NAEP Grade 8 Academic Preparedness Research:

Establishing a Statistical Relationship between the NAEP and EXPLORE® Grade 8 Assessments in Reading and Mathematics for Tennessee Students

Adrienne Sgammato
Mei-Jang Lin
Laura Jerry
David Freund
Rochelle Michel
Nuo Xi
Andreas Oranje

NCES Project Officer: Bill Tirre, Senior Technical Advisor
Governing Board Project Officer: Sharyn Rosenberg,
Assistant Director for Psychometrics

Prepared by Educational Testing Service with funding by the National Assessment Governing Board, under National Center for Education Statistics contract ED-IES-13-C-0017, Task 9, Option 9(C).

Introduction

Starting in early 2003, the National Assessment Governing Board embarked on an ambitious mission to redesign grade 12 assessments and reporting as recommended by the National Commission on 12th Grade Assessment and Reporting. Most importantly, the commission recommended that a state program should be implemented (similar to 4th and 8th grade) and that NAEP should start reporting on the readiness of 12th graders for college, training for employment, and entrance into the military. As a result of the second recommendation, a number of studies were conducted to assess whether and in what ways NAEP could report on *academic preparedness*. To be "academically prepared for college", 12th graders should have the knowledge and skills in reading and mathematics to qualify for placement into entry-level, credit-bearing, non-remedial courses in broad access 4-year institutions and, for 2-year institutions, the general policies for entry-level placement, without remediation, into degree-bearing programs designed to transfer to 4-year institution. After various content alignment studies, judgmental standard setting, secondary analyses, data collections, and statistical linking research (National Assessment Governing Board, 2009), potential benchmarks were identified on the 12th grade Reading and Mathematics scales to indicate what level of performance would correspond to a reasonable probability of being academically prepared for postsecondary education. As a result, a national postsecondary education preparedness percentage could be estimated and reported for the 2013 assessments in Reading and Mathematics. Details about this work can be found on a section of the National Assessment Governing Board website dedicated to preparedness (http://www.nagb.gov/what-we-do/commission.html).

As part of the initial statistical linking research, Florida participated (and continues to participate) at the 12th grade level and was a critical component for the validity evaluation of the benchmarks offering SAT®/ACT® data, Grade Point Averages, and ACCUPLACER® College Placement Exam results as well as longitudinal data into Florida public postsecondary institutions, including Remedial Course Placement and First Year Grade Point Average.

Moving forward, one focus of the second phase of the NAEP academic preparedness research is to study the extent to which grade 8 students are on track for being academically prepared for college by the end of high school. Several states, including Tennessee, participated in the statistical linking research and provided data on students who were part of the NAEP grade 8 sample during the 2012-2013 school year. Some state partners will continue to provide longitudinal data as these students progress through high school and beyond, to be analyzed and reported in future reports.

In this report we will describe the NAEP and EXPLORE® assessments in Reading and Mathematics, discuss the linking methodology (and refer the interested reader to more technical references), and provide the results. A summary will complete this report.

Linking Assessments

The ACT EXPLORE® Assessment

The EXPLORE® test[1] developed by ACT was administered to nearly all 8th graders in Tennessee during the 2012-2013 school year (with the testing window in Sep through Nov, 2012). The assessment includes four multiple-choice tests. Each test measures student's achievement in one of the following four areas: English, Mathematics, Reading, and Science. Students had 30 minutes to finish each test. The number of items in the test varies by subject. For reading and mathematics, both tests have 30 items. EXPLORE® scores provide evidence about the knowledge and skills that students are likely to have in each of the four aforementioned areas. The distribution of item difficulties was selected so that the tests will effectively differentiate among students who vary widely in the level of achievement. A composite score is provided, which is calculated as the average of the four test scores. The individual test scores, as well as the composite score, range from 1 to 25 and are disseminated to students and schools directly. In this study, only the Reading and Mathematics scores were used to link with the NAEP Reading and Mathematics assessments.

The ACT EXPLORE® assessments were designed to assess a specific student's academic progress at the 8th and 9th grade levels, especially with respect to college and career readiness. To help students translate test scores into a clear indicator of their current level of college readiness, ACT derived the ACT College Readiness Benchmarks based on a review of normative data, college admissions criteria, and information obtained through ACT's Course Placement Services. Students who meet a benchmark on the ACT test have approximately a 50% chance of obtaining a B or higher and approximately a 75% chance of obtaining a C or higher in the corresponding credit-bearing first-year college courses (ACT EXPLORE® 2013/2014 Technical Manual, p. 17). In addition, there are corresponding benchmarks for the ACT EXPLORE®, which are linked to the ACT College Readiness Benchmarks. Students who meet a benchmark on the EXPLORE® test have approximately a 50% chance of meeting the ACT Benchmark in the same subject, and are likely to have approximately the same chance of earning a B or better grade in the corresponding college course(s) by the time they graduate high school. The current College Readiness Benchmarks for the EXPLORE® Reading test for grade 8 is 16 and for the EXPLORE® Mathematics test is 17 (ACT EXPLORE® 2013/2014 Technical Manual, p. 17). These benchmarks were used in this investigation. Note that the EXPLORE® reading benchmark was adjusted in 2013. Previously the reading EXPLORE® benchmark was 15. The math EXPLORE® benchmark remained unchanged.

