2017 Nebraska Standard Setting for the ACT
July 10–11, 2017
Technical Report

Prepared September, 2017

Joann Moore, Jeffrey Steedle, Wayne Camara, & Teri Fisher

Table of Contents

2017 Nebraska Standard Setting for the ACT
July 10–11, 2017
Technical Report

Prepared September, 2017
Joann Moore, Jeffrey Steedle, Wayne Camara, & Teri Fisher

Introduction

ACT staff conducted an empirical standard setting at the request of the Nebraska Department of Education (NDE) on July 10–11, 2017. The process resulted in three recommended cut scores (Developing/Approaching, Approaching/On Track, and On Track/Exceeding) on the ACT defining four performance levels (Developing, Approaching, On Track, and Exceeding) for three subject areas: Mathematics, ELA[1], and Science. Performance Level Descriptors (PLDs) are planned to be developed in the fall of 2017. Following the standard setting meeting, the cut score recommendations were provided to the NDE Commissioner for review and approval. The final cut scores have not been approved at the time this report was prepared. The cut scores are intended to be applied to results from the spring 2017 and subsequent state administrations of the ACT to 11th grade students in Nebraska.

Context for New Cut Scores

Nebraska adopted College and Career Ready Standards for English Language Arts (ELA)[2] in 2014 and College and Career Ready Standards for Mathematics in 2015. In April 2016, the Nebraska State Legislature passed legislation requiring the use of a college admissions test as a replacement for the Nebraska State Assessments (NeSA) at the 11th grade. The ACT was selected based on ACT, Inc.'s response to a state-issued RFP. The ACT was administered to Nebraska 11th grade students in spring 2017.

Prior to adopting the new standards, NeSA had two cut scores that defined three performance levels (Below the Standards, Meets the Standards, and Exceeds the Standards). After Nebraska adopted new standards, the Nebraska ACT Standard Setting Advisory Committee recommended setting new performance levels for the NeSA grades 3-8 ELA assessments to define three performance levels (Developing, Established, and Distinguished). Math and Science standards have yet to be determined at the time this report was prepared.

Because the ACT and NeSA are different assessments, different approaches were considered for setting cut scores on the ACT. The Nebraska ACT Standard Setting Advisory Committee made recommendations related to issues such as number of performance levels and standard setting methodology. These recommendations, summarized in the following section, were reviewed by NDE and ACT staff, the Nebraska State Board of Education, Nebraska's Technical Advisory Committee, and the Nebraska Assessment and Accountability Advisory Committee.

[1] The ELA score is a combination of the ACT English, reading, and writing scores.
[2] https://www.education.ne.gov/AcademicStandards/index.html

Nebraska ACT Standard Setting Advisory Committee

The Nebraska ACT Standard Setting Advisory Committee made recommendations about the intended interpretations of the performance standards. Specifically, when NDE makes claims based on ACT results, those claims should focus on "college readiness" because that is what the ACT was designed to measure. The committee recommended that performance standards for the ACT be set using a process grounded in empirical evidence of the relationship between ACT scores and college readiness, where college readiness is manifested by grades in the first-year, credit-bearing college courses. Given the empirical evidence supporting the ACT College Readiness Benchmarks (Allen, 2013), the committee recommended that NDE use ACT's benchmarks in the reporting scheme rather than setting their own cut scores indicative of college readiness. Empirical evidence has been used by other states employing national college admissions tests for accountability, as well as states setting cut scores on their own high school assessments.

Regarding opportunity to learn, the committee recommended that, prior to standard setting, NDE should request research from ACT describing when students take the high school courses that teach content covered by the ACT. The committee suggested that this research reflect the ACT core curriculum (4 years of English, 3 year each of math, science, and social studies) and non-core reporting, and it would ideally be based on Nebraska schools. Additional research should be conducted to examine the alignment of Nebraska state standards to the content measured by the ACT.

The committee also made recommendations concerning alignment between ACT, NeSA, and other assessments. That is, NDE should not link ACT reporting to NeSA reporting until alignment evidence is available to support a connection between them. Related to this point, the Nebraska Technical Advisory Committee indicated that it is the responsibility of the state to create coherence. The committee also asserted that performance goals for the ACT cannot be set in isolation of the goals for grades 3–8; there should be a progression from grades 3–8 as students move toward the high school goals. To achieve this, standards should be set on the ACT first, then on Grade 8 NeSA, and finally on Grades 3–7 NeSA.

Whereas the previous standards aligned with proficiency, the new standards would align to college readiness. Given this, the committee anticipated possible decreases in pass rates. Consequently, the committee recommended that NDE carefully communicate the meaning of the new standards and the importance of assessment participation to schools. Again, the committee affirmed that the standard setting should focus on ACT's College Readiness Benchmarks.

Finally, the committee made recommendations regarding testing accommodations. That is, NDE should develop a communication plan to help schools understand the process for requesting accommodations and help students and parents/guardians make informed decisions about using accommodations not permitted for college-reportable scores. The committee also recommended a release of ACT accommodations, processes, videos, and deadlines for special education students and English language learners. These materials would include a crosswalk between ACT and former NeSA accommodations, a clarification of processes (e.g., evidence that issues with accommodations are infrequent), and a list of ELL accommodations (or "supports") to be added in spring 2018.

Following the convening of the Nebraska ACT Standard Setting Advisory Committee, ACT staff presented the empirical standard setting methodology to the NDE and the Governor's Technical Advisory Committee on March 10, 2017. In accord with the recommendations of the committee, a study was conducted in May 2017 to evaluate alignment between the Nebraska College and Career Readiness standards and the content measured on the ACT. That study found strong alignment in ELA and math, but weaker alignment in science. On June 15, 2017, NDE and ACT staff presented the empirical standard setting methodology to the Nebraska ACT Pre-Standard Setting Committee, and the committee was asked to make recommendations related to the number of performance levels and the names of the performance levels to be used when reporting ACT results for Nebraska's accountability system. The committee recommended that Nebraska adopt three cut scores reflecting four performance levels (Developing, Approaching, On Track, and Exceeding).

Methodology

Thee recommended cut scores for the ACT were determined using an empirical standard setting process, rather than a traditional content-based standard setting. In a content-based standard setting, like that used to establish NeSA cut scores and performance level descriptors for grades 3–8, cut scores and performance levels are established based on test content and content standards, and they are described in that context. In an empirical standard setting, panelists make judgments based on established relationships between test scores and a variety of educational outcomes. To inform the selection of ACT cut scores for Nebraska 11th graders, panelists viewed comparative evidence from the ACT, NeSA, and NAEP, other measures of high school achievement, impact on the percentages of students classified in different performance levels for several ACT-tested student populations, and impact on college success outcomes. This evidence is described in detail in the Orientation, Training, and Discussion section below.

Empirical standard setting approaches have been the primary method used to establish cut scores and benchmarks on admissions test for several reasons. The primary purpose of college admissions tests is to identify students who are likely to succeed in postsecondary academic environments. Such assessments are used for admissions, placement, recruitment, and talent identification because they predict GPA and grades in specific courses. Educators in secondary schools use the tests to determine if students are on track to being college ready at the end of high school, to identify strengths and weaknesses to address, and to aid in postsecondary planning for students. All such information prioritizes the empirical relationship between test scores and outcomes such as enrollment, course grades, GPA, and retention. College readiness benchmarks for the ACT and SAT have been established exclusively based on such empirical relationships (e.g., 50% chance of a B or higher in college algebra), and ACT has employed empirical standard setting methods when assisting states to set upper and lower cut scores, as well as validating or establishing their own proficient (or 'meets') level.

ACT College Readiness Benchmarks

The primary evidentiary sources for the Nebraska standard setting were the ACT College Readiness Benchmarks and the probabilities of C or higher, B or higher, or A grades in first-year college courses derived in the research undergirding the development of the Benchmarks.

