September 2016

NAEP Grade 12 Academic Preparedness Research:

Establishing a Statistical Relationship between the NAEP and ACT Assessments in Reading and Mathematics for Grade 12 Tennessee Students

Nuo Xi
Mei-Jang Lin
Laura Jerry
David Freund
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 (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*. The Governing Board's working definition of academic preparedness for college is the knowledge and skills in reading and mathematics needed 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, scale scores of 302 on the NAEP grade 12 reading assessment (equivalent to the *Proficient* cut score) and 163 on the NAEP grade 12 mathematics assessment (between the *Basic* cut score of 141 and the *Proficient* cut score of 176) were identified to project a reasonable probability of being academically prepared for college. As a result, the percentage of 12th grade students in the U.S. who were academically prepared for college was estimated and reported for the 2013 and 2015 assessments in reading and mathematics. Extensive details about this work can be found on a section of the National Assessment Governing Board website dedicated to preparedness (https://www.nagb.org/what-we-do/preparedness-research.html).

As part of the first phase of the Governing Board's preparedness research, Florida participated in the research by providing (via a data sharing agreement) longitudinal data that could be linked to 2009 NAEP grade 12 performance in reading and mathematics. These data were 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.

In the current (second) phase of the Governing Board's academic preparedness research, additional state partners have agreed to provide longitudinal data that can be linked to the 2013 NAEP reading and mathematics assessments at grades 8 and 12. Tennessee, as one of the state partners, participated in the state-level statistical linking research connecting NAEP and ACT and provided data on students who were part of the NAEP grade 12 sample during the 2012-2013 school year, as well as their ACT data. 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 ACT 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 Assessment

The ACT® test was administered to almost all 11th graders in Tennessee in the spring of 2012. It is a curriculum- and standards-based assessment that measure students' academic readiness for college (http://www.act.org/content/act/en/products-and-services/the-act.html). 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. The testing time and the number of items in the test vary by subject. For reading, students have 35 minutes to finish 40 multiple-choice items. For math, the test has 60 multiple-choice items, and students have 60 minutes to finish. 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 36 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 tests were designed to measure students' knowledge and skills needed for first-year college success. 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 (https://www.act.org/content/act/en/education-and-career-planning/college-and-career-readiness-standards/benchmarks.html). The College Readiness Benchmarks for the ACT reading test is 22 and for the ACT mathematics is also 22 (ACT, 2013). These benchmarks were used in this investigation.

The National Assessment of Educational Progress (NAEP)

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. The NAEP test was administered to a representative sample of 12th graders in Tennessee public schools during the 2012-2013 school year (with the testing

window from the last week of January to the first week of March in 2013). 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 12 reading, each booklet contains two blocks of about 10 items each. For grade 12 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 (Mislevy, Beaton, Kaplan, & Sheehan, 1992). For a comprehensive description of NAEP estimation procedures, the reader is referred to Mislevy et al. (1992).

For the linking study, this requires that the relationship between NAEP and other measures (e.g., ACT 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 12 reading and from 0 to 300 in grade 12 mathematics, and the goal is to express the aforementioned ACT benchmarks in terms of these scales. Students sampled for participation in NAEP are assessed in only one subject. Consequently, each student in the matched or linking sample had ACT 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 content 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. It serves as the foundation for most of the preparedness research, especially for the statistical relationship studies. The content alignment studies between NAEP and ACT reading and mathematics were conducted by ACT in 2009, under subtask 4.3 of contract ED-06-CO-0098 with the National Assessment Governing Board. The studies found similar content in NAEP and ACT, and the content overlap was more extensive in mathematics than in reading (https://www.nagb.org/what-we-do/preparedness-research/types-of-research/content-alignment.html).

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 12th grade reading or mathematics in 2013 and had also taken the ACT. From late January through early March of 2013, NAEP assessments in reading and mathematics were administered. Thirteen states participated in the pilot state assessment at grade 12, including Tennessee. About 3,000 and 3,200 students at grade 12 were assessed in reading and mathematics, respectively, in Tennessee. 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 ACT assessment was required in Tennessee at the 11th grade level, meaning that almost all students who were sampled for NAEP also participated in ACT and have associated scores. The reverse is not true, given that NAEP is sample-based (i.e., not every student who participated in ACT also participated in NAEP). Notice that the two tests were not administered concurrently. There could be a nine- to eleven-month time span between the state-wide ACT administration (spring of 2012) and the NAEP administration (first quarter of 2013).