[1] ACT will discontinue the use of the EXPLORE® test after fall 2015 for existing users and no new users are now being accepted.

The National Assessment of Educational Progress (NAEP)

The NAEP test was administered to selected 8th graders in Tennessee during the 2012-2013 school year (with the testing window from the last week of January to the first week of March in 2013). NAEP is the only nationally representative assessment of 4th, 8th, and 12th grade students in public and private schools in the U.S. in a variety of academic subjects. Subjects such as Reading, Mathematics, and Science are also assessed at the state- and even large urban district-level, particularly in grades 4 and 8. Samples of schools and students are selected from a sampling frame in order to produce results that are nationally representative and also representative of participating states and urban districts. Selected students had 50 minutes to complete the cognitive items (i.e., test questions) contained in the NAEP test booklets that were randomly assigned to them. The number and type of items in each booklet vary by subject and by grade. For grade 8 reading, each booklet contains two blocks of about 10 items each. For grade 8 math, each booklet contains two blocks of about 15 items each. A mix of multiple-choice and constructed response items is administered and blocks are systematically paired across booklets (i.e., matrix sampling design). The NAEP assessment is based on broad frameworks developed by the National Assessment Governing Board. By law, no student or school results are estimated or reported using the NAEP assessment. In fact, the assessment is designed in a way that no reliable score *can* be computed at the student level while minimizing the burden of any individual student selected to participate in the assessment. Instead, the main objective of NAEP is to report on the achievement of policy-relevant population groups, estimated directly using marginal estimation latent regression methods. For a comprehensive description of NAEP estimation procedures, the reader is referred to Mislevy, Beaton, Kaplan, & Sheehan (1992).

For the linking study, this requires that the relationship between NAEP and other measures (e.g., EXPLORE® scores) must be directly estimated using this latent regression methodology since there are no appropriate student-level scores available. In the methodology section we will discuss some of the steps that were required to complete this part of the research. NAEP reports results on scales that range from 0 to 500 in grade 8 Reading and Mathematics and the goal is to express the aforementioned ACT EXPLORE® benchmarks in terms of these scales. Students sampled for participation in NAEP are assessed in only one assessment subject. Consequently, each student in the matched or linking sample had EXPLORE® scores in both reading and mathematics, but results for only one NAEP assessment, either reading or mathematics.

Linking

When linking scales of different assessments, it is important to be precise about what that exactly entails. Usually, the two instruments under a linking study do not measure the same construct and have not been designed for that purpose, but generally there is some overlap. The greater the overlap, as evidenced by a higher correlation between the two scales, the more confident we can be that the instruments can be used to predict each other well. When the relationship is very strong and

the instruments have a similarly high reliability, we would be able to claim that the two scales are largely interchangeable and, therefore, that there is a one-to-one relationship between scores on the one scale and scores on the other scale. When this relationship is moderate, then we can do a 'best' projection of one scale onto the other or the reverse, which would not necessarily lead to similar results. In that case, the outcome would be of a probabilistic nature (e.g., "at score level X, students have a reasonably high probability to be prepared"). In the case of the preparedness linking studies, and taking past studies into account, a moderate relationship is most probable. We will elaborate further on this in subsequent sections.

Typically, a content alignment precedes statistical alignment to assess the extent to which the instruments were designed to measure the same or different constructs. Content alignment studies between NAEP and EXPLORE® Reading and Mathematics are being conducted by the National Opinion Research Center (NORC) at the University of Chicago (under contract ED-NAG-14-C-0002 with the National Assessment Governing Board) and will provide an important context for the statistical linking results presented here.

Methodology

In this section we will discuss the data and the linking methodology. The purpose is to give the reader some insight into the procedures that were followed and, therefore, the opportunity to evaluate the results within that context.

Data

This study used data from students who were sampled and assessed in NAEP 8th grade reading or mathematics in 2013 and had also taken the EXPLORE® assessment. From late January through early March of 2013, NAEP assessments in reading and mathematics were administered to samples of 8th grade students that were representative of each state, and together of the nation. As a result, about 2,700 public school students in Tennessee were sampled for each subject. Sample sizes are rounded to the nearest hundred as required in the NCES Statistical Standards (https://nces.ed.gov/statprog/2002/stdtoc.asp). Because only a sample is assessed and for efficiency purposes schools are sampled proportionally to size (in addition to other adjustments), sampling weights have to be used to appropriately represent all student groups of interest and, consequently, calculate unbiased results. The EXPLORE® assessment is required in Tennessee at the 8th grade level, meaning that almost all students who were sampled for NAEP also participated in EXPLORE® and have associated scores. The reverse is obviously not true, given that NAEP is sample-based (i.e., not every student who participated in EXPLORE® also participated in NAEP).