In 2005, ACT established College Readiness Benchmarks representing the ACT assessment scores of students in 11th and 12th grade associated with a 50% chance of earning a B or higher grade in common first-year credit-bearing courses at a typical postsecondary institution (Allen & Sconing, 2005). The Benchmarks also correspond to an approximate 75% chance of earning a C or higher grade in these courses. The first-year credit-bearing courses studied were English Composition, College Algebra, Social Science courses (including American History, Other History, Psychology, Sociology, Political Science, and Economics), and Biology. The original Benchmarks corresponded to scores of 18 on the English test, 22 in Mathematics, 21 in Reading, and 24 in Science.

The Benchmarks were updated in fall 2013 (Allen, 2013) to address possible changes in college grading standards, aggregate college student performance, course taking patterns of first-year college students, and alignment between secondary and postsecondary course content that transpired since the original benchmarks were established. Using a large sample of first-year students attending two- and four-year institutions, the study detected no changes in the English and Mathematics Benchmarks (18 and 22, respectively), the Reading Benchmark increased from 21 to 22, and the Science Benchmark decreased from 24 to 23. The distribution of institutions by state is shown in Figure 1. The sample included more institutions in states that typically enroll higher proportions of ACT-tested students (i.e., states in the South and Midwest) and fewer institutions from states that typically enroll fewer ACT-tested students (i.e., states on the East and West coasts).

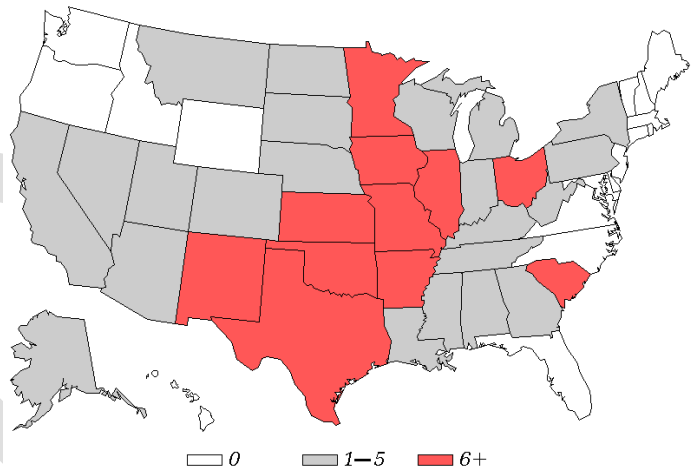

Figure 1. Number of benchmark institutions by state.

Compared to ACT-tested students nationally who enroll in college, students in the course samples were more likely to be female, less likely to be Hispanic or African American and more likely to be White, less likely to have extreme ACT Composite scores, and less likely to have lower high school GPAs. Moreover, fewer students in the samples enrolled at selective and highly selective institutions. To address this issue, the samples were adjusted statistically to make results approximate what would be observed with a nationally representative sample of ACT-tested college-going students. **Error! Reference source not found.** summarizes the characteristics of the institutions used in the study.

Table 1
Institutional Samples Used in Benchmark Development

Characteristic	College course			
	English Composition I	**College Algebra**	**Social Science**	**Biology**
N (Institutions)	136	125	129	90
N (Students)	96,583	70,461	130,954	41,651
Type:				
2-year	50%	42%	42%	44%
Less selective 4-year	43%	48%	49%	46%
More selective 4-year	7%	10%	9%	10%
Control:				
Public	88%	92%	92%	87%
Private	13%	8%	8%	13%

Error! Reference source not found. illustrates the overall success rate by course, which ranged from 47% in Biology (ACT Science benchmark) to 59% in English Composition I (ACT English benchmark) for the B or higher criterion and from 72% in College Algebra (ACT Math benchmark) to 81% in English Composition I for the C or higher criterion. Across all courses, B was the modal course grade. For additional details about the regression models, see Allen (2013).

Table 2
Success Rates by Course

College course type	Percentage of course grades					Success criteria	
	A	**B**	**C**	**D**	**F**	**≥B**	**≥C**
English Composition I	27%	32%	22%	7%	13%	59%	81%
College Algebra	24%	25%	23%	11%	18%	49%	72%
Social Science	25%	27%	23%	10%	14%	53%	76%
Biology	20%	27%	26%	12%	16%	47%	73%

The B or higher grade is used by ACT as the criterion for the benchmarks for several reasons. First, the statistical models used to develop the benchmarks are affected by courses and institutions where grades below a C are uncommon. Courses in English and the social sciences, in particular, frequently have 80% to 90% of students earning grades of C or higher. In addition, establishing a policy wherein students with only a 50% chance of earning a C or higher are placed into a class could be problematic because students would also have a 50% chance of earning a D or F. Moreover, the B or higher criterion best reproduced the original grade distribution.

In 2015, ACT began reporting an English Language Arts (ELA) score, which is an average of the ACT English, Reading, and Writing scores (after the Writing score is transformed from a 2–12 scale to a 1–36 scale). Students must take the ACT Writing test to obtain an ELA score. In 2017, ACT developed an ELA Benchmark (Radunzel, Westrick, Bassiri, & Li, 2017). The methodology used to develop the ELA Benchmark was similar to that used to develop the benchmarks for the four ACT subject tests. The ELA Benchmark is the score associated with a 50% chance of earning a B or higher grade in English Composition I, American History, Other History, Psychology, Sociology, Political Science, and Economics courses (the same courses used to develop the English and Reading Benchmarks, respectively). The ELA Benchmark also corresponds to an approximate 75% chance of earning a C or higher grade in these courses. **Error! Reference source not found.** is a summary of the institutional samples used to develop the ELA Benchmark. The institutions represented in ACT research used to set the benchmarks approximately reflected the composition of colleges and universities in the U.S. in terms of selectivity and 2-year vs. 4-year institutions. It should be noted that few differences have been detected between 2-year and 4-year institutions in setting college readiness benchmarks.

Table 3
Institutional Samples Used in ELA Benchmark Development

	College course		
Characteristic	**English Composition I**	**Combined Social Science**	**Total Sample**
N (Institutions)	200	154	233
N (Students)	107,142	91,133	198,275
Type			
2-year	42%	43%	40%
Less selective 4-year	51%	50%	53%
More selective 4-year	7%	7%	7%
Control			
Public	89%	94%	88%
Private	11%	6%	12%

Error! Reference source not found. show the overall success rates by course in the ELA benchmarking study. Across all courses, B was the modal grade. The overall percentage of students earning a B or higher was approximately 52%, and the overall percentage of students earning a C or higher was approximately 77%.

Table 4
Success Rates by Course

College course type	Percentage of course grades					Success criteria	
	A	B	C	D	F	≥B	≥C
English Composition I	20%	35%	26%	7%	12%	55%	81%
Social Science	21%	28%	25%	11%	16%	49%	73%
Total Sample	20%	32%	25%	9%	14%	52%	77%

ACT Performance in Nebraska

Figure 2 provides a comparison of the percentage of students in the 2016 ACT-tested high school graduating cohort[3] who met or exceeded the Benchmarks in Nebraska and the nation. Eighty-eight percent of Nebraska 2016 graduates took the ACT. Nebraska's performance is higher than the national average. Note that the Nebraska-specific results include the most recent test scores of all students completing the ACT (n=18,598), not just students testing in 11th grade. This distinction is important to consider when comparing the 2016 Nebraska high school graduates to students testing in 11th grade in 2017. *The Condition of College & Career Readiness 2016* state report contains additional information about how Nebraska graduates performed on the ACT (ACT, 2016).

[3] The graduation cohort represents all students in a state or the nation who completed the ACT at any point during high school. This includes students attending private or public high schools. The last ACT test score is used in reporting results for cohorts. Therefore, results from the NE graduating cohort will differ from results reported for all juniors who tested in public schools as part of the spring census testing event.

