	**4,488**	**100.00**	**5,293**	**100.00**

The percentages of both unconditionally exempt and conditionally exempt status for each of the three EAP tests from the matched sample were slightly higher than those of the overall EAP samples, except for the unconditionally exempt status for the EAP-HSM test. In general, the distributions of the exemption status were similar between the overall sample and the matched sample for each of the three EAP tests.

In addition, Table A.1 through Table A.4 in Appendix A present the demographic profiles of the SB FT overall sample, SB FT samples removed due to the restriction of the SB FT scale score range [2300, 2800] for this study, EAP overall samples, and the three matched samples in comparison with the enrollment data for California eleventh graders in 2013. As shown in Table A.2 through Table A.4, the demographic profile of the EAP-ELA matched sample was close to the overall enrollment for California eleventh graders in 2013, except that there were more white students (31.46% vs. 27.80%), more English-only students (59.23% vs. 55.37%), and fewer English learners (6.75% vs. 11.50%) in the EAP-ELA matched sample compared to the enrollment data for California eleventh graders in 2013.

The demographic profiles of the EAP-ALG II and EAP-HSM matched samples have some discrepancies compared to the enrollment data for California eleventh graders in 2013. There were more Hispanic students (55.06% vs. 49.74%) and more reclassified fluent English proficient (R-FEP) students (31.91% vs. 25.16%) in the EAP-ALG II matched sample compared to the enrollment data for California eleventh graders in 2013. There were more female students

(55.68% vs. 49.08%), fewer Hispanic students (41.53% vs. 49.74%), more Asian students (20.05% vs. 9.27%), more R-FEP students (35.63% vs. 25.16%), fewer English learners (1.78% vs. 11.50%) and fewer students with reported disabilities (1.23% vs. 8.55%) in the EAP-HSM matched sample compared to the enrollment data for California eleventh graders in 2013.

Table 3.6 through Table 3.8 present the percentage of students at each level on the SB FT by exemption status as determined by the EAP assessments for each of the matched samples. For the EAP-ELA matched sample, among the students who were above the Smarter Balanced Level 4 cut, 76.02 percent were unconditionally exempt and 15.35 percent were conditionally exempt. Among the students who were between the Smarter Balanced Level 3 and 4 cuts, 35.74 percent were unconditionally exempt and 23.65 percent were conditionally exempt.

For the EAP-ALG II matched sample, among the students who were above the Smarter Balanced Level 4 cut, 30 percent were unconditionally exempt and 46.67 percent were conditionally exempt. Among the students who were between the Smarter Balanced Level 3 and 4 cuts, 11.04 percent were unconditionally exempt and 34.18 percent were conditionally exempt.

Finally, for the EAP-HSM matched sample who were above the Smarter Balanced Level 4 cut, 40.57 percent were unconditionally exempt and 58.23 percent were conditionally exempt. Among the students who were between the Smarter Balanced Level 3 and 4 cuts, 12.65 percent were unconditionally exempt and 82.07 percent were conditionally exempt.

Table 3.6 Percentages of Students Receiving Exemption Status on the EAP-ELA by the Smarter Balanced FT Performance Level

Smarter Balanced FT Scale Score (SS)	N	EAP-ELA		
		Unconditionally Exempt	Conditionally Exempt	Not Exempt
Below Level 2 cut [2300, 2493)	3,044	3.98	5.62	90.41
Between Level 2 & 3 cuts [2493, 2583)	4,359	10.23	14.25	75.52
Between Level 3 & 4 cuts [2583, 2682)	4,580	35.74	23.65	40.61
Above Level 4 cut [2682, 2800]	1,739	76.02	15.35	8.63

Table 3.7 Percentages of Students Receiving Exemption Status on the EAP-ALG II by the Smarter Balanced FT Performance Level

Smarter Balanced FT Scale Score (SS)	N	EAP- ALG II		
		Unconditionally Exempt	Conditionally Exempt	Not Exempt
Below Level 2 cut [2300, 2543)	2,063	0.63	6.93	92.44
Between Level 2 & 3 cuts [2543, 2628)	1,613	2.48	17.67	79.85
Between Level 3 & 4 cuts [2628, 2718)	752	11.04	34.18	54.79
Above Level 4 cut [2718, 2800]	60	30.00	46.67	23.33

Table 3.8 Percentages of Students Receiving Exemption Status on the EAP-HSM by the Smarter Balanced FT Performance Level

Smarter Balanced FT Scale Score (SS)	N	EAP-HSM		
		Unconditionally Exempt	Conditionally Exempt	Not Exempt
Below Level 2 cut [2300, 2543)	905	1.33	48.62	50.06
Between Level 2 & 3 cuts [2543, 2628)	1,539	3.44	72.71	23.85
Between Level 3 & 4 cuts [2628, 2718)	2,181	12.65	82.07	5.27
Above Level 4 cut [2718, 2800]	668	40.57	58.23	1.20

Table 3.9 presents the correlations between the scale scores of all the three EAP tests and the linked SB FT scale scores. All three correlations were statistically significant at the 0.01 level. Additionally, Figure B.1 through Figure B.3 in Appendix B present the scatterplots of the EAP scale scores with the SB FT scale scores for the three EAP tests. These scatterplots also show a large degree of spread around the regression line, indicating that the corresponding EAP test and the SB FT measure similar but not identical constructs. The moderate correlation between the tests may also suggest that there are one or more factors leading to variability in the results, such as different item types (constructed response [CR] vs. multiple choice [MC]), an overall lower probability of answering a selected response question correctly by guessing, and by the difference in the testing modality (computer-based vs. paper and pencil).

The correlation information is an essential component in determining the most appropriate method for linking scores of the two assessments. Since the correlation between the EAP scale scores and the SB FT scale scores are lower than 0.87, which is the minimum requirement for equating two tests, a prediction linking method is more appropriate (Dorans & Walker, 2007). The predicted linear regression lines are also included in the scatterplots in Appendix B. Values of R^2 are included in Table 3.9, along with the correlations for each matched sample. The effect sizes for the correlations were medium for the relationship between the EAP-ALG II and SB FT-Mathematics scaled scores; and moderate for the relationships between EAP-HSM and SB FT-Mathematics scaled scores and EAP-ELA and SB FT-ELA scaled scores (Cohen, 1988)[4].