The process of matching EXPLORE® scores to NAEP participants was carried out through an agreement between the National Assessment Governing Board and the National Center for

Education Statistics (NCES) to have NAEP contractors Westat and ETS conduct the preparedness research work. In addition, data confidentiality agreements were established between all parties involved and the Tennessee Department of Education. A process for matching the student records was developed to protect students' identity and confidentiality. Confidentiality of state supplied scores (e.g., EXPLORE® scores) was assured through the assignment of a pseudo ID for students taking that assessment and using that pseudo ID as a way to transfer scores to ETS *without* the need to include Personally Identifiable Information (PII) such as names or birthdates. Similarly, the pseudo ID was appended to NAEP files by Westat who then provided that file to ETS, again *without* any PII. Via the pseudo ID, ETS subsequently matched EXPLORE® scores to NAEP files. In the case of Tennessee, EXPLORE® scores were matched at 93% for reading and 94% for mathematics, which is extraordinarily high. The matching rates for various student subgroups (by gender, by race/ethnicity, etc.) were at or above 91%. Table 1 provides weighted percentages by gender and race/ethnicity for the matched sample and overall match rates.

Table 1. Weighted percentages by gender and race of the Tennessee linking samples

Reading								
	White	*Black*	*Hispanic*	*Asian*	*American Indian /Alaskan Native*	*Pacific Islander*	*2+ races*	***Total*[2]**
Male	36%	10%	3%	1%	#[1]	#	#	**51%**
Female	34%	11%	3%	1%	#	#	#	**49%**
***Total*[2]**	**71%**	**21%**	**6%**	**1%**	**#**	**#**	**1%**	**100%**
							Overall Match Rate	**93%**
Mathematics								
	White	*Black*	*Hispanic*	*Asian*	*American Indian /Alaskan Native*	*Pacific Islander*	*2+ races*	***Total*[2]**
Male	36%	11%	3%	1%	#	#	#	**51%**
Female	35%	10%	3%	1%	#	#	#	**49%**
***Total*[2]**	**71%**	**21%**	**6%**	**2%**	**#**	**#**	**#**	**100%**
							Overall Match Rate	**94%**

NOTES: [1]# Rounds to zero.
[2] Detail may not sum to totals because of rounding.

Given the fact that the two assessments that are linked have very different purposes and, possibly, different stakes, an outlier analysis is in order. For instance, if there are participants that scored very high on a *higher* stakes test (i.e., EXPLORE® test) and very low on the *lower* stakes test, the low performance can be reasonably attributed to motivation rather than performance level. Such cases would be considered 'outliers' and removed from further analyses. An initial examination of the joint

distribution of NAEP and EXPLORE® revealed very few potential outlier cases. After this more cursory inspection, standardized residuals from robust regression (Huber, 1973) were used to identify approximately 0.6% of cases in reading and approximately 0.8% of cases in mathematics (cases with absolute standardized residuals greater than 3 were considered outliers and removed). We refer to Huber (1973) for details about the procedure and the criteria applied. These outliers were excluded from the final linking samples and were not used in subsequent analyses.

Analysis Approach

After preparatory data identification, matching, merging, and data reconciliation, the linking analyses were conducted. The current study was designed to pursue five[2] specific analysis questions that guide the choices in methodology for the linking and validation:

1) What are the correlations between the grade 8 NAEP and EXPLORE® scores in reading and mathematics?
2) What scores on the grade 8 NAEP reading and mathematics scales correspond to the EXPLORE® benchmarks?
3) What are the average grade 8 NAEP reading and mathematics scores (and the difference between the 75th and 25th percentiles) and the IQR for students below, at, and above the EXPLORE® benchmarks?
4) What scores on the EXPLORE® reading and mathematics scales correspond to the grade 8 NAEP *Proficient* cut scores in reading and mathematics?
5) What are the average EXPLORE® reading and mathematics scores and inter quartile ranges (IQRs) for students below and at or above the NAEP *Proficient* cut score?

We will describe pertinent methodological details about the analysis followed by the results of the analyses in the final section. The key steps of the analysis are (a) estimating the correlation between NAEP and EXPLORE®, which includes use of the aforementioned latent regression methodology (b) determining the appropriate methodology for linking based on those correlations (c) applying procedures to effectively estimate cumulative probability functions and (d) calculating impact data as part of the results.

A satisfactory treatment of the latent regression methodology is outside the scope of this report and the interested reader is referred to Mislevy, Beaton, Kaplan, and Sheehan (1992). The basic notion is that NAEP measures constructs that are represented on item response theory based latent scales, which are not measured reliably at the student level. However, pertinent data from students in specified groups of interest can be pooled to estimate reliable scores at the group level. EXPLORE® scores, on the other hand, are reliably estimated at the individual level and can be treated as a set of

[2] Research questions 4) and 5) were added by the Committee on Standards, Design and Methodology (COSDAM) at the August 2015 NAGB meeting, in order to explore the relationship between grade 8 NAEP and EXPLORE® scores in reading and mathematics in a reverse direction.

consecutive (semi-continuous) groups. Correlations between NAEP and EXPLORE® can be directly estimated at the overall level and the result showed that the (true score) correlation for reading is 0.73 and for mathematics is 0.81. While these are not low correlations, they do suggest that there is enough uncertainty in the relationship that a direct one-to-one correspondence of scale score points is not advisable.