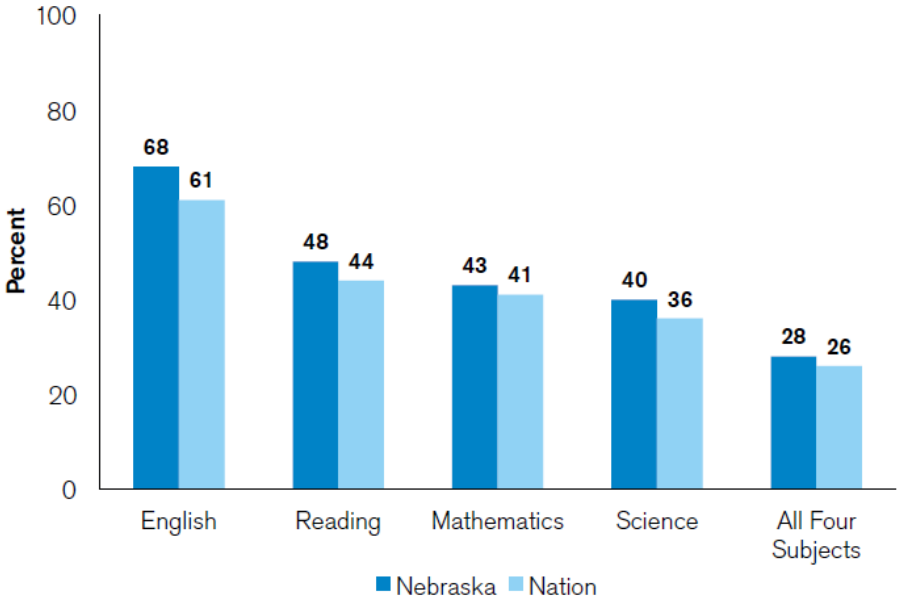

Figure 2. Percentage of 2016 Nebraska and national ACT-tested high school graduates meeting ACT's College Readiness Benchmarks, by subject.

Panelists

The NDE was responsible for all logistical details such as identifying panelists, communicating with panelists about the event, reimbursement for travel and stipends, and meeting coordination. Panelists were invited via email to participate during the spring of 2017. This process produced 11 panelists each for ELA and mathematics, and 10 panelists for science. The panel for each subject area included a mix of teachers, administrators, district curriculum or assessment coordinators, counselors, higher education faculty, and NDE staff.

In terms of demographics, the panel was 63% female, 100% non-Hispanic, and 97% White. The panel included 7 classroom teachers (3 in math, 2 in ELA, and 2 in science), with 4 (57%) teaching grade 9, 3 (43%) teaching grade 10, 7 (100%) teaching grade 11, and 6 (86%) teaching grade 12. Three of the teachers taught ELL students (43%), 2 (29%) taught special education students, and 3 (43%) taught gifted and talented students. Five panelists (16%) were principals or administrators, 6 (19%) were curriculum/assessment coordinators, 2 (6%) were counselors, 3 (9%) were higher education faculty (representing both two- and four-year colleges), and 9 (28%) were NDE staff or consultants.

Twenty seven panelists (84%) provided responses to questions pertaining to position and 26 (81%) reported their highest degree earned. All respondents indicated working at least 6 or more years in education, with more than half (17 panelists; 63%) indicating that they have worked in education for 20 years or more. Most respondents indicated that they had been in their current positions for 1–5 years (9 panelists; 33%) or 6–10 years (10 panelists; 37%). Three respondents (11%) were in their current positions for less than a year, and the rest had been in their current positions for more than 10 years (5 panelists; 19%). As for highest educational degree attained, 58% of respondents (15 panelists) completed a master's degree and 42% completed a doctoral degree (11 panelists). The background questionnaire and summary of results can be found in Appendix C.

Process

The two-day standard setting process was structured as follows:

- Overview of Nebraska Standards and context for standard setting
- ACT College Readiness Benchmarks and probabilities of success in first-year college courses
- Impact and comparative impact data (percent at/above each score point on the ACT, NeSA, and NAEP assessments)
- Judgment Round 1: On Track Level
  - Overview, discussion, and On Track rating to identify borderline achievement in terms of probability of success
- Judgment Round 2: On Track Level
  - Review Round 1 results
  - Review content-specific comparative and impact evidence
  - Discussion and Round 2 On Track rating to identify borderline achievement in terms of probability of success and ACT score
- Judgment Round 3: Approaching and Exceeding Levels
  - Review content-specific comparative and impact evidence
  - Review Round 2 results
  - Overview, discussion, and Approaching and Exceeding ratings to identify borderline achievement in terms of probabilities of success and ACT scores
- Judgment Round 4: On Track, Approach, and Exceeding Levels
  - Review Approaching, On Track, and Exceeding results from Round 3
  - Review content-specific comparative and impact evidence
  - Discussion and final ratings for Approaching, On Track, and Exceeding

Slides and other reference materials presented during the standards setting can be found in the Appendix.

Day 1

Orientation, Training, and Discussion

Day one of the standard setting meeting began with introductions of the key organizations and staff, followed by introductions of the panelists. The purpose and context of the meeting was summarized by NDE staff. ACT provided examples of performance level descriptors from Smarter Balanced, PARCC, NAEP, and NeSA to illustrate the differences between content and performance standards and differences between content and empirical standard settings. Information about the ACT test and the standard setting task was also provided.

The percentages of students at or above each performance level on the NeSA Math, Reading, Writing, and Science assessments in grades 8 and 11 in 2015 and 2016 were presented to show historical performance before the introduction of the new Nebraska College and Career Readiness Standards (see Tables M1, ELA1, and S1 in Appendices E, F, and G, respectively). The percentages of students at or above each performance level on grade 8 NAEP in Nebraska and nationally, and on grade 12 NAEP nationally (there were no Nebraska state NAEP results for grade 12) in math, reading, and science were also presented as points of reference (see Tables M2, ELA2, and S2 in Appendices E, F, and G, respectively). The comparison to NAEP results

revealed very different percentages of students considered Proficient depending on the assessment and the definition of Proficient. Specifically, a much higher percentage of students met the NeSA standard than met the NAEP definition of Proficient. For example, in 2015, 68% of Nebraska 8th grade students met or exceeded the NeSA standard in math, whereas only 38% of Nebraska 8th grade students met or exceeded the NAEP Proficient standard in math.

ACT staff presented several sources of information to provide context about why college readiness is important, including the percentages of ACT-tested 2016 high school graduates enrolled in college in Nebraska (73%) and nationally (64%). Additional evidence included median earnings of students who completed high school ($30,500) or attained associate ($36,900), bachelor's ($50,000), or a master's degree or higher ($60,000), obtained from *The Condition of Education 2017* report from the National Center for Education Statistics (McFarland, Hussar, de Brey, Snyder, Wang, Wilkinson-Flicker, Gebrekristos, Zhang, Rathbun, Barmer, Bullock Mann, and Hinz, 2017), and remedial coursework rates in college (25% of students at 4-year colleges and 61% of students at 2-year colleges), obtained from *The Condition of Education 2004* report (U.S. Department of Education, National Center for Education Statistics, 2004) and a research report from ACT (Noble and Sawyer, 2013).

ACT also shared the results of a study from the National Center for Education Statistics mapping state grade 8 reading and mathematics standards to the NAEP scale (Bandeira de Mello, Bohrnstedt, Blankenship, and Sherman, 2015). While different states use different assessments for accountability, NAEP is administered nationally, so the percentage of students meeting the NAEP standard is a useful point of reference for interpreting the percentage of students meeting a given state's standard. This study revealed that some states set more challenging standards than others, with many states setting their Proficient standard at a level that is comparable to the NAEP definition of Basic, and very few states setting their Proficient standards at a level that is comparable to the NAEP definition of Proficient. Nebraska's standards in 2013 were near the middle of the distribution across states, with their Meets the Standard performance level nearly equivalent to Basic on NAEP for reading and above Basic but below Proficient for mathematics.

ACT provided a summary of the ACT scores used for postsecondary course placement, both nationally and in a sample of Nebraska colleges. The national data were obtained from a published study (Fields & Parsad, 2012) in which 23% of institutions reported using ACT Math scores for placement, and 16% reported using ACT Reading scores for placement. Nebraska-specific data were gathered by placing phone calls to the testing centers or advising offices of postsecondary institutions in Nebraska. Of the 11 institutions contacted, 5 reported using ACT scores for first-year course placement. The sample included two community college systems (3 campuses each) and three baccalaureate institutions (one awarding only bachelor's degrees, one also awarding master's degrees, and one also awarding Ph.D. degrees). The results indicate that ACT Math scores are more commonly used for course placement, and College Algebra placement scores are typically close to the ACT College Readiness Benchmark of 22, while lower scores of 17–19 could place a student into lower level credit-bearing math courses. ACT English and Reading were used less frequently, with placement scores near the ACT College Readiness Benchmarks of 18 in English, and no institutions reported using ACT Science scores for course placement. It should be noted that the Nebraska-specific data were a small sample of all of Nebraska's 2-year and 4-year colleges, and were not representative of Nebraska postsecondary institutions.