Table 3.9 Correlation of EAP Scale Scores and the Smarter Balanced FT Scale Scores

	SB ELA		SB Mathematics	
	P	R^2	ρ	R^2
EAP-ELA	.68 *	.46	–	–
EAP-ALG II	–		.49 *	.24
EAP-HSM	–		.61 *	.37

* $p < .01$

[4] Pearson correlations of 0.10, 0.30, and 0.50 indicate small, medium, and large effect, respectively (Cohen, 1988).

3.B. Prediction Linking

Table 3.10 through Table 3.12 summarize the results of the MLR to predict CSU ELM/EPT exemption status, based on EAP performance, using the standardized SB FT scale scores for each of the three EAP tests.

For the multinomial logistic regression model, a level of "not exempt" was set as the reference level of each model. The coefficient estimates for the model refer to the log odds ratio of the outcome category (unconditionally exempt or conditionally exempt) relative to the reference category (not exempt), where odds are defined as the ratio of the probability of the outcome category (unconditionally exempt or conditionally exempt) to the probability of the reference category (not exempt).

The likelihood ratio chi-square test results at the bottom of Table 3.10 through Table 3.12 show there was a statistically significant relationship between the standardized SB FT scale scores and the exemption classification with *p*-values less than .0001. In other words, the existence of a relationship between the SB FT scale score and the exemption classification was supported by the data. The overall model fit was moderate for predicting exemption status for the EAP-ELA test and EAP-HSM test since the pseudo R^2 values were 0.20 or higher[5]. The overall model fit was worse for predicting the exempt status of the EAP-ALG II test since the pseudo R^2 was lower than 0.20. Note that the range of pseudo R^2 values is from 0 to 1. However, it is not exactly analogous to the R^2 statistics for linear regression models—how it translates to an equivalence of R^2 is an empirical question related to each specific dataset. Table 3.9 is used to directly assess the R^2 statistics of linear regression models.

In addition, the odds ratio shows the relationship between standardized SB FT scale score and the exemption classification. One standardized unit increase on the SB-ELA scale indicates the odds of being conditionally exempt relative to being not exempt is 3.177 times more likely, and the odds of being unconditionally exempt relative to being not exempt is 8.215 times more likely. For the EAP-ALG II test, one standardized unit increase on the SB-mathematics scale indicates the odds of being conditionally exempt relative to being not exempt is 2.866 times more likely, and the odds of being unconditionally exempt relative to being not exempt is 6.532 times more likely. For the EAP-HSM test, one standardized unit increase on the SB-mathematics scale indicates the odds of being conditionally exempt relative to being not exempt are 3.299 times more likely, and the odds of being unconditionally exempt relative to being not exempt are 16.525 times more likely.

[5] McFadden's Pseudo $R^2 = 1\text{-}\ln(L_M)/\ln(L_0)$, defined as the change in terms of log-likelihood from the null model to the fitted model. McFadden's Pseudo R^2 can be interpreted as an approximate variance in the outcome accounted for by the independent variable(s). Values from 0.2 to 0.4 for McFadden's Pseudo R^2 indicate moderate model fit; values below 0.1 indicate poor model fit (Domencich & McFadden, 1975).

Table 3.10 MLR Results to Predict Exemption Status Given the Performance on EAP-ELA Using the Standardized Smarter Balanced FT Scale Score as Predictor

	Conditionally Exempt					Unconditionally Exempt				
Variable	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**
Standardized SB_ELA SS	1.156	0.034	<.0001	3.177	(2.973, 3.396)	2.106	0.038	<.0001	8.215	(7.623, 8.854)
(Intercept)	-1.257	0.027	<.0001	–	–	-1.330	0.031	<.0001	–	–

$\chi^2 = 5682.52$, df = 2, $p < .0001$; pseudo $R^2 = 0.22$

B = regression coefficient C.I. = confidence interval

SE = standard error

Table 3.11 MLR Results to Predict Exemption Status Given the Performance on EAP-ALG II Using the Standardized Smarter Balanced FT Scale Score as Predictor

	Conditionally Exempt					Unconditionally Exempt				
Variable	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**
Standardized SB_MATH SS	1.053	0.055	<.0001	2.866	(2.575, 3.189)	1.877	0.127	<.0001	6.532	(5.097, 8.372)
(Intercept)	-1.902	0.052	<.0001	–	–	-4.167	0.156	<.0001	–	–

$\chi^2 = 703.49$, df = 2, $p < .0001$; pseudo $R^2 = 0.14$

B = regression coefficient C.I. = confidence interval

SE = standard error

Table 3.12 MLR Results to Predict Exemption Status Given the Performance on EAP-HSM Using the Standardized Smarter Balanced FT Scale Score as Predictor

	Conditionally Exempt					Unconditionally Exempt				
Variable	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**	**B**	**SE**	***p*-value**	**Odds Ratio**	**95% C.I.**
Standardized SB_MATH SS	1.194	0.046	<.0001	3.299	(3.016, 3.608)	2.805	0.091	<.0001	16.525	(13.835, 19.738)
(Intercept)	1.850	0.049	<.0001	–	–	-0.837	0.088	<.0001	–	–

$\chi^2 = 1669.22$, df = 2, $p < .0001$; pseudo $R^2 = 0.20$

B = regression coefficient C.I. = confidence interval

SE = standard error

Figure C.1 through Figure C.3 in Appendix C present the conditional probabilities of achieving some level of exemption on EAP tests predicted by the SB FT scale scores for each of the three matched samples (EAP-ELA, EAP-ALG II, and EAP-HSM). Frequency distributions of the SB FT scale scores are also included in these figures. Note that for the [2300, 2800] SB scale score range, there were relatively large numbers of students at the higher end of the frequency distribution of the SB-ELA scale scores.