The process of matching ACT 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., ACT 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 ACT scores to NAEP files. In the case of Tennessee, ACT scores were matched at 89% for reading and 90% for mathematics, which is very high. The matching rates for various student subgroups (by gender, by race/ethnicity, etc.) were at or above 81%. Table 1 provides weighted percentages by gender and race/ethnicity for the matched sample and overall match rates. The matched samples appear to be NAEP representative. In terms of ACT, the weighted average ACT reading and math scores of the matched sample are very close to the average ACT scores of the Tennessee graduating class 2013, which are released in the ACT Profile Report (https://forms.act.org/newsroom/data/2013/pdf/profile/Tennessee.pdf).

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%	2%	1%	#[1]	#	#	**49%**
Female	35%	13%	2%	1%	#	#	#	**51%**
***Total*[2]**	**70%**	**23%**	**4%**	**2%**	**#**	**#**	**1%**	**100%**
							Overall Match Rate	**89%**
Mathematics								
	White	*Black*	*Hispanic*	*Asian*	*American Indian /Alaskan Native*	*Pacific Islander*	*2+ races*	***Total*[2]**
Male	36%	10%	2%	1%	#	#	#	**50%**
Female	35%	12%	2%	1%	#	#	#	**50%**
***Total*[2]**	**71%**	**22%**	**4%**	**2%**	**#**	**#**	**1%**	**100%**
							Overall Match Rate	**90%**

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 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., ACT 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 ACT revealed very few potential outlier cases. After this more cursory inspection, standardized residuals from robust regression (Huber, 1973) were used to identify approximately 1.3% of cases in reading and approximately 1.4% 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 three specific analysis questions that guide the choices in methodology for the linking and validation:

1) What are the correlations between the grade 12 NAEP and ACT scores in reading and mathematics?
2) What scores on the grade 12 NAEP reading and mathematics scales correspond to the ACT benchmarks?
3) What are the average grade 12 NAEP reading and mathematics scores and IQRs (i.e., the difference between the 75th and 25th percentiles) for students below, at, and at or above the ACT benchmarks?

Questions 2) and 3) have been specified in one particular direction to estimate an academic preparedness cutpoint on the NAEP scale. Conversely and as a complement to these questions, the same analyses can be conducted in the opposite direction to verify: 2*) what scores on the ACT reading and mathematics scales correspond to the grade 12 NAEP *Proficient* cut scores in reading and mathematics and 3*) what the average ACT reading and mathematics scores and IQRs are for students below and at or above the NAEP *Proficient* cut scores.

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

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. ACT scores, on the other hand, are reliably estimated at the individual level and can be treated as a set of consecutive (semi-continuous) groups. Correlations between NAEP and ACT 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 0.83. 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 ACT 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 ACT. 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 ACT reading (r =0.73) and mathematics (r =0.83) 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)[1]. Consequently, projection was used in this study. 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-ACT distributions[2].

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[3]

[1] Note that if the two assessments were administered closer to each other, the correlation might have been somewhat higher.
[2] For reading, as part of the loglinear smoothing procedure we preserved the first 3 moments for the NAEP distribution, 4 moments for the ACT distribution, and 4 cross-moments. For math, we preserved the first 2 moments for the NAEP distribution, 5 moments for the ACT 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.
[3] 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.

in order to produce reliable comparisons. For the Tennessee linking samples, both gender groups met that criterion. For the race/ethnicity groups, only White student subgroup met the criterion. Separate linking functions were established for these subgroups. 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, the linking functions for Male and White student subgroups were slightly higher than the overall linking function, and the linking function was slightly lower for Female student subgroup. For NAEP math, no substantial deviation from the overall linking function was detected for White student subgroup. The linking function for Female student subgroup was slightly higher than the overall linking function, and it was slightly lower for Male student subgroup. 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 projection tables containing conditional cumulative distributions of NAEP proficiencies for ACT scores. The range of possible NAEP scores below, at, and at or above the ACT benchmark (22 on the ACT reading scale and 22 on the ACT 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 ACT benchmarks. In addition, the direction of the linking relationship was reversed and the point on the ACT measure that corresponds most closely to the NAEP *Proficient* cut score was identified using the conditional cumulative distributions of the ACT scores for the NAEP proficiencies. We will discuss the results of the linking study in the following section.

Results

ACT benchmarks projected on the NAEP scale

The second and third analysis questions ask what scores on the NAEP reading and mathematics scales correspond to the ACT benchmarks. In other words, what would be the scale score on NAEP that corresponds most reasonably to an established benchmark of academic preparedness for college (i.e., the ACT).