ACT staff presented background information to the panelists about the ACT College Readiness Benchmarks, as described above. In addition, the panelists received data books containing comparative demographic information and test results for several ACT-tested student populations of interest:

- 2017 ACT-tested juniors in Nebraska (census tested)
- 2015 and 2016 ACT-tested juniors in Nebraska
- 2015 and 2016 ACT-tested juniors in census-tested states[4]
- 2015 and 2016 ACT-tested juniors nationally
- 2015 and 2016 Nebraska graduate cohorts
- 2015 and 2016 graduate cohorts in census-tested states[5]
- 2015 and 2016 graduate cohorts nationally

The 2017 ACT-tested juniors in Nebraska (census tested) composed the primary population of interest since their data will be reported for accountability purposes. The main comparison groups included Nebraska juniors in 2016 (prior to statewide adoption), census-tested juniors in other states in 2016, and the national population of ACT-tested juniors in 2016. Other populations of interest were summarized in less detail, and were provided for context and comparison with the primary populations.

ACT participation rates, percent tested in 11th grade, percent who took the ACT Writing test, percent female, percent white, average ACT scores, and percent meeting the ACT Benchmarks were provided for each group (see Table 5 and Tables M4.1, M4.2, ELA4.1, ELA4.2, S4.1, and S4.2 in Appendices E, F, and G, respectively). Differences between populations were emphasized because those differences are important to consider when comparing different populations. For example, when comparing ACT-tested juniors in Nebraska in 2016 (prior to statewide adoption) and 2017 (after statewide adoption), several differences are important to account for. In 2016, 71% of Nebraska juniors took the ACT, compared to 99% in 2017. Juniors who took the ACT in 2016 were more likely to be higher-achieving, college-bound students, whereas in 2017 nearly all juniors tested. This is one factor that can explain the decrease in scores from 2016 to 2017. Additionally, while 99% of 2017 juniors took the ACT Writing test, only 10% did so in 2016. The ACT Writing test is required for students to receive an ELA score; the large decrease in the percentages of students meeting the ELA Benchmark from 2016 to 2017 can be explained in part by the fact that the small population of Nebraska students who took the ACT Writing test in 2016 were higher-achieving students than those who did not take the Writing test.

[4] Census-tested states include only those that administered the ACT with Writing and therefore received ELA scores. This includes Alabama, Colorado, Hawaii, Illinois, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Montana, Nevada, North Carolina, North Dakota, South Carolina, Utah, Wisconsin, and Wyoming in 2015 and Alabama, Arkansas, Colorado, Hawaii, Kentucky, Louisiana, Mississippi, Missouri, Montana, Nevada, North Carolina, North Dakota, South Carolina, Utah, Wisconsin, and Wyoming in 2016.

[5] Graduate cohorts in census-tested states in 2015 and 2016 include only states that administered the ACT with Writing in 2014 and 2015. This includes Alabama, Colorado, Hawaii, Illinois, Kentucky, Louisiana, Michigan, Montana, North Carolina, North Dakota, Utah, and Wyoming in 2014 and Alabama, Colorado, Hawaii, Illinois, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Montana, Nevada, North Carolina, North Dakota, South Carolina, Utah, Wisconsin, and Wyoming in 2015

When comparing ACT-tested juniors to ACT-tested graduating cohorts, it is important to note that the graduating cohorts are based on students' most recent ACT scores, for students who tested more than once. Because some students may have retested between their junior year and the end of their 12th grade year, scores are slightly higher for the graduating cohorts than for the junior-tested populations. For example, 34% of the Nebraska 2016 graduating cohort tested in 11th grade, compared to 100% of the Nebraska 2016 junior-tested cohort. Results from all states using the ACT (with Writing) for census testing of juniors was provided to panelists for comparison purposes because it eliminates the self-selection bias that exists with national data.

ACT staff presented the results of a recent study investigating score gains of students by the timing of when they take ACT test, as well as expected gains by months of instruction between tests (Camara & Allen, 2017). They found that approximately 57% of students increase their scores upon retesting, with score gains of about one point on average. The lower the initial score, the more likely there will be an increase upon retest. Score gains increase on average approximately 0.20 points per additional month of instruction, yielding gains of approximately one point between students' junior and senior year of high school.

Table 5
Summary of Demographic and Test Results by Student Population

	Nebraska Graduates		Nebraska State Juniors		Census State Juniors	
	2016	**2015**	**2017**	**2016**	**2016**	**2015**
N	18,598	18,347	21,711	16,419	384,406	686,137
Juniors	88%	88%	100%	100%	100%	100%
ACT Participation	34%	32%	99%	71%	99%	99%
Took Writing	10%	10%	99%	10%	99%	99%
Male	47%	48%	51%	47%	50%	50%
Female	53%	52%	49%	53%	50%	50%
African American	4%	4%	6%	3%	17%	15%
American Indian	1%	1%	1%	1%	1%	1%
White	73%	74%	68%	75%	57%	59%
Hispanic	13%	12%	17%	12%	12%	12%
Asian American	3%	3%	2%	3%	3%	4%
Pacific Islander	0%	0%	0%	0%	1%	0%
Two or more races	4%	4%	4%	3%	5%	5%
Mean ACT English (% met ACT Benchmark)	20.9 (68%)	21.1 (69%)	18.4 (50%)	20.7 (67%)	17.8 (47%)	18.5 (52%)
Mean ACT Reading (% met ACT Benchmark)	21.8 (48%)	21.9 (49%)	19.5 (34%)	21.8 (50%)	19.2 (33%)	19.6 (33%)
Mean ACT Math (% met ACT Benchmark)	20.8 (43%)	21.0 (44%)	19.4 (31%)	21.0 (44%)	18.9 (28%)	19.4 (32%)
Mean ACT Science (% met ACT Benchmark)	21.5 (40%)	21.6 (42%)	19.5 (30%)	21.5 (41%)	19.3 (26%)	19.5 (28%)
Mean ACT ELA (% met ACT Benchmark)	23.8 (74%)	24.2 (78%)	18.4 (41%)	22.2 (67%)	17.6 (35%)	18.2 (40%)
Mean ACT Writing	7.5	7.6	6.3	7.1	5.8	6.0

ACT staff then presented the probabilities of achieving an A, B or higher, or C or higher grade in first-year credit-bearing English, mathematics, social science, and science courses by ACT score. These probabilities were developed as part of the Benchmark update study (Allen, 2013). The impact of selecting a particular probability of success was illustrated using test score

distributions for Nebraska 2017 ACT-tested juniors, Nebraska 2016 ACT-tested juniors, juniors from other states who administered the ACT with writing statewide, and juniors nationally who took the ACT in 2016, as shown in Table 6 in an abbreviated form. Several slides were presented focusing on the distinction between the probabilities of success and the impact data because an understanding of these elements is vital to the standard setting task.

Table 6

Success Probabilities and Percentage of Students Scoring At or Above

Probability of Success			**Percentage At/Above**			
A	**B or higher**	C or higher	**Juniors Census 2017**	**NE Juniors 2016**	**State Juniors 2016**	**National Juniors 2016**
	Mathematics					
0.19	**0.48**	0.72	32	46	29	43
0.20	**0.50**	0.73	31	44	27	41
0.22	**0.52**	0.74	30	42	26	40
	ELA					
0.13	**0.48**	0.75	45	70	39	57
0.14	**0.50**	0.77	41	67	36	54
0.15	**0.52**	0.78	38	65	33	51
	Science					
0.17	**0.48**	0.76	30	34	38	40
0.19	**0.50**	0.78	27	31	35	37
0.20	**0.52**	0.79	25	28	33	34

Data books were provided to panelists containing impact evidence by subject area and ACT test score; each panelist received a data book specific to his or her subject area. The data books contained secure information and panelists were not allowed to remove them from the meeting rooms, but they could reference them throughout the standard setting process. A sample page from the mathematics data book is provided in Table 7; complete data books for mathematics, ELA, and science are included in Appendices E, F, and G, respectively.