Also shown in the figures is that the probability of exemption and the SB FT scale score has a monotonic relationship, meaning the probability increases as the SB FT scale score increases or vice versa. Table 3.13 lists the corresponding SB FT scale score cutoffs and standard errors

associated with achieving some level of exemption on CSU EPT/ELM assessments with a 0.70 probability or greater and with a 0.80 probability or greater respectively. SB scale score cuts for Levels 3 and 4 and corresponding proportions of the EAP matched samples are also included in Table 3.13.

Table 3.13 Estimated Smarter Balanced FT Scale Score Cutoffs for Achieving Conditional Exemption Status or Greater Based on EAP Performance with 0.70 and 0.80 Probability

	Smarter Balanced FT Scale Score Cutoff (standard error)		SB SS Level 3 cut		SB SS Level 4 cut	
EAP Test	Prob = 0.70	Prob = 0.80	Score	%[1]	Score	%[2]
ELA	2653 (26.21)	2681 (26.41)	2583	46.05	2682	10.04
ALG II	2738 (30.90)	2773 (35.00)	2628	18.09	2718	1.34
HSM	2549 (45.48)	2587 (40.96)	2628	53.83	2718	12.62

[1] Percentages achieving SB Level 3 and above for the corresponding EAP matched sample
[2] Percentages achieving SB Level 4 and above for the corresponding EAP matched sample

3.C. Issues and Risks Associated with Smarter Balanced Field-Test Data

Using psychometrically sound methods, this study identified positive correlation between EAP exemption status and SB FT scale score. This study used data obtained from the results of the SB Field Tests and matched them to the results data obtained from EAP assessments. Although it provides an opportunity to investigate the relationship between EAP and SB tests, which do not have any items in common, using the performance data of the students taking both tests has some potential issues, especially in association with the use of the FT data in this study. Those issues might lessen or block the true relationship between EAP exemption status and SB FT scale score partially.

Content Coverage:

An important assumption is that a representative sample of EAP students will take a full-length Smarter Balanced Field Test that *closely resembles the operational Smarter Balanced test*. However, the number of items scored and that contributed to student ability estimates differed due to the Field Test administration and sampling plan implemented specifically for California.

Student Motivation:

Unlike the EAP tests, since no stakes were associated with the SB FT, motivation and the ability to detect it might be an issue. Guessing may cause an overall lower probability of answering a selected response question correctly. Although it is very difficult to measure motivation, the potential lack of motivation may have hindered the establishment of a completely valid and interpretable statistical relationship between the Smarter Balanced and EAP results.

Student Exposure to New Standards:

The 2014 administration was the first time that California eleventh graders took the SB FT. Current eleventh grade students are transitioning to the Common Core State Standards (CCSS) at a time when they are exiting the K–12 system. Consequently, some students taking these Field Tests may have been tested on material for which they had not yet received instruction, which could be reflected in their results. Therefore, the results might differ in future cohorts as they would have had more exposure to the CCSS.

Test Design:

Unlike the EAP tests, the SB FT was designed to include CR items and were computer-based tests. Different item types (CR vs. MC) and the difference in the testing modality (computer-based tests [CBT] vs. paper-pencil tests [PPTs]) may lead to variability in SB FT measure and EAP measure. However, results from the previous science Computer-based Testing Tryout study (CDE, 2013) suggested that different item types may not indicate different dimensions, and that CBT may lead to differential impact at the item and test level for some subgroups, but the effect sizes of the differences between CBT and PPT across subgroups were small.

Section 4: Conclusion

4.A. Relationship

The results show a positive moderate relationship between the SB FT scale scores and the EPT/ELM exempt statuses derived from EAP results. However, the SB FT and the corresponding EAP test do not measure identical constructs. Students who do well on the EAP may not do well on the SB FT assessments and vice versa, as shown in Table 3.6 through Table 3.8. In order to truly determine whether it is predictive of college readiness, future follow-up studies that examine the relationship between the Smarter Balanced tests and other tests, including the CSU's EPT and ELM, are warranted.

References

California Department of Education. (2013). *Computer-based testing try out report*. Retrieved from ~~http://www.cde.ca.gov/ta/tg/sr/documents/cbttryoutrpt.pdf~~. [Note: the preceding Web address is no longer valid.]

California State University. (2012) *Early Assessment Program (EAP) assessment and accountability information meeting.* Retrieved October 30, 2014, from http://www.calstate.edu/eap/documents/presentation_cde.ppt

California State University. (n.d.) *Early Assessment Program (EAP) frequently asked questions.* Retrieved October 27, 2014, from http://www.calstate.edu/eap/documents/eapfaqfinal.pdf

Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.

Domencich, T. A., & McFadden, D. (1975). *Urban Travel Demand: A behavioral analysis.* Amsterdam: North-Holland Publishing Company.

Dorans, N. J., & Walker, M. E. (2007). Sizing up linkages. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 179–198). New York: Springer.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking (2nd Ed.)*. New York, NY: Springer.

Moran, R., Oranje, A., & Freund, D. (2009). *NAEP 12th grade preparedness research: Establishing a statistical relationship between NAEP and SAT®*. Retrieved March 13, 2013, from ~~http://www.nagb.org/content/nagb/assets/documents/what-we-do/preparedness-research/statistical-relationships/SAT-NAEP_Linking_Study.pdf~~. [Note: the preceding Web address is no longer valid.]

Torlakson, T. (2013). *Recommendations for transitioning California to a future assessment system*. Retrieved April 12, 2014, from ~~http://www.cde.ca.gov/ta/tg/sa/documents/suptrecrptjan13.pdf.~~ [Note: the preceding Web address is no longer valid.]