Table 2 provides descriptive statistics to get an initial sense of where the benchmark most likely will be located on the NAEP scales as well as some distributional properties as context to these results. The average scores and percentile estimates for students below, at, and at or above the ACT 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, and At or Above the ACT Benchmarks

Subject	ACT Benchmark	Mean	Percentage	SD	Percentile		IQR[1]
					25th	*75th*	
Reading	*Below*	269	63%	29	250	289	39
	At	295	5%	23	280	310	30
	At or Above	311	37%	25	294	328	34
Mathematics	*Below*	135	73%	23	120	151	31
	At	164	4%	14	154	173	19
	At or Above	181	27%	18	167	192	25

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

To determine the NAEP scale score point that most reasonably corresponds to the ACT 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 ACT benchmark for pertinent score levels on NAEP. Colored vertical lines indicate where the NAEP achievement levels are located. Finally, and as mentioned previously, a proportion level has to be chosen commensurate with the confidence required to indicate whether students have passed the benchmark or not. A red dotted line shows above which point students are more likely to have reached the benchmark than not (i.e., the conditional proportion is set at 0.50). Given the moderate relationships between the two scales, this seems a reasonable location for indicating sufficient chance to be academically prepared for college. For context, a secondary, light orange line indicates when the conditional proportion *p* is set at 0.80, indicating a relatively high level of confidence that students have attained the ACT benchmark.

From the graphs it can be deduced that the location on the NAEP reading scale students have a reasonable probability to be academically prepared for college could be at a NAEP scale score of 301, slightly lower than the *Proficient* achievement level. The corresponding location on the NAEP math scale could be at 168, about 8 points below the *Proficient* achievement level.

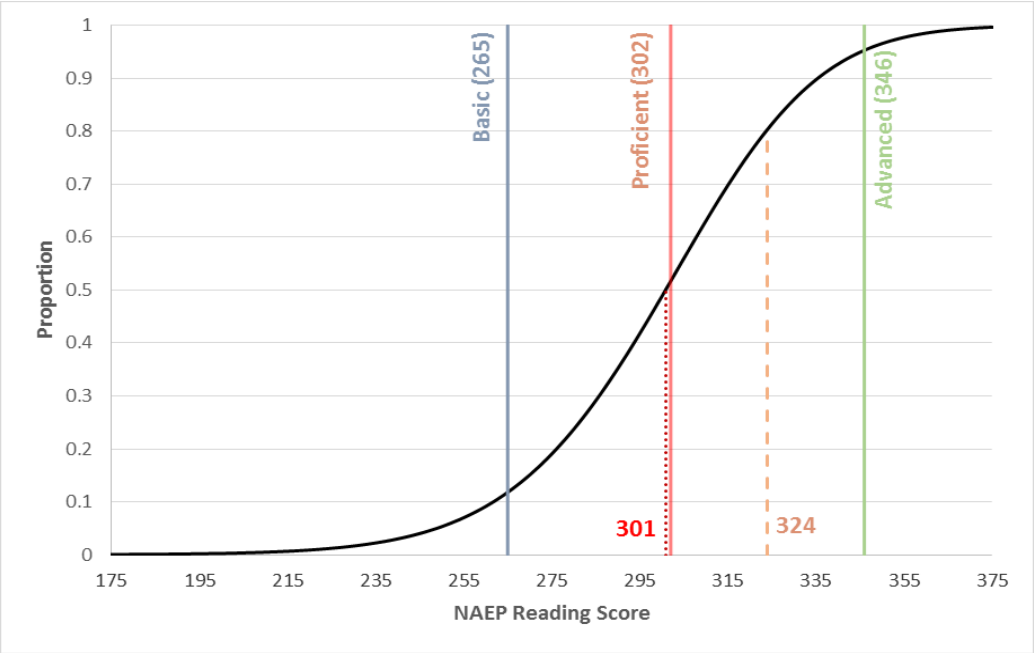

Figure 1: Proportion of students meeting the ACT reading benchmark of 22 in Tennessee for NAEP reading scores