Table 7
Sample Page from Mathematics Data Book

Percentage of Students At or Above Each Success Probability
ACT Subject: Mathematics
College Course: Algebra

Probability of Success			**Percentage At/Above**			
			NE Juniors Census	**NE Juniors**	**Census State Juniors**	**National Juniors**
A	**B or higher**	C or higher	**2017**	**2016**	**2016**	**2016**
0.04	0.20	0.48	84	93	83	89
0.05	0.22	0.51	78	89	75	84
0.06	0.24	0.53	70	83	67	77
0.07	0.26	0.55	62	78	58	71
0.07	0.28	0.57	56	73	54	67
0.08	0.30	0.59	51	68	49	62
0.09	0.32	0.60	48	64	45	59
0.10	0.34	0.62	46	60	41	55
0.11	0.36	0.64	43	57	38	53
0.12	0.38	0.65	40	55	36	50
0.13	0.40	0.67	38	53	34	49
0.15	0.42	0.68	37	51	33	47
0.16	0.44	0.69	36	50	32	46
0.17	0.46	0.71	34	48	30	44
0.19	0.48	0.72	32	46	29	43
0.20	0.50	0.73	31	44	27	41
0.22	0.52	0.74	30	42	26	40
0.24	0.54	0.76	28	40	25	38
0.25	0.56	0.77	27	38	24	36
0.27	0.58	0.78	25	35	22	34
0.29	0.60	0.79	23	33	20	32
0.32	0.62	0.80	21	30	18	30
0.34	0.64	0.81	18	27	16	27
0.36	0.66	0.82	17	24	14	25
0.39	0.68	0.83	15	22	12	23
0.42	0.70	0.84	13	19	11	21
0.45	0.72	0.85	11	16	9	18
0.48	0.74	0.86	9	14	7	16
0.51	0.76	0.87	8	11	6	14
0.54	0.78	0.88	7	9	5	12
0.58	0.80	0.89	5	7	4	10
0.62	0.82	0.90	4	6	3	8
0.66	0.84	0.91	3	4	2	7
0.70	0.86	0.92	2	3	2	5
0.74	0.88	0.93	2	3	1	4
0.79	0.90	0.94	1	2	1	3
N-count	70,461		21,711	16,419	384,406	1,739,885

Round 1 Rating

Instructions were given for the first round of making cut score judgments. Table 7 also illustrates the Round 1 Rating Form for math; there were corresponding forms for ELA and science (see forms ELA5 and S5, respectively in Appendices F and G). Panelists were instructed to highlight the row of probabilities that corresponded to their conception of a minimally On Track student in their respective subject areas. After the panelists made their judgements, the session ended and the panelists went to lunch. The first rating task was designed to be completed without access to how ACT scores correspond to impact data. ACT feels it is important to have initial ratings based on grades and probabilities of success and to ensure that initial ratings are not overly influenced by rater's perception about the meaning of specific ACT scores. However, during the discussion, panelists requested access to such data prior to the second rating and NDE determined it would be best to share this information with panelists prior to the first rating. In retrospect, this decision appeared to prematurely focus panelists on what ACT score was defensible as a benchmark rather than focus them on the impact and performance data as the empirical methodology proposes.

Round 1 Results and Discussion

The day one afternoon session began with a review of the Round 1 results, followed by a review of the subject area-specific comparative and impact evidence. Figure 3 and Table 8 show the results of the first round of ratings. Median probability ratings were calculated within subject areas. The resulting medians reflected notable variability across subject areas. The median probability of earning a B or higher grade was .26 for Mathematics, .38 for ELA, and .46 for Science. Probabilities of earning an A or C or higher grade were also provided in the data books (see Table 7), and panelists were instructed to use the probabilities that made the most sense to them when making judgements. Based on discussion during the standard setting, quite a few panelists focused on the C or higher probabilities. Because panelists were instructed to highlight the entire row on the rating form, it ultimately did not matter in terms of the medians which probability (A, B or higher, or C or higher) was their focus.

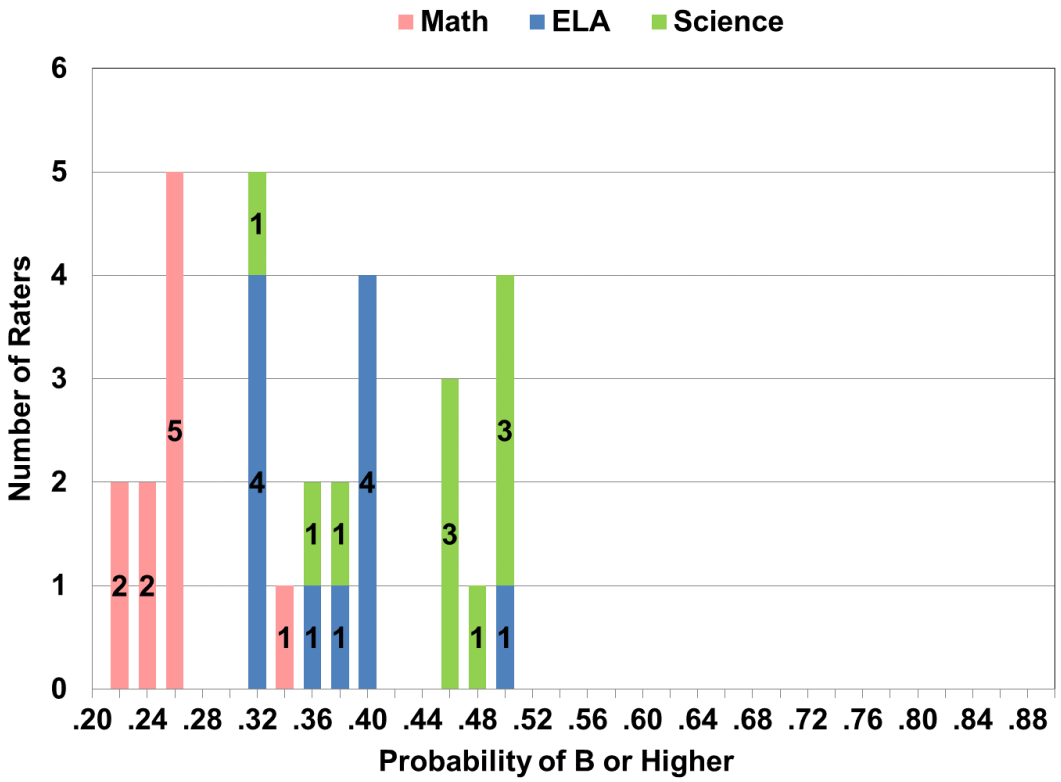

Figure 3. Round 1 On Track Ratings by B or higher grade probabilities.

Table 8
Descriptive Statistics for Round 1 On Track Ratings

On Track Round 1	Mathematics		ELA		Science	
	B or higher prob.	ACT score	B or higher prob.	ACT score	B or higher prob.	ACT score
Range	0.22–0.34	16–19	0.32–0.50	15–20	0.32–0.50	19–23
Median	0.26	17	0.38	17	0.46	22
Mode	0.26	17	--	17	--	23
SD	0.03	0.92	0.06	1.51	0.06	1.40

Table 9 shows the median Round 1 ratings and the ACT scores associated with those ratings. Table 9 also provides impact data for Nebraska census testing juniors in 2017, Nebraska juniors in 2016, all census-tested juniors in 2016, and the national sample of ACT-tested juniors in 2016. All evidence presented *after* the Round 1 ratings was anchored to ACT scores, rather than the success probabilities because the ultimate objective was to choose ACT cut scores for to the Approaching, On Track, and Exceeding achievement levels.