Appendix A: Demographic Information for the FT and EAP Samples

Table A.1 Demographic Information for the Smarter Balanced Field Test Sample, Grade Eleven (2300 ≤ SS ≤ 2800) and the SB FT sample removed due to the restriction of the SB scale score range applied to this study

		SS < 2300 or SS > 2800				2300 ≤ SS ≤ 2800			
		ELA		Mathematics		ELA		Mathematics	
Category	**Subgroup**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Gender	Male	157	72.02	379	50.80	8,598	50.30	7,963	48.73
	Female	61	27.98	367	49.20	8,496	49.70	8,378	51.27
Ethnicity	White	49	22.48	130	17.43	5,691	33.29	3,943	24.13
	African American	13	5.96	54	7.24	751	4.39	822	5.03
	Hispanic	129	59.17	423	56.70	8,096	47.36	8,591	52.57
	Asian *	18	8.26	112	15.01	1,845	10.79	2,479	15.17
	Pacific Islander	2	0.92	2	0.27	83	0.49	89	0.54
	American Indian **	0	0.00	3	0.40	119	0.70	71	0.43
	Multirace	7	3.21	22	2.95	509	2.98	346	2.12
English Proficiency Level	English speaker	138	63.30	604	80.97	15,860	92.78	14,972	91.62
	English learner	80	36.70	142	19.03	1,234	7.22	1,369	8.38
Disability Status	No disability	148	67.89	654	87.67	15,965	93.40	15,470	94.67
	Disability	70	32.11	92	12.33	1,129	6.60	871	5.33
	Total	**218**	**100.00**	**746**	**100.00**	**17,094**	**100.00**	**16,341**	**100.00**

* Includes Filipino students

** Includes Alaskan Natives

Table A.2 Demographic Information for the EAP-ELA Overall Sample and Matched Sample, Grade Eleven (2300 ≤ SS ≤ 2800)

		2013 STAR		EAP-ELA			
		Overall		Overall		Matched	
Category	**Subgroup**	**N**	**%**	**N**	**%**	**N**	**%**
Gender	Male	238,808	50.83	163,434	49.56	6,787	49.46
	Female	230,618	49.08	166,083	50.37	6,929	50.50
Ethnicity	White	130,630	27.80	83,616	25.36	4,317	31.46
	African American	30,970	6.59	19,909	6.04	617	4.50
	Hispanic	233,722	49.74	169,528	51.41	6,716	48.94
	Filipino	13,696	2.91	11,400	3.46	377	2.75
	Asian *	43,559	9.27	33,638	10.20	1,138	8.29
	Pacific Islander	2,665	0.57	1,929	0.58	60	0.44
	American Indian **	3,476	0.74	1,902	0.58	85	0.62
	Multirace	10,253	2.18	7,826	2.37	412	3.00
English Proficiency Level ***	EO	260,178	55.37	176,075	53.40	8,128	59.23
	I-FEP	33,958	7.23	26,366	8.00	883	6.43
	R-FEP	118,216	25.16	94,799	28.75	3,761	27.41
	EL	54,022	11.50	30,080	9.12	926	6.75
Disability Status	No disability	428,734	91.25	309,722	93.93	13,013	94.83
	Disability	40,177	8.55	20,008	6.07	709	5.17
	Total	**469,854**	**100.00**	**329,748**	**100.00**	**13,722**	**100.00**

* Excludes Filipino students

** Includes Alaskan Natives

*** EO = English only; I-FEP = Initially fluent English proficient; R-FEP = Reclassified fluent English proficient; EL = English learner

Table A.3 Demographic Information for the EAP-ALG II Overall Sample and Matched Sample, Grade Eleven (2300 ≤ SS ≤ 2800)

		2013 STAR		EAP-ALG II			
		Overall		Overall		Matched	
Category	Subgroup	N	%	N	%	N	%
Gender	Male	238,808	50.83	46,436	47.90	2,160	48.13
	Female	230,618	49.08	50,443	52.04	2,327	51.85
Ethnicity	White	130,630	27.80	23,681	24.43	1,125	25.07
	African American	30,970	6.59	6,258	6.46	241	5.37
	Hispanic	233,722	49.74	53,764	55.46	2,471	55.06
	Filipino	13,696	2.91	3,402	3.51	143	3.19
	Asian *	43,559	9.27	6,584	6.79	389	8.67
	Pacific Islander	2,665	0.57	623	0.64	24	0.53
	American Indian **	3,476	0.74	534	0.55	20	0.45
	Multirace	10,253	2.18	2,091	2.16	75	1.67
English Proficiency Level ***	EO	260,178	55.37	50,924	52.53	2,305	51.36
	I-FEP	33,958	7.23	7,070	7.29	399	8.89
	R-FEP	118,216	25.16	30,258	31.21	1,432	31.91
	EL	54,022	11.50	8,132	8.39	349	7.78
Disability Status	No disability	428,734	91.25	93,079	96.02	4,314	96.12
	Disability	40,177	8.55	3,856	3.98	174	3.88
	Total	**469,854**	**100.00**	**96,937**	**100.00**	**4,488**	**100.00**

* Excludes Filipino students

** Includes Alaskan Natives

*** EO = English only; R-FEP = Reclassified fluent English proficient; I-FEP = Initially fluent English proficient; EL = English learner

Table A.4 Demographic Information for the EAP-HSM Overall Sample and Matched Sample, Grade Eleven (2300 ≤ SS ≤ 2800)

		2013 STAR		EAP-HSM			
		Overall		Overall		Matched	
Category	**Subgroup**	**N**	**%**	**N**	**%**	**N**	**%**
Gender	Male	238,808	50.83	52,181	46.44	2,346	44.32
	Female	230,618	49.08	60,141	53.52	2,947	55.68
Ethnicity	White	130,630	27.80	31,994	28.47	1,379	26.05
	African American	30,970	6.59	4,441	3.95	206	3.89
	Hispanic	233,722	49.74	44,183	39.32	2,198	41.53
	Filipino	13,696	2.91	5,685	5.06	312	5.89
	Asian *	43,559	9.27	22,271	19.82	1,061	20.05
	Pacific Islander	2,665	0.57	523	0.47	19	0.36
	American Indian **	3,476	0.74	457	0.41	5	0.09
	Multirace	10,253	2.18	2,815	2.51	113	2.13
English Proficiency Level ***	EO	260,178	55.37	60,392	53.74	2,667	50.39
	I-FEP	33,958	7.23	12,715	11.32	646	12.20
	R-FEP	118,216	25.16	36,255	32.26	1,886	35.63
	EL	54,022	11.50	2,393	2.13	94	1.78
Disability Status	No disability	428,734	91.25	111,134	98.90	5,228	98.77
	Disability	40,177	8.55	1,232	1.10	65	1.23
	Total	**469,854**	**100.00**	**112,369**	**100.00**	**5,293**	**100.00**