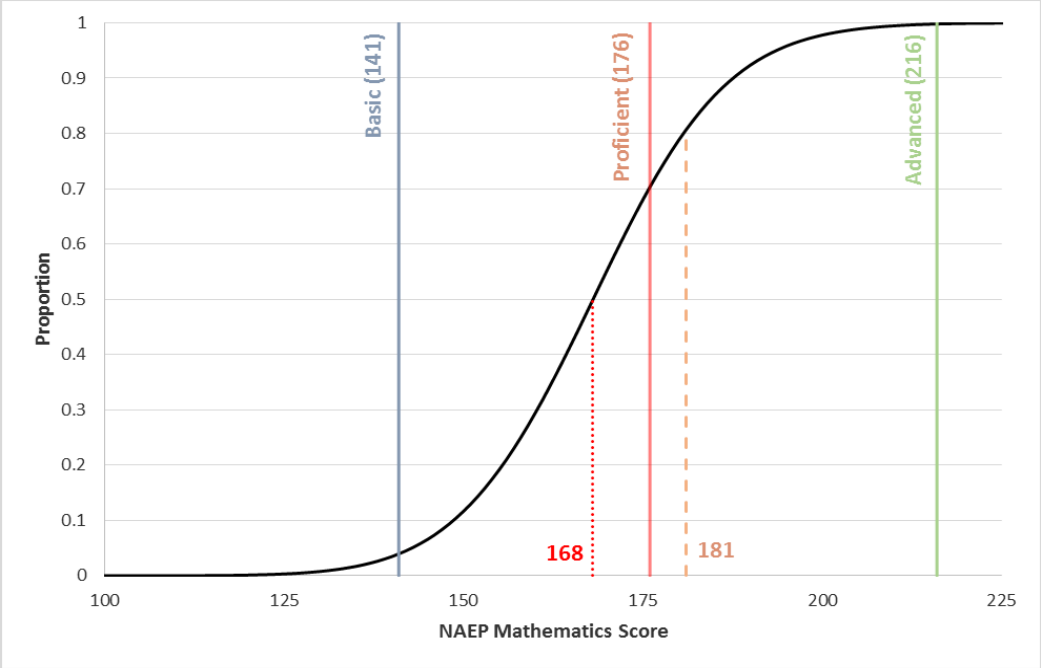

Figure 2: Proportion of students meeting the ACT mathematics benchmark of 22 in Tennessee for NAEP mathematics scores

NAEP *Proficient* cut scores projected on the ACT scale

To conduct the complementing analyses, we find the point on the ACT measure that corresponds most closely to the NAEP *Proficient* cut score, essentially reversing the direction of the linking relative to the previous analyses. Table 3 provides descriptive statistics of the ACT reading and mathematics scores for students below and at or above the grade 12 NAEP *Proficient* achievement level. The grade 12 NAEP *Proficient* level cut score was set at 302 for reading and 176 for mathematics.

Table 3: Descriptive ACT Statistics for Students Below, and At or Above the Grade 12 NAEP Proficient Level.

Subject	NAEP *Proficient*	Mean	Percentage	SD	Percentile		IQR[1]
					25th	*75th*	
Reading	*Below*	18	68%	5	14	20	6
	At or Above[2]	25	32%	5	21	28	7
Mathematics	*Below*	18	82%	3	15	19	4
	At or Above	26	18%	4	23	28	5

NOTES: [1]IQR is the Inter Quartile Range or the difference between the 75th and 25th percentiles.
[2]The "At" category has fewer than 1% students due to the non-continuous nature of the reporting ACT scale scores.

Following the same methodology of statistical projection (see Figures 3 and 4) we identified an ACT reading score of 23 and a mathematics score of 25 as cut points. The cut points are about 1 and 3 points higher than the ACT benchmarks for reading and mathematics tests, respectively, for grade 12 students.

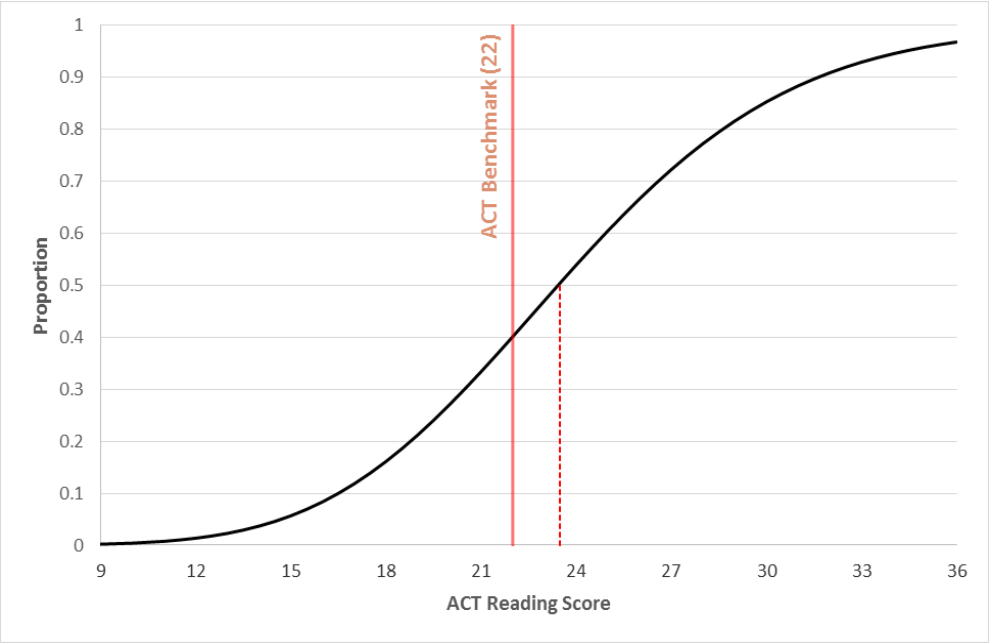

Figure 3: Proportion of students meeting the NAEP reading Proficient achievement level of 302 in Tennessee for ACT reading scores