Table 9
Probabilities and Impact Data Associated with Round 1 On Track Cut Scores

	ACT score	Probability			Percentage at or above			
		A	**B or higher**	C or higher	**NE Juniors Census**	**NE Juniors**	**Census States Juniors**	**National Juniors**
					2017	**2016**	**2016**	**2016**
	Mathematics							
Round 1	17	0.07	**0.26**	0.55	62	78	58	71
	ELA							
Round 1	17	0.09	**0.38**	0.69	62	83	55	72
	Science							
Round 1	22	0.16	**0.46**	0.76	33	48	30	43

Multiple sources of comparative evidence were summarized on a number line representing percentage at or above for each subject area (Figure 4, Figure 5, and Figure 6). The figures allowed easy comparisons of various measures of achievement with color coding to differentiate evidence sources by type and population. Evidence shown in white, green, and light blue reflected the percentage meeting or exceeding certain ACT scores for the primary population of interest (Nebraska 11th graders in 2017). Evidence sources shown in red reflected performance on NeSA of Nebraska 11th graders in 2016, and percentages of students at or above Basic and Proficient on NAEP were shown in dark blue. This collateral evidence (e.g., NAEP, college placement cut scores) was meant to provide additional context for panelists in setting cut scores beyond ACT impact data alone.

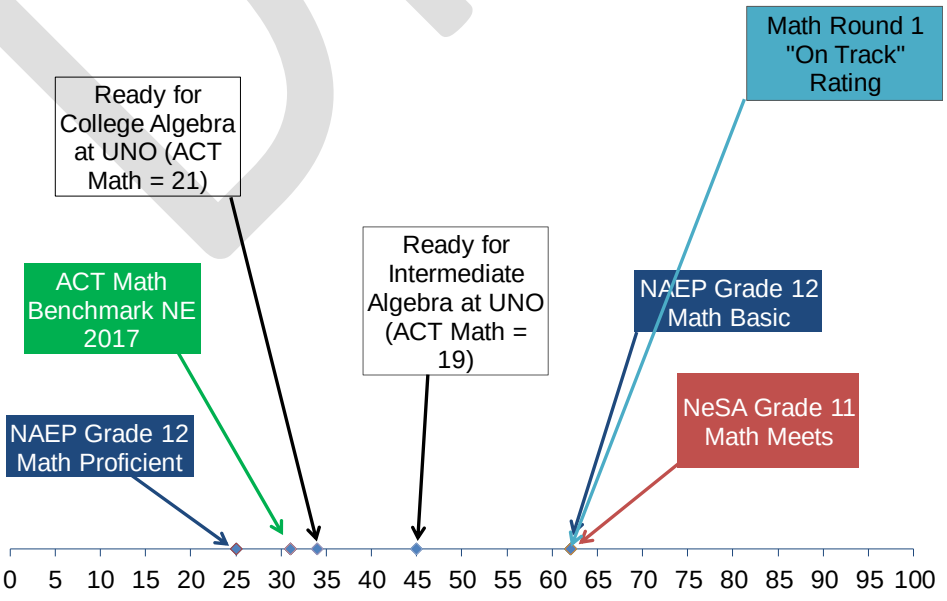

Figure 4. Math On Track level comparative information.

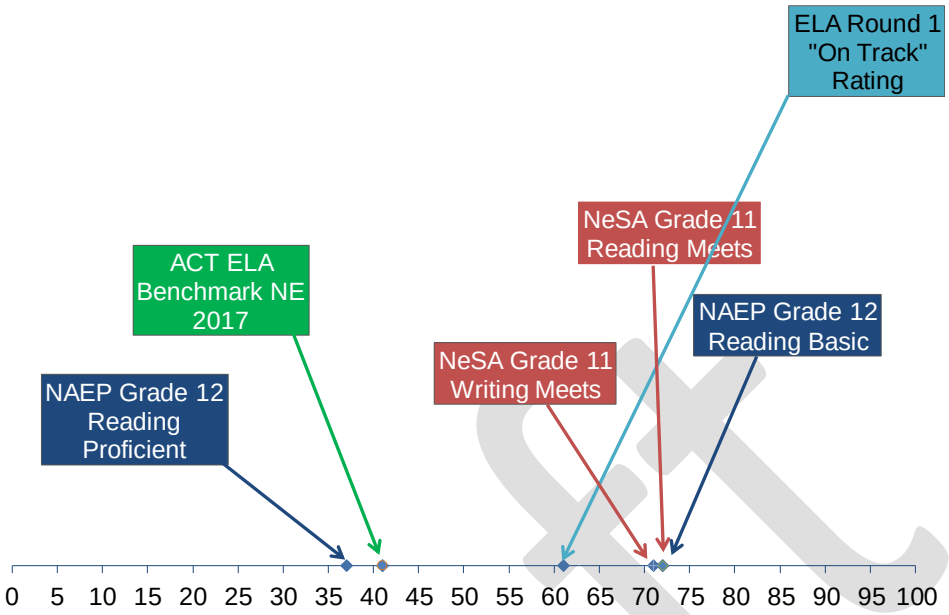

Figure 5. ELA On Track level comparative information.

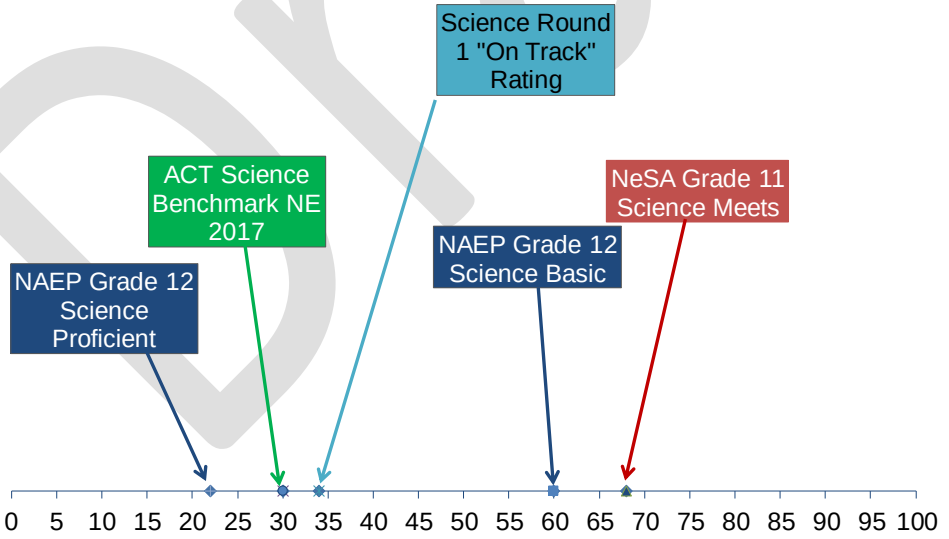

Figure 6. Science On Track level comparative information.

The percentages of students scoring at or above each ACT Math, ELA, and Science score were provided for Nebraska juniors in 2016 and 2017, census-tested juniors in 2016, and national juniors in 2016, and were disaggregated by numbers of years of coursework in math, social studies, or science (categories of 2+, 3+, and 4+ years; coursework in English was not presented because nearly all students take four years of English in high school). The distributions of

highest courses taken in math and science for the four comparison groups were also presented as context about the highest level of coursework taken at the time students take the ACT. The presentation also included college enrollment rates for students scoring at the Round 1 cut scores for 2015 and 2016 Nebraska graduate cohorts, 2016 national graduates, and 2016 graduates in census-tested states. In addition, panelists viewed the probability of long-term college success based on a study of 2003 ACT-tested high school graduates nationally (Noble & Radunzel, 2012). This included the probability of returning to any college in year 2 for students who first enrolled in a 2-year college, the probability of returning to any college in year 2 for students who first enrolled in a 4-year college, the probability of completing a bachelor's degree in 6 years for students who first enrolled in a 4-year college, and the probability of completing an associate or bachelor's degree in 6 years. All of the tables and slides containing comparative and impact evidence presented during the standard setting can be found in the Appendix.

After the comparative and impact evidence were presented, panelists were given the opportunity to discuss their ratings and the evidence. Panelists were instructed to focus their discussion on how their cut score judgements compared to others in their respective subject areas and in other subject areas, how the impact information may have influenced their On Track cut score judgements, and which information was the most important in deciding the On Track cut score.