* Excludes Filipino students

** Includes Alaskan Natives

*** EO = English only; I-FEP = Initially fluent English proficient; R-FEP = Reclassified fluent English proficient; EL = English learner

Appendix B: EAP Scale Scores and FT Scale Scores Relationships

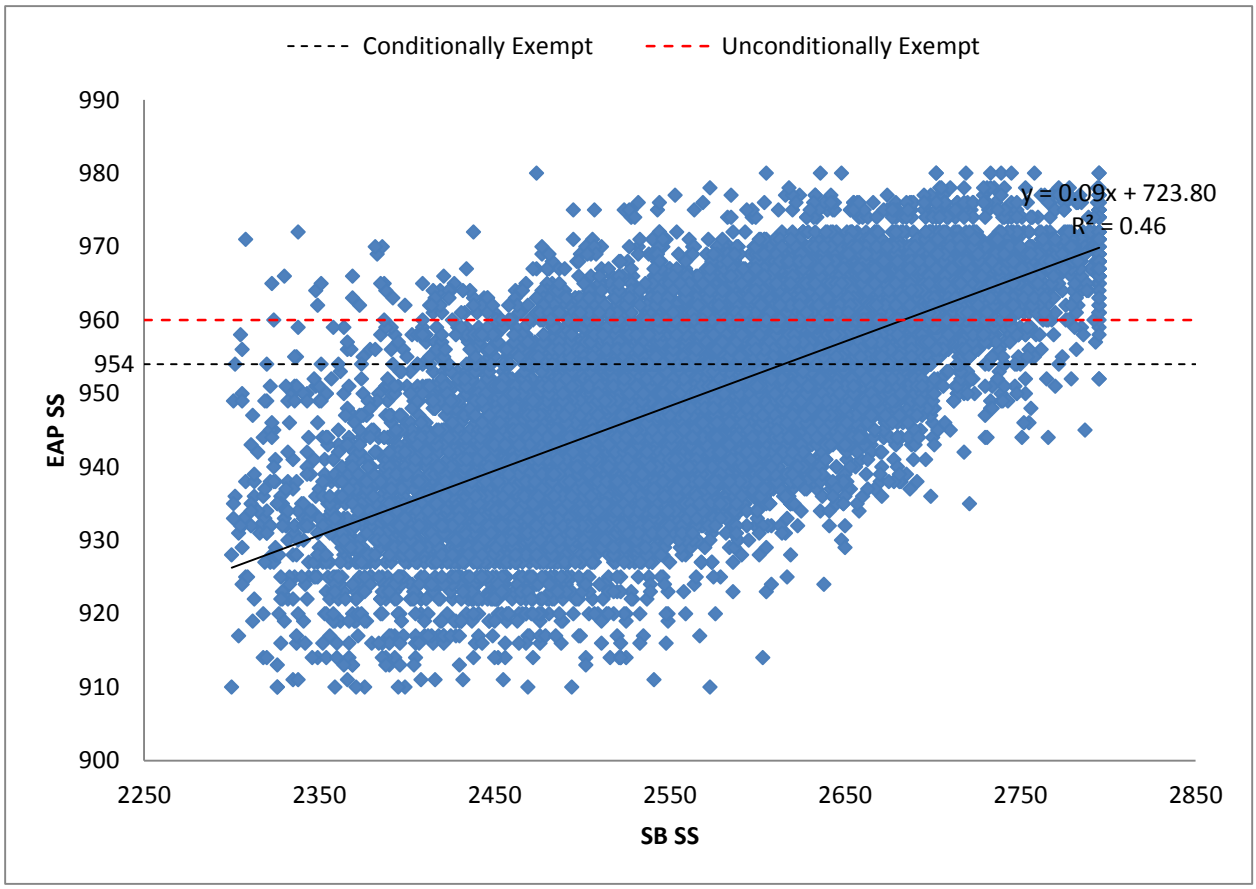

Figure B.1 Scatterplot of the EAP-ELA Scale Scores with the Smarter Balanced FT Scale Scores

The solid line in Figure B.1 represents the linear concordance between the EAP-ELA scale scores and the Smarter Balanced FT scale scores.