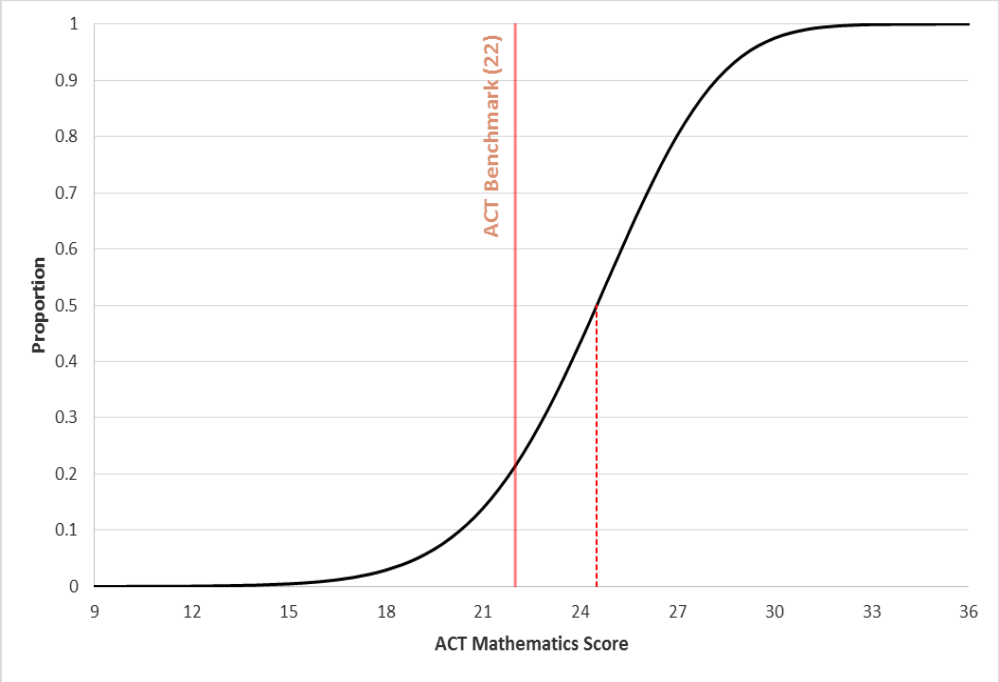

Figure 4: Proportion of students meeting the NAEP mathematics Proficient achievement level of 176 in Tennessee for ACT mathematics scores

Summary

The goal of this study was to statistically relate NAEP and ACT and use that relationship to identify a reference point or range on the NAEP 12th grade reading and mathematics scales reasonably associated with ACT benchmarks for reading and mathematics measures. Identifying such points would potentially allow NAEP to report on the percentage of students at 12th grade who are academically prepared for college for the nation and for states. The state of Tennessee participated in this study and graciously provided the critical ACT data necessary to conduct the linking study 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 academic preparedness reporting (see Figures 1 and 2 for examples of how the markers were determined).

In addition, we identified the point on the ACT measure that corresponds most closely to the NAEP *Proficient* achievement level cut score, for grade 12 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. In addition, in Tennessee the grade 12 NAEP assessment was administered almost a year later than the state-wide ACT administration, making interpretation somewhat more challenging. The results showed that, in the state of Tennessee, the ACT College Readiness Benchmarks and the NAEP *Proficient* achievement level cut scores correspond well to each other for reading in both linking directions (i.e., the projection results are 1 scale score point different from the ACT benchmark/NAEP *Proficient* level), but differ more for mathematics. In particular, the NAEP reading scale score of 301 could form a reasonable basis for reporting on academia preparedness for college, while the mathematics counterpart is 168 on the NAEP scale. On the other hand, the projection result of the NAEP *Proficient* reading cut score on the ACT scale is close to the existing ACT College Readiness Benchmark for reading, and about 3 points higher for mathematics. To what extent these results generalize to other states or the nation is an empirical question.

References

ACT (2013). *What are the ACT College Readiness Benchmarks?* (http://www.act.org/content/dam/act/unsecured/documents/benchmarks.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*.