To elaborate on that observation and as briefly introduced earlier, different classes of statistical relationships can be established between various tests, and the distinctions correspond to the extent to which the tests are similar with respect to the constructs measured, populations, and measurement characteristics of the tests (Feuer, Holland, Green, Bertenthal, & Hemphill, 1999; Holland & Dorans, 2006). In this study, two types of statistical linking were originally considered: concordance and projection. Concordance establishes a score linkage between two tests by matching the corresponding score distributions. The claims that can be made based on concordance are also commensurately strong. Essentially, the claim is made that a score *x* on NAEP exactly corresponds to a score *y* on EXPLORE® and vice versa. Projection is a less stringent type of correspondence in which scores on one test are related, typically via a linear or nonlinear regression, to a conditional distribution of scores on the other test. Projection relationships are not symmetric, and do not assume or result in a one-to-one correspondence. The claim is made that a score of *x* on NAEP corresponds to the proportion *p* of students attaining the benchmark score of *y* or higher on EXPLORE®. Subsequently, a choice for *p* has to be made, where a more conservative claim requires a higher *p*. This means that if one wants to have a very high degree of confidence that students at a certain NAEP score pass the benchmark, then a relatively high *p* has to be set, a relatively high score level is identified, and, likely, the percent of students that actually pass the benchmark is under-estimated. The reverse is true when a lower degree of confidence is acceptable. Needless to say, concordance assumes and requires a much stronger relationship than projection.

The relationships between NAEP and EXPLORE® reading (r =0.73) and mathematics (r = 0.81) are not sufficiently strong to support concordance, given that a generally accepted minimum correlation for concordance is r = 0.866 (Dorans, 1999; Dorans & Walker, 2007). Consequently, projection was used in this study. As mentioned before, typically a smoothing process is applied in order to produce more accurate probability distributions, particularly when the underlying population distribution of test scores may contain irregularities (Moses & Liu, 2011), for example due to a non-continuous nature of the scale. Bivariate loglinear smoothing (Holland & Thayer, 2000) was applied to the joint NAEP-EXPLORE® distributions[3].

[3] For reading, as part of the loglinear smoothing procedure we preserved the first 3 moments for the NAEP distribution, 4 moments for the EXPLORE® distribution, and 4 cross-moments. For math, we preserved the first 6 moments for the NAEP distribution, 6 moments for the EXPLORE® distribution, and 4 cross-moments. These loglinear smoothing models mostly resulted in the smallest value of the Akaike Information Criterion (AIC) statistic (Moses & von Davier, 2006), although model complexity and sample size was also taken into consideration.

An important tool for evaluating statistical links between tests is sensitivity analysis, which is intended to examine the extent to which the linking relationship is invariant across key student groups, such as gender and race/ethnicity groups. These analyses require a minimum sample size[4] in order to produce reliable comparisons. For the Tennessee linking samples, both gender groups met that criterion. For the race/ethnicity groups, only White and Black student subgroups met the criterion. Separate linking functions were established for these subgroups and deviations from the overall linking function indicated violation of invariance. It should be noted though that the purpose of this linking is to establish a specific benchmark for preparedness. In that sense, substantial variability across student groups for parts of the scale that does not entail the benchmark could be quite harmless. For NAEP reading, no substantial deviation from the overall linking function was detected for Male, Female, or White student subgroups. The linking function for Black students was slightly lower than the overall linking function. For NAEP math, no substantial deviation from the overall linking function was detected for Female or White student subgroups. The linking functions for Male and Black students were slightly lower than the overall linking function. Even though the comparison between the linking functions indicated some variance among different subgroups, the difference was not large enough to discredit the linking study. In fact, it should be emphasized that some subgroups considered here had a much smaller sample size than the overall linking sample, and therefore the difference observed between the linking functions should be interpreted with great caution.

Finally, for both reading and mathematics, the probabilities from the smoothed joint distributions were used to create projections tables containing conditional cumulative distributions of NAEP proficiencies for EXPLORE® scores. The range of possible NAEP scores below, at, and above the EXPLORE® benchmark (16 on the EXPLORE® reading scale and 17 on the EXPLORE® mathematics scale) were estimated and, subsequently, for each subject area the projected conditional distributions were used to identify the NAEP scale scores associated with the EXPLORE® benchmarks. In addition, the direction of the linking relationship was reversed and the point on the EXPLORE® measure that corresponds most closely to the NAEP *Proficient* cut score was identified using the conditional cumulative distributions of the EXPLORE® scores for the NAEP proficiencies. We will discuss the results of the linking study in the following section.

[4] The minimum was set at 500 as a rule of thumb, but based on the idea that there is at least one observation below -3 and above +3 standard deviations (in a standard normal distribution) in expectation.

Results

EXPLORE® benchmarks projected on the NAEP scale

The most important result, following the second and third analysis questions, is to determine what scores on the NAEP reading and mathematics scales correspond to the EXPLORE® benchmarks. In other words, what would be the 'on track to be prepared' score level on NAEP that corresponds most reasonably to an established 'on track' benchmark.

Table 2 provides descriptive statistics to get an initial sense of where the benchmark most likely will be located as well as some distributional properties as context to these results. The average scores and percentile estimates for students below, at, and above the EXPLORE® benchmarks are spread out, though more so for students below the benchmark than above. Note that the mean *at* the benchmark is not necessarily the same as the NAEP score equivalent for the benchmark, but rather a characterization of the students at this level. Also note that these results are based on the statistical linking (i.e., projection methodology).