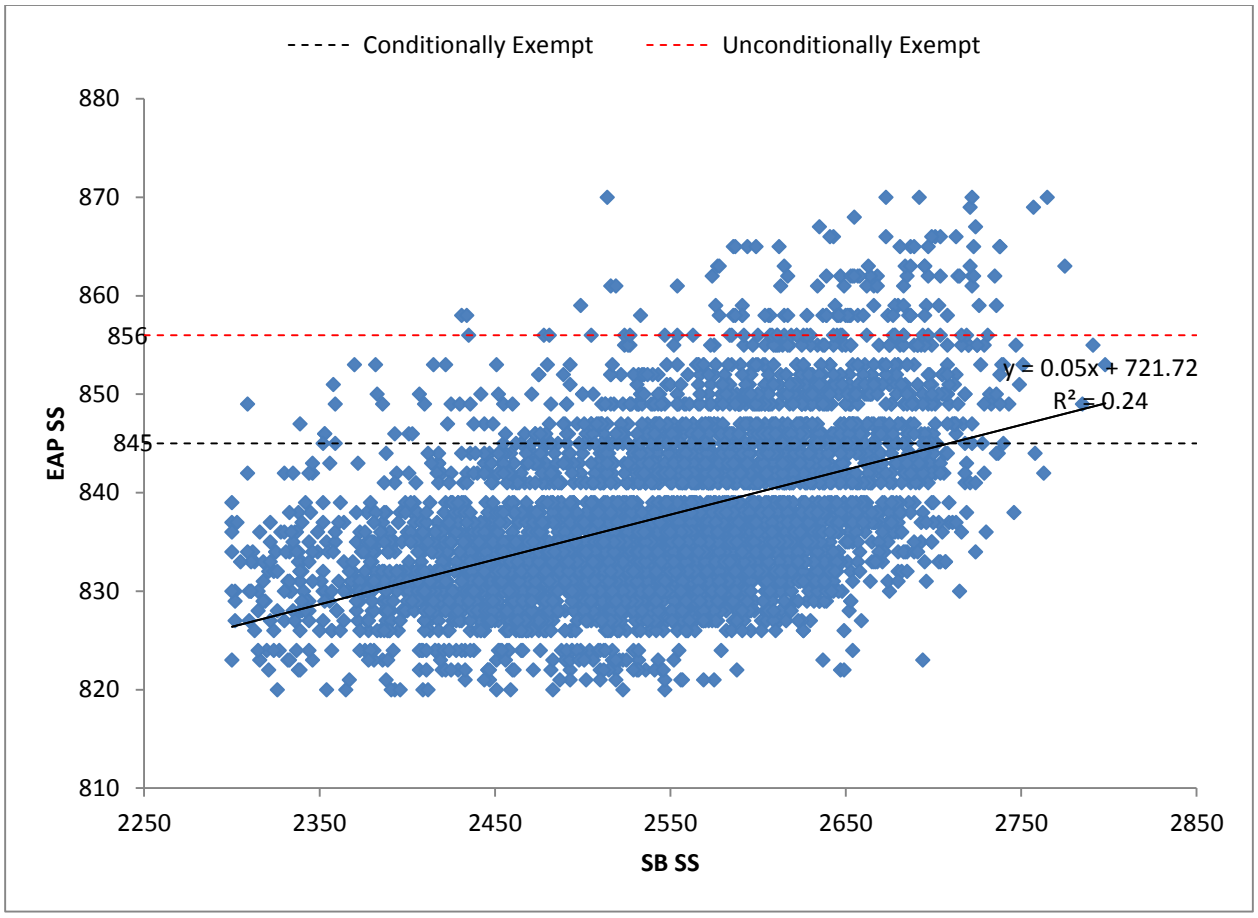

Figure B.2 Scatterplot of the EAP-ALG II Scale Scores with the Smarter Balanced FT Scale Scores

The solid line in Figure B.2 represent the linear concordance between the EAP-ALG II scale scores and the Smarter Balanced FT scale scores. Non-linear concordance was not shown here because the regression coefficient for the quadratic form was very small and the increase of R^2 from the linear regression to non-linear regression was trivial. (Note: non-linear regression equation: $y = 0.0001x^2 - 0.66x + 1613.20, R^2 = 0.27$)

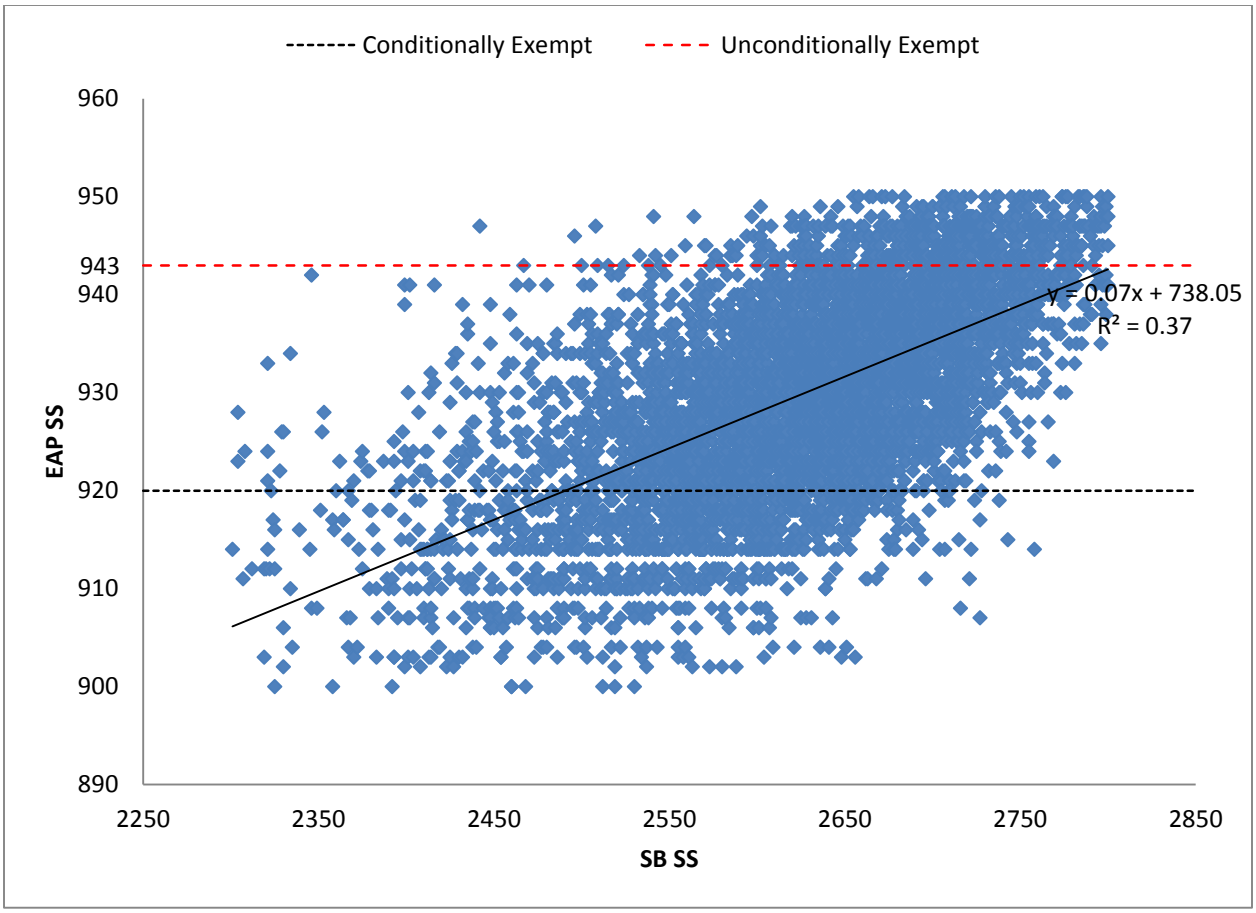

Figure B.3 Scatterplot of the EAP-HSM Scale Scores with the Smarter Balanced FT Scale Scores

The solid line in Figure B.3 represents the linear concordance between the EAP-HSM scale scores and the Smarter Balanced FT scale scores.