Round 2: Proficient Level–Identifying Borderline Achievement by Probability of Success and ACT Score

A second round of On Track ratings followed the discussion. The Round 2 ratings process followed the same general procedures as those of the Round 1 ratings, with panelists highlighting a single row of scores and their associated probabilities on the rating sheet. The rating sheets for Round 2 differed from the rating sheets for Round 1 in that each row represented a unique ACT score, with the probabilities of success and percentages at or above anchored to the ACT scores. However, as noted earlier, panelists had access to this additional information prior to Round 1 and the only systematic difference between rounds resulted from discussion following the initial ratings. The rating sheets for Round 2 can be found on pages M7, ELA7, and S7, respectively, in Appendices E, F, and G.

Day Two

Day two began with a debriefing, during which panelists discussed the standard setting process so far and voiced any concerns or issues. One panelist indicated that the On Track cut score should signify the level of knowledge and skills needed for college access rather than success in specific college courses. Similarly, another math panelist suggested that the ACT Benchmarks were too rigorous, and that it would be preferable for the On Track cut scores to indicate the level of knowledge and skills needed for placement into the lowest level credit-bearing math course rather than College Algebra. Panelists were also interested in additional outcomes, such as career success, which were not available.

Round 2 Results and Discussion

The primary data sources of interest were briefly reviewed, then panelists reviewed the Round 2 results as shown in Figure 7, Table 10, and Table 11. Compared to Round 1, the Round 2 On Track cut score for Mathematics was the same, the cut score for ELA increased by 1 point, and

the cut score for Science decreased by 1 point. By comparing Figure 3 to Figure 7 and Table 8 to Table 10, it can be seen that within group consensus was greater in Round 2 than in Round 1.

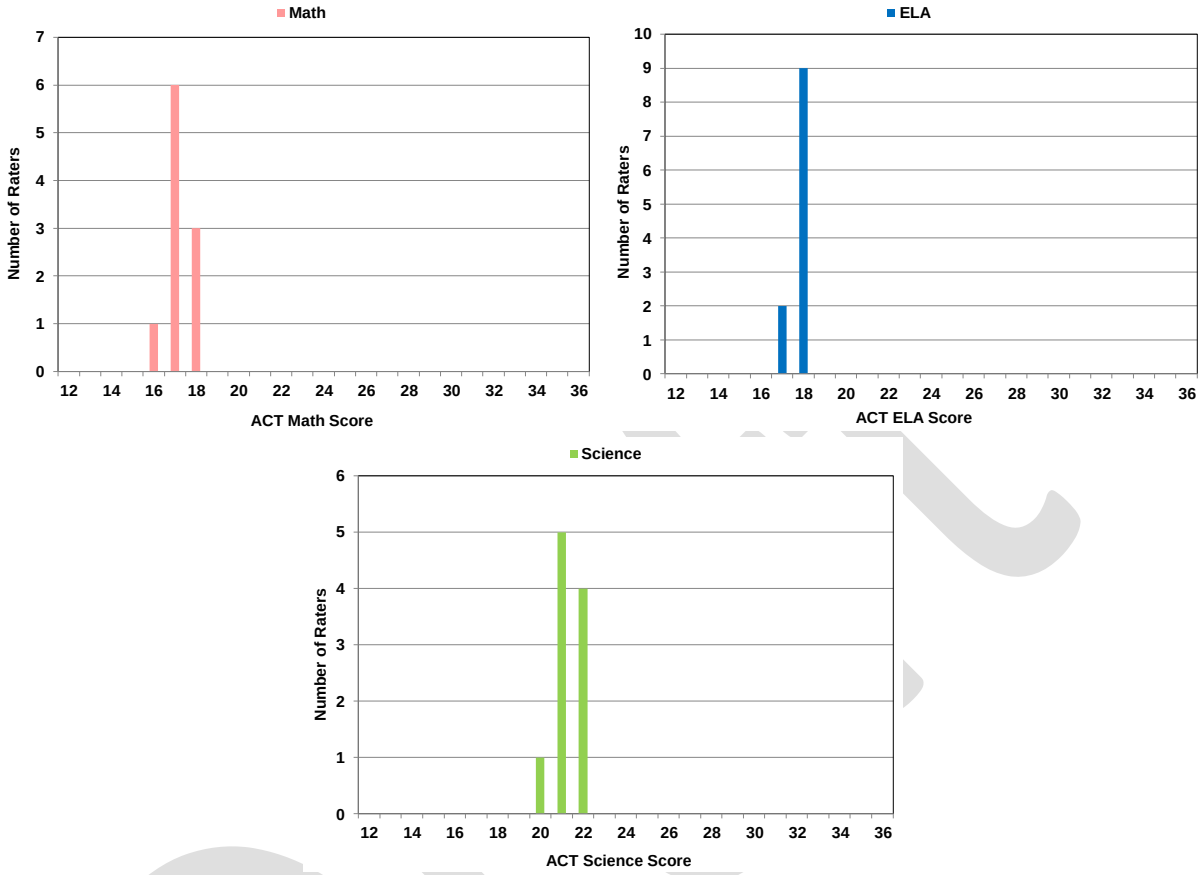

Figure 7. Round 2 On Track ratings by ACT score.

Table 10
Descriptive Statistics for Round 2 On Track Ratings

On Track Round 2	ACT Mathematics score	ACT ELA score	ACT Science score
Range	16–18	20–22	17–18
Median	17	21	18
Mode	17	21	18
SD	0.63	0.67	0.4

Table 11
Probabilities and Impact Data Associated with Round 1 and Round 2 On Track Cut Scores and ACT Benchmarks

	ACT score	Probability			Percentage at or above			
		A	**B or higher**	C or higher	**NE Juniors Census 2017**	**NE Juniors 2016**	**Census States Juniors 2016**	**National Juniors 2016**
Mathematics								
Round 1	17	0.07	0.26	0.55	62	78	58	71
Round 2	17	0.07	0.26	0.56	62	78	59	71
ACTB	22	0.20	0.51	0.73	31	44	28	41
ELA								
Round 1	17	0.09	0.38	0.69	62	83	55	72
Round 2	18	0.14	0.43	0.71	54	78	47	66
ACTB	20	0.19	0.51	0.76	41	67	35	54
Science								
Round 1	22	0.16	0.46	0.75	33	48	30	43
Round 2	21	0.12	0.41	0.71	42	57	39	52
ACTB	23	0.18	0.51	0.79	30	41	26	38

As in Round 1, multiple sources of comparative evidence were summarized on one line representing percentage at or above by subject (Figure 8, Figure 9, and Figure 10). The only difference was that the new figures also included the Round 2 cut score recommendations, illustrating the change in impact as the recommended cut scores increased or decreased.

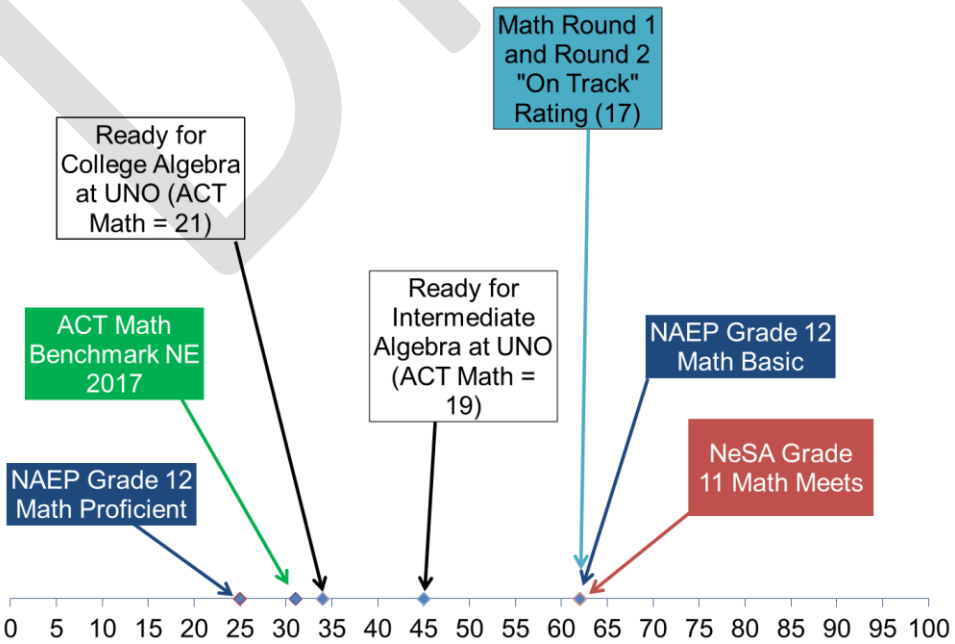

Figure 8. Math Round 2 On Track level comparative information.