Table 2: Descriptive NAEP Statistics for Students Below, At, or Above the EXPLORE® Benchmarks

Subject	EXPLORE® Benchmark	Mean	Percentage[2]	SD	Percentile		IQR[1]
					25th	*75th*	
Reading	*Below*	252	64%	28	234	271	37
	At	279	8%	22	264	293	29
	Above	297	27%	21	282	311	29
Mathematics	*Below*	264	65%	28	247	284	37
	At	294	12%	18	282	306	24
	Above	317	23%	22	301	331	30

NOTES: [1]IQR is the Inter Quartile Range or the difference between the 75th and 25th percentiles.
[2] Detail may not sum to totals because of rounding.

To determine the NAEP scale score point that most reasonably corresponds to the EXPLORE® benchmarks, it is most illustrative to graphically represent the relationship. Figures 1 and 2 show the relationship based on statistical projection for students at the respective benchmarks. The black curved line shows the proportion of students meeting the EXPLORE® benchmark for pertinent score levels on NAEP. Colored vertical lines indicate where the NAEP achievement levels are located. Finally, and as mentioned before, a proportion level has to be chosen commensurate the confidence required to indicate whether students have passed the benchmark or not. A red dotted line shows at which point students are more likely to have reached the benchmark than not (i.e., the probability is set at 0.50). Given the moderate relationships between the two scales, this seems a reasonable

location for indicating sufficient chance to be 'on track to preparedness'. For context, a secondary, lighter red line indicates when the probability is set at 0.80, indicating a relatively high level of confidence that students have attained the EXPLORE® benchmark.

From the graphs it can be deducted that the location where students have a reasonable probability to be on track for reading could be set at a NAEP scale score of 284, slightly above the *Proficient* achievement level. The mathematics counterpart could be set at 296, slightly below the *Proficient* achievement level.

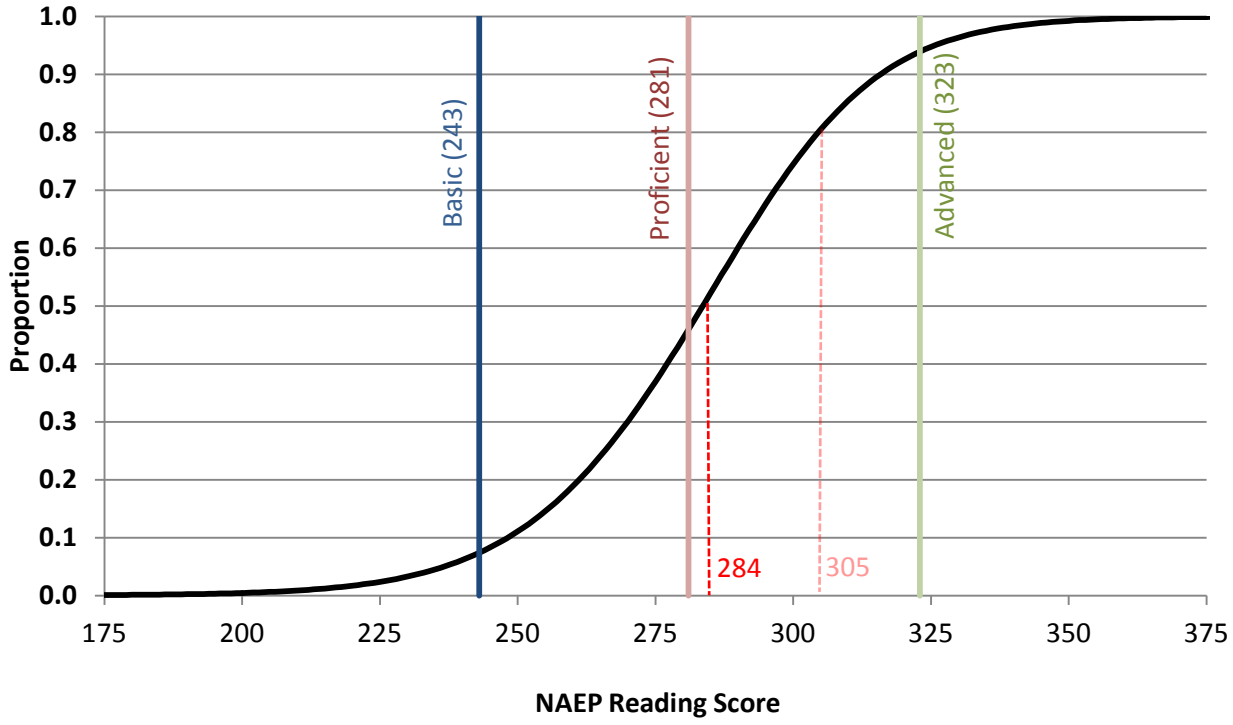

Figure 1: Proportion of students meeting the Reading EXPLORE® benchmark of 16 in Tennessee for NAEP Reading levels