Appendix C: Conditional Probabilities

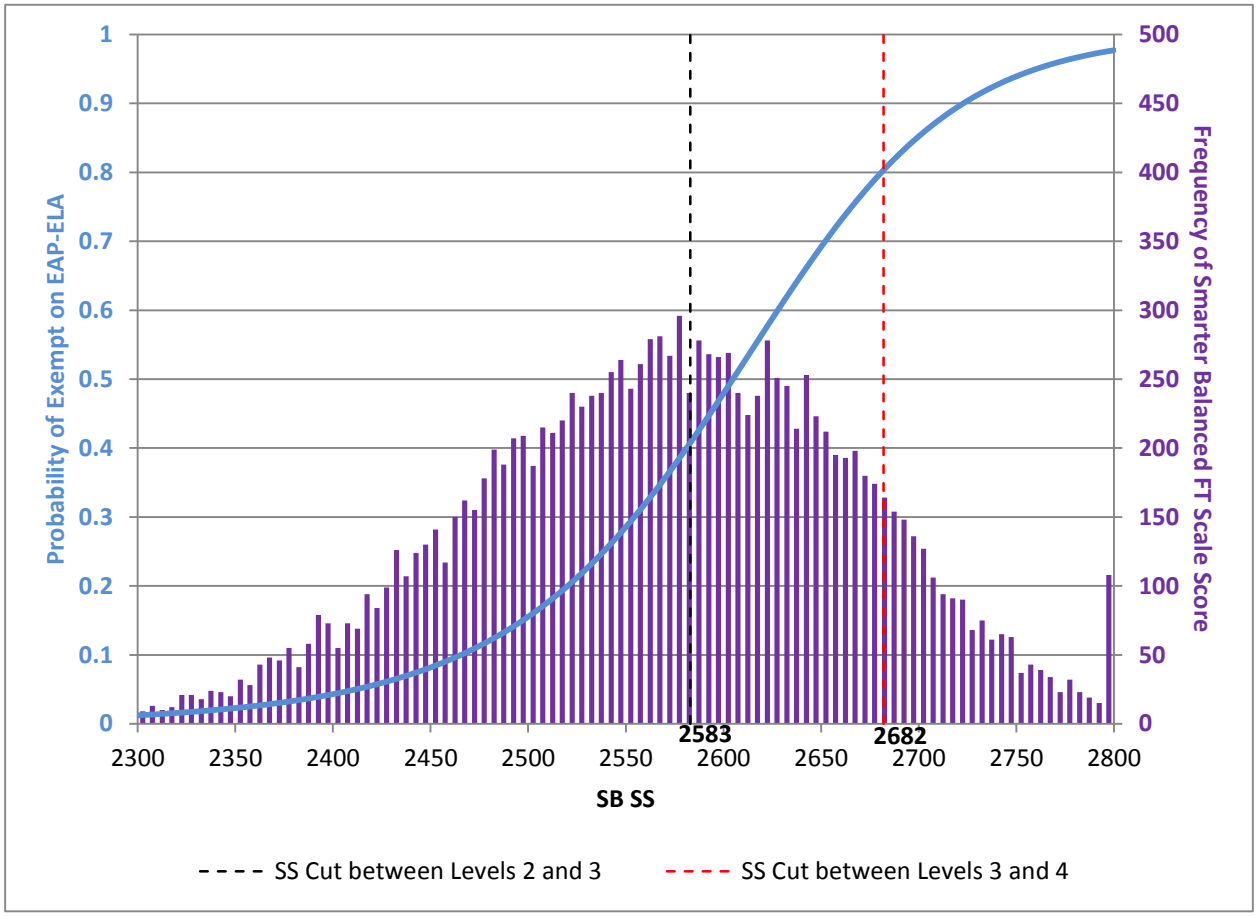

Figure C.1 Conditional Probability of EAP-ELA Exemption Predicted by Smarter Balanced FT Scale Score