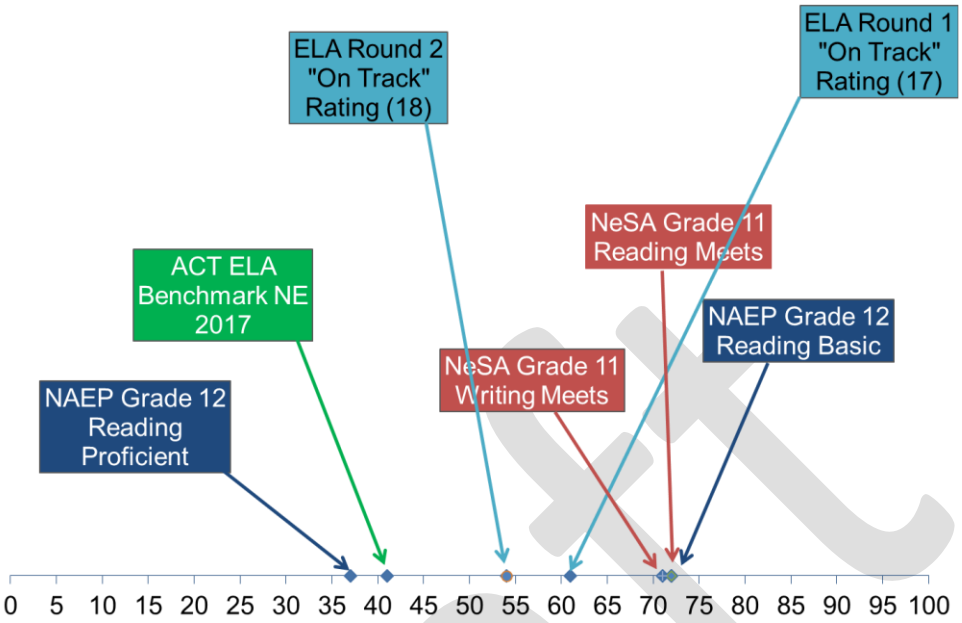

Figure 9. ELA Round 2 On Track level comparative information.

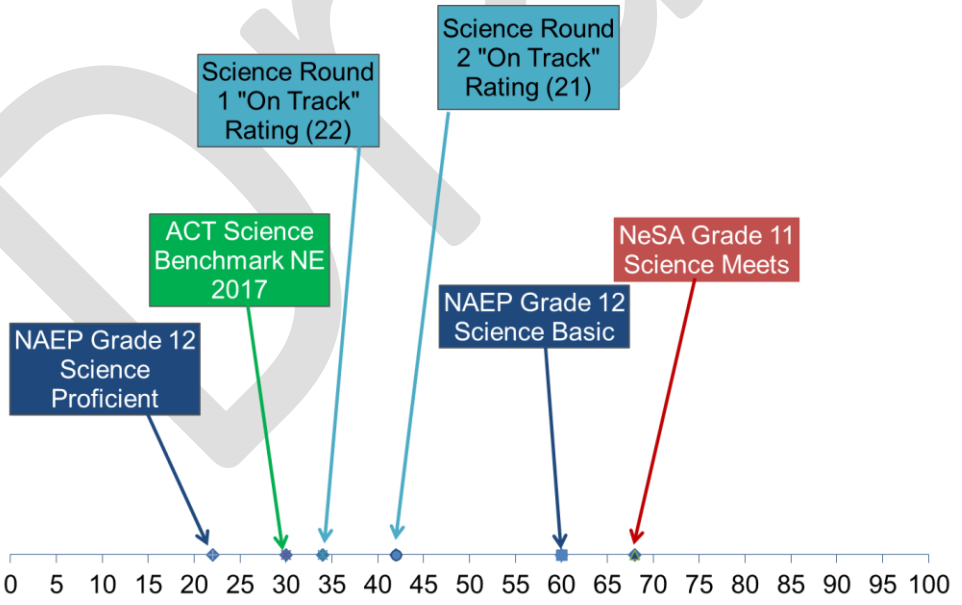

Figure 10. Science Round 2 On Track level comparative information.

Panelists reviewed ACT performance data for Nebraska juniors in 2015, 2016, and 2017. As expected, given the change in testing population, average ACT performance for Nebraska juniors declined between 2016 and 2017. The percentages of Nebraska juniors meeting the ACT College Readiness Benchmarks were shown next to the percentages of students attaining the Meets or Exceeds performance levels on NeSA in 2015 and 2016. Then, to address questions about

possible academic growth between junior and senior year of high school, the panelists reviewed average ACT score differences between juniors and graduating cohorts, which reflect most recent ACT scores. Consistent with the Camara and Allen (2017) study, those results suggested very little improvement in ACT scores between junior and senior year. As a final point of reference, panelists reviewed the cut score decisions made in four other states using the ACT for accountability. Two states adopted the ACT Benchmarks, one state adopted Proficient cut scores within one point of the ACT Benchmarks, and one state adopted Proficient cut scores 2-3 points lower than the ACT Benchmarks. Panelists participated in whole-group discussion after reviewing the Round 2 results and comparative impact data and additional reference information.

Round 3: Approaching and Exceeding Levels–Identifying Borderline Achievement by Probability of Success and ACT Score

The process for identifying Approaching and Exceeding achievement and establishing corresponding cut scores was summarized, along with a review of comparative examples of NAEP, SBAC, and PARCC Basic and Advanced achievement level descriptors. Time was allowed for group discussion, and then panelists split into subject area-specific groups to discuss what it means to be minimally Approaching and minimally Exceeding in their respective subject areas. The panelists were instructed to consider what the minimally Approaching student knows, how the minimally Approaching student differs from the Developing student, and the minimally Approaching student's chances of earning a B or higher and a C or higher in a credit-bearing college course. Likewise, panelists considered what the minimally Exceeding student knows, how the minimally Exceeding student differs from the On Track student, and the minimally Exceeding student's chances of earning a B or higher and a C or higher in a credit-bearing college course. ACT staff also explained to the panelists that the standard error of measurement on ACT subject-area tests is approximately 2 points. If adjacent cut scores were set 2 or fewer points away from each other, the cut scores would not be statistically distinguishable and numerous classification errors would likely result. Thus, it would be difficult to defend having adjacent cut scores that are 2 or fewer points apart.

After discussion, ACT staff gave instructions for the Round 3 Approaching and Exceeding ratings process. The rating sheets for Round 3 were in the same format as those used for Round 2, and can be found on pages M7, ELA7, and S7, respectively, in Appendices E, F, and G. Panelists were instructed to highlight two rows: one for their rating of minimally Approaching, and one for their rating of minimally Exceeding.

Round 3 Results and Discussion

Figure 11 shows the distributions of the Round 3 Approaching and Exceeding ratings. All distributions showed a clear central tendency except for Math Exceeding, which was bimodal. Table 12 summarizes the distribution of Round 3 Approaching and Exceeding ratings, and Table 13 shows the probabilities and impact data associated with the Round 2 and Round 3 ratings. The science panel exhibited the highest agreement during Round 3. Ratings for the ELA panel had greater variability, especially for the Exceeding cut score. The math ratings for Approaching were tightly distributed, but the distribution of Exceeding ratings indicated significant division among the math panelists in terms of defining the minimally Exceeding student.

As a reminder, subject area scores should not be directly compared because the ACT Mathematics, ELA, and Science scales are independently scaled (e.g., a 23 in Math is not necessarily the same achievement level as a 23 in ELA). However, cut scores may be compared in by their associated probabilities of earning an A, B or higher, and C or higher.