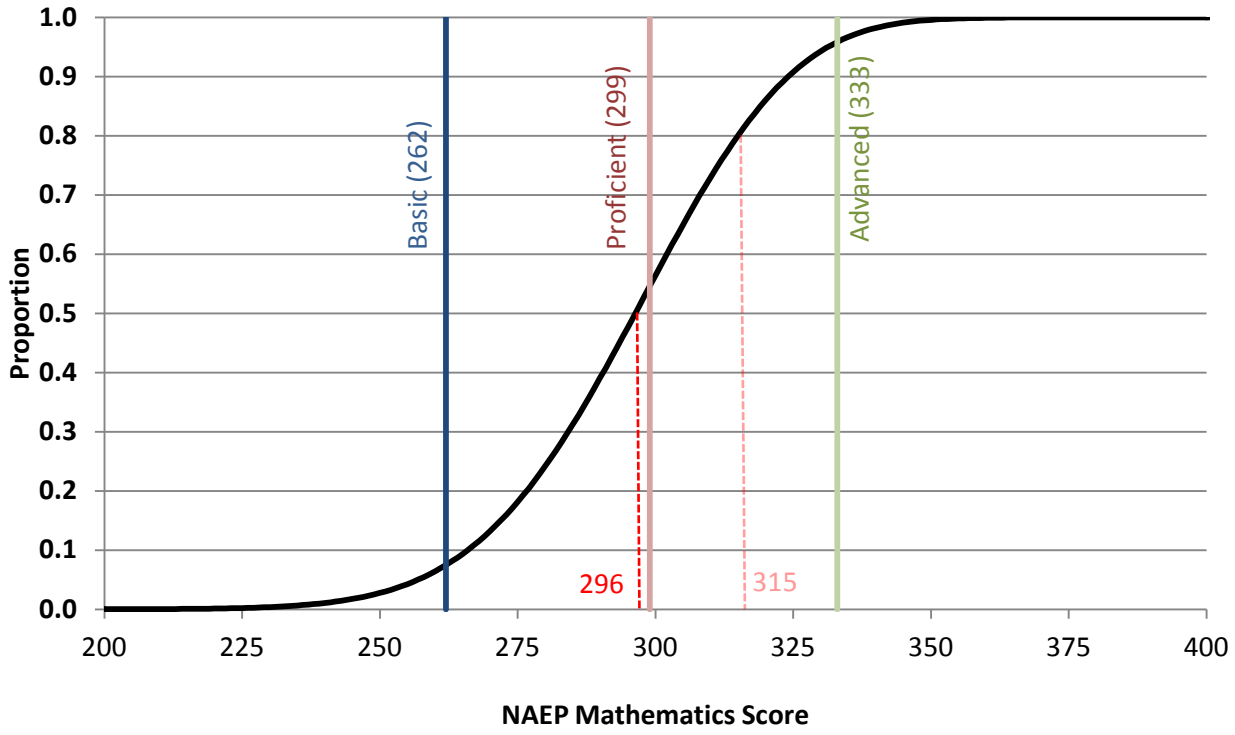

Figure 2: Proportion of students meeting the Mathematics EXPLORE® benchmark of 17 in Tennessee for NAEP Mathematics levels

Impact

Now that potential points have been identified, it is important to show what percentage of students in Tennessee are deemed to have a reasonable probability (i.e., the probability set at 0.50) of being on track in grade 8 across various student groups. Table 3 provides those percentages, based on the potential points identified on the NAEP scales, as well as the EXPLORE® benchmarks. Table 3 indicates that overall about 31 to 36 percent of students are on track, but the results differ across different subgroups. No significance testing has been conducted to compare these percentages and, therefore, no comparative statements will be made.

Table 3: Percentage of the Tennessee linking samples that have a reasonable probability to be on track to be academically prepared based on the potential points identified on the NAEP scale, compared to the percentage of the same sample meeting the Reading EXPLORE® benchmark of 16 and Mathematics EXPLORE® benchmark of 17.

	Reading		Mathematics	
Student Group	**NAEP ≥ 284**	**EXPLORE® ≥ 16**	**NAEP ≥ 296**	**EXPLORE® ≥ 17**
Total	31%	33%	32%	36%
Male	27%	31%	33%	35%
Female	34%	35%	31%	36%
White	35%	38%	37%	41%
Black	15%	17%	12%	18%
Hispanic	26%	26%	26%	29%

NAEP *Proficient* cut scores projected on the EXPLORE® scale

To answer the fourth and fifth research questions, we find the point on the EXPLORE® measure that corresponds most closely to the NAEP *Proficient* cut score, essentially reversing the direction of the linking relative to the previous analyses. Table 4 provides descriptive statistics of the EXPLORE® reading and mathematics scores for students below and at or above the grade 8 NAEP *Proficient* achievement level. The grade 8 NAEP *Proficient* level cut score was set at 281 for reading and 299 for mathematics.

Table 4: Descriptive EXPLORE® Statistics for Students Below, At or Above the Grade 8 NAEP Proficient Level.

Subject	NAEP *Proficient*	Mean	Percentage[2]	SD	Percentile		IQR[1]
					25th	*75th*	
Reading	*Below*	13	65%	3	11	14	3
	At or Above	18	35%	4	15	20	5
Mathematics	*Below*	14	71%	2	12	15	3
	At or Above	19	29%	3	16	20	4

NOTES: [1]IQR is the Inter Quartile Range or the difference between the 75th and 25th percentiles.