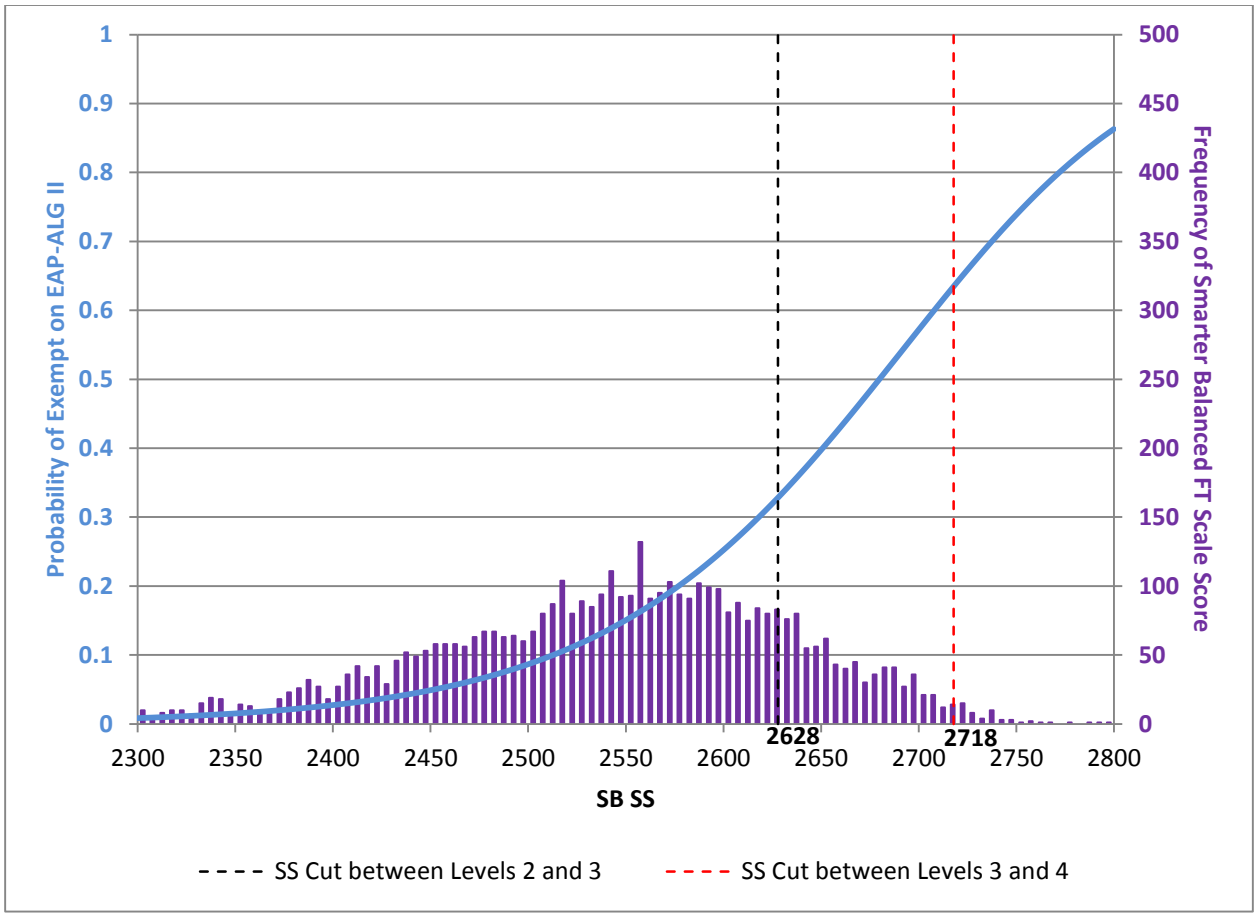

Figure C.2 Conditional Probability of EAP-ALG II Exemption Predicted by Smarter Balanced FT Scale Score

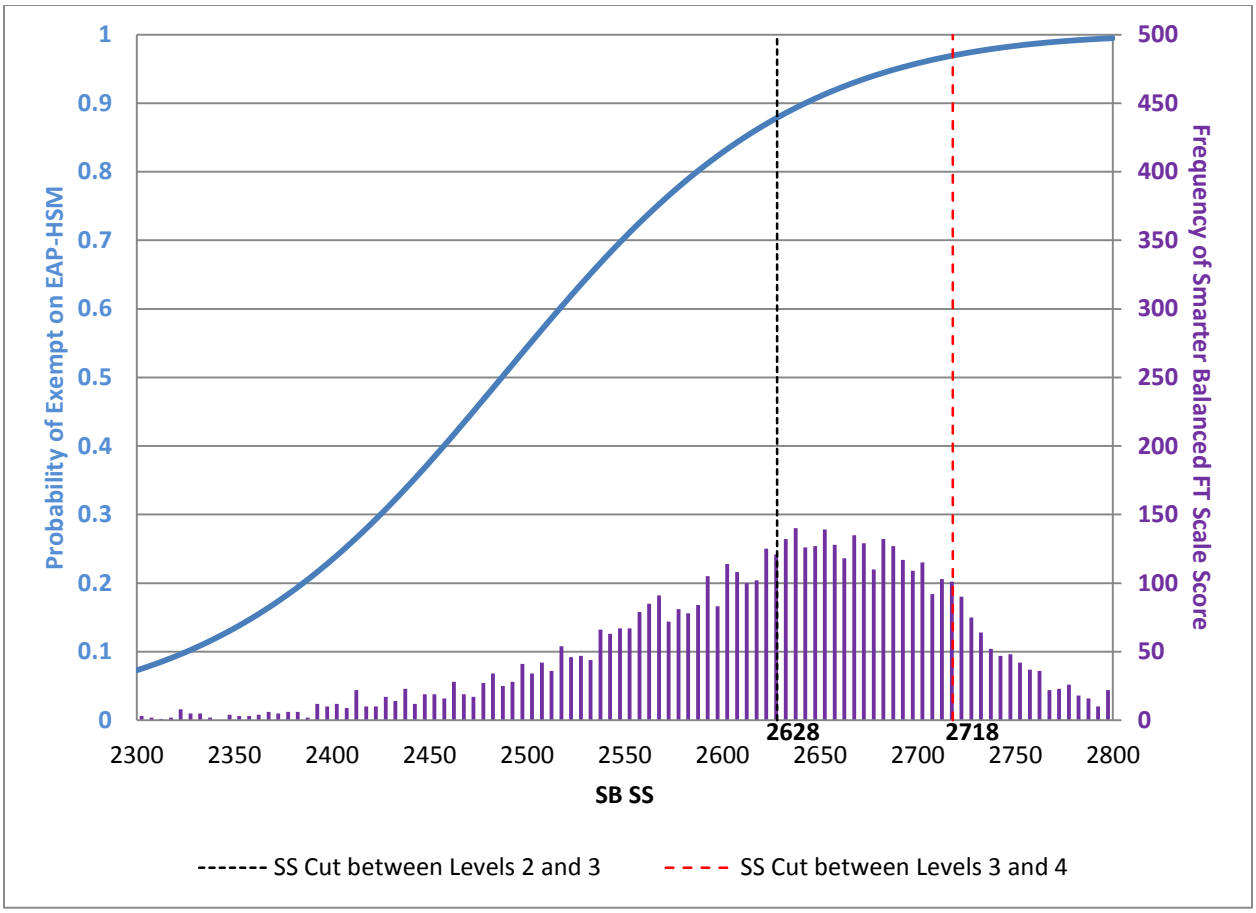

Figure C.3 Conditional Probability of EAP-HSM Exemption Predicted by Smarter Balanced FT Scale Score