Following the same methodology of statistical projection (see Figures 3 and 4) we identified an EXPLORE® reading score of 16 and a mathematics score of 17 as cut points. These points coincide with the College Readiness Benchmarks for EXPLORE® reading and mathematics tests for grade 8 students.

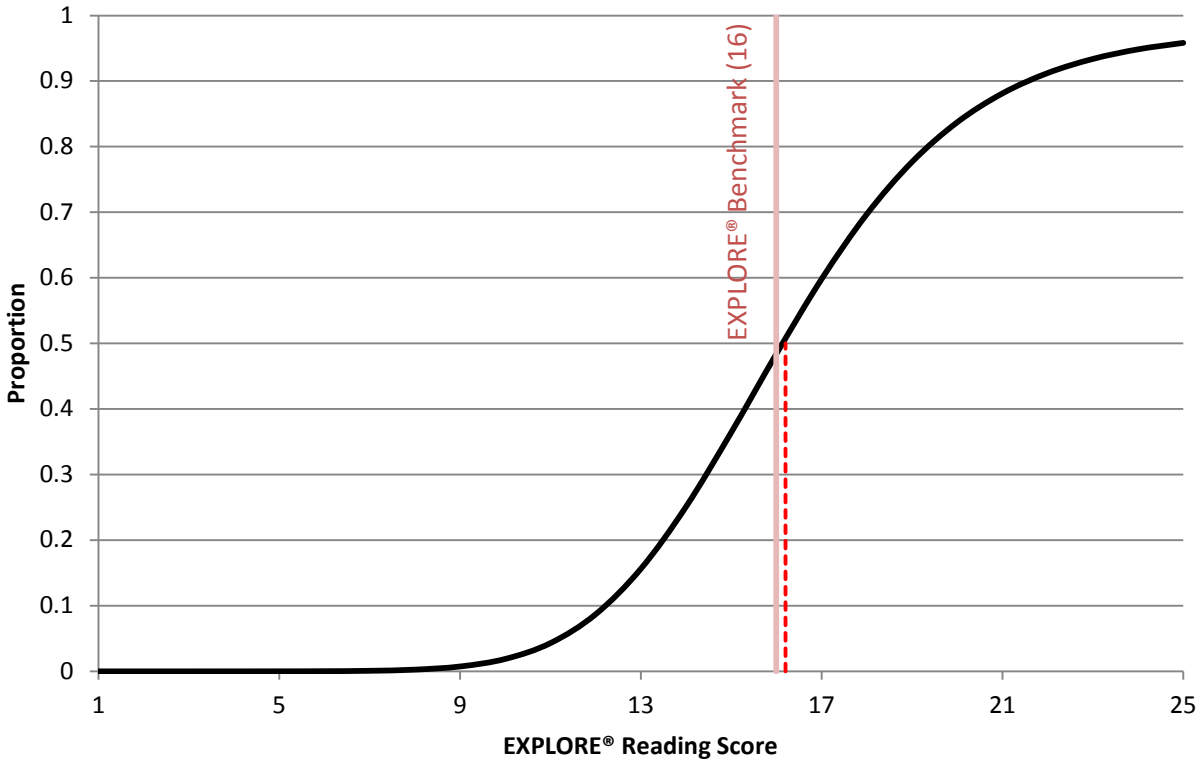

Figure 3: Proportion of students meeting the NAEP Reading Proficient achievement level of 281 in Tennessee for EXPLORE® Reading scores

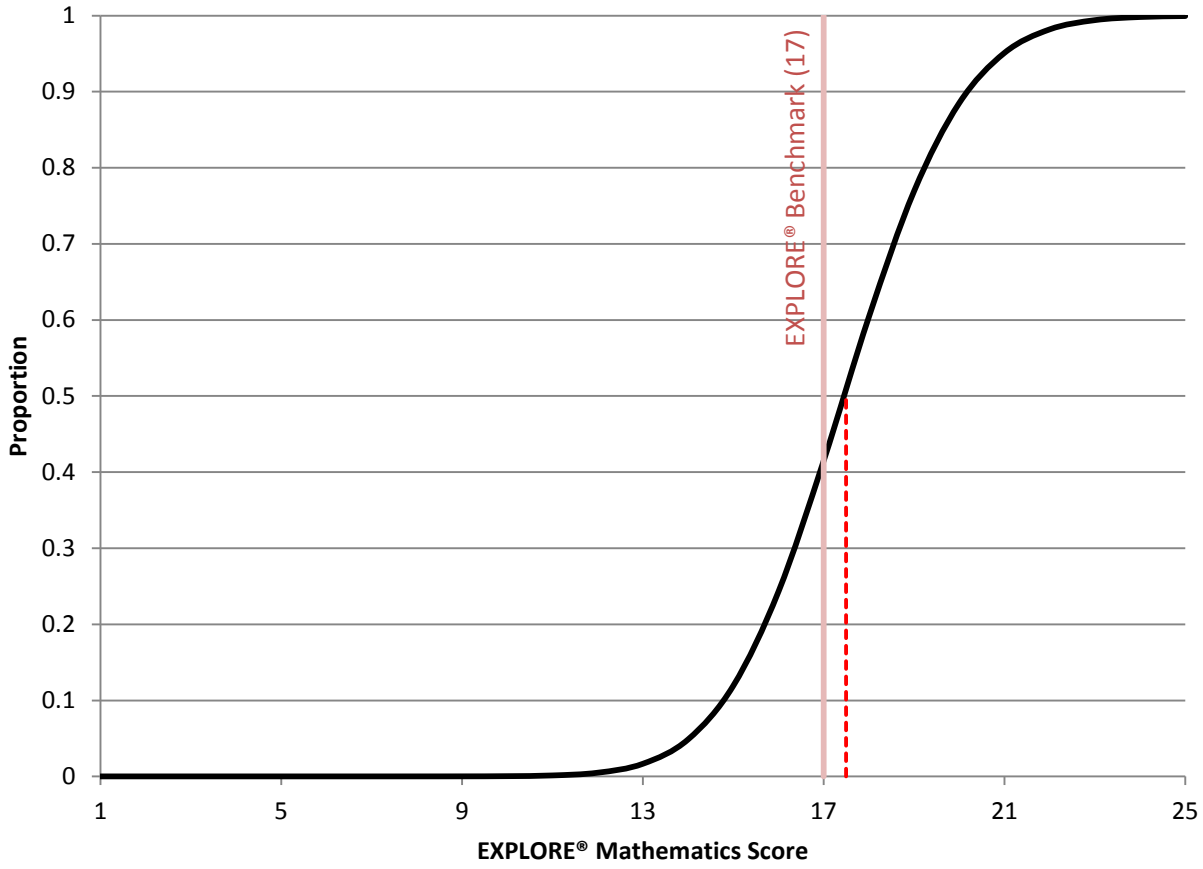

Figure 4: Proportion of students meeting the NAEP Mathematics Proficient achievement level of 299 in Tennessee for EXPLORE® Mathematics scores

Summary

The goal of this study was to statistically relate NAEP and EXPLORE® and use that relationship to identify a reference point or range on the NAEP 8th grade reading and mathematics scales reasonably associated with ACT's preparedness benchmarks on the EXPLORE® reading and mathematics measures. Identifying such points would potentially allow NAEP to report on the percentage of students at 8th grade who are on track to be prepared for college for the nation and for states. The first step involves three participating states, including Tennessee, who have graciously provided the critical EXPLORE® data necessary to calculate the relationship with NAEP. In this study, various statistical techniques, including latent regression, smoothing, and statistical projection were used to establish the relationship and identify potential markers on the NAEP scale that could form the basis

for 'on track to preparedness' reporting (see Figures 1 and 2 for examples of how the markers were determined).

In addition, we identified the point on the EXPLORE® measure that corresponds most closely to the NAEP *Proficient* achievement level cut score, for grade 8 reading and mathematics scales, in order to explore the relationship between the two measures in the reverse direction (see Figures 3 and 4 for the linking results).

A key finding was that the relationship between the two scales is moderate, meaning that the kind of relational statements that can be made need to be presented in terms of probability rather than direct one-to-one relationships. This is not surprising because the instruments are not intended to measure the exact same construct. However, it does make interpretation somewhat more challenging. The results showed that the College Readiness Benchmarks for EXPLORE® and the NAEP *Proficient* achievement level cut scores correspond well to each other in both linking directions. In particular, NAEP scale score points near the *Proficient* achievement levels could form a reasonable basis for reporting 'on track for preparedness'. Approximately 31% of Tennessee 8th graders met that criterion for reading and 32% met the criterion for math. On the other hand, the projection results of the NAEP *Proficient* cut score on the EXPLORE® scale are very close to the existing EXPLORE® benchmarks for reading and mathematics. Further content alignment work, which is conducted independently from this study, should provide further context to these results.

References

ACT EXPLORE Technical Manual 2013/2014. (http://www.act.org/explore/pdf/TechManual.pdf)

Dorans, N. J. (1999). *Correspondences between ACT and SAT I scores* (Research Report No. 99-2). Princeton, NJ: Educational Testing Service.

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.

Feuer, M.J., Holland, P.W., Green, G.F., Bertenthal, M.W., & Hemphill, F.C. (1999). *Uncommon measures: Equivalence and linkage among educational tests* (Report of the Committee on Equivalency and Linkage of Educational Tests, National Research Council). Washington, DC: National Academy Press.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 187-220). Washington, DC: American Council on Education.

Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, *25*, 133-183.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, *1*, 799-821.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29* (2), 133-161.

Moses, T.P., & Liu, J. (2011). *Smoothing and Equating Methods Applied to Different Types of Test Score Distributions and Evaluated With Respect to Multiple Equating Criteria* (Research Report No. 11-20). Princeton, NJ: Educational Testing Service.

Moses, T. P., & von Davier, A. A. (2006). *An SAS macro for loglinear smoothing: Applications and implications* (Research Report No. 06-05). Princeton, NJ: Educational Testing Service.

National Assessment Governing Board (2009). *Making New Links, 12th Grade and Beyond: Technical Panel on 12th Grade Preparedness Research Final Report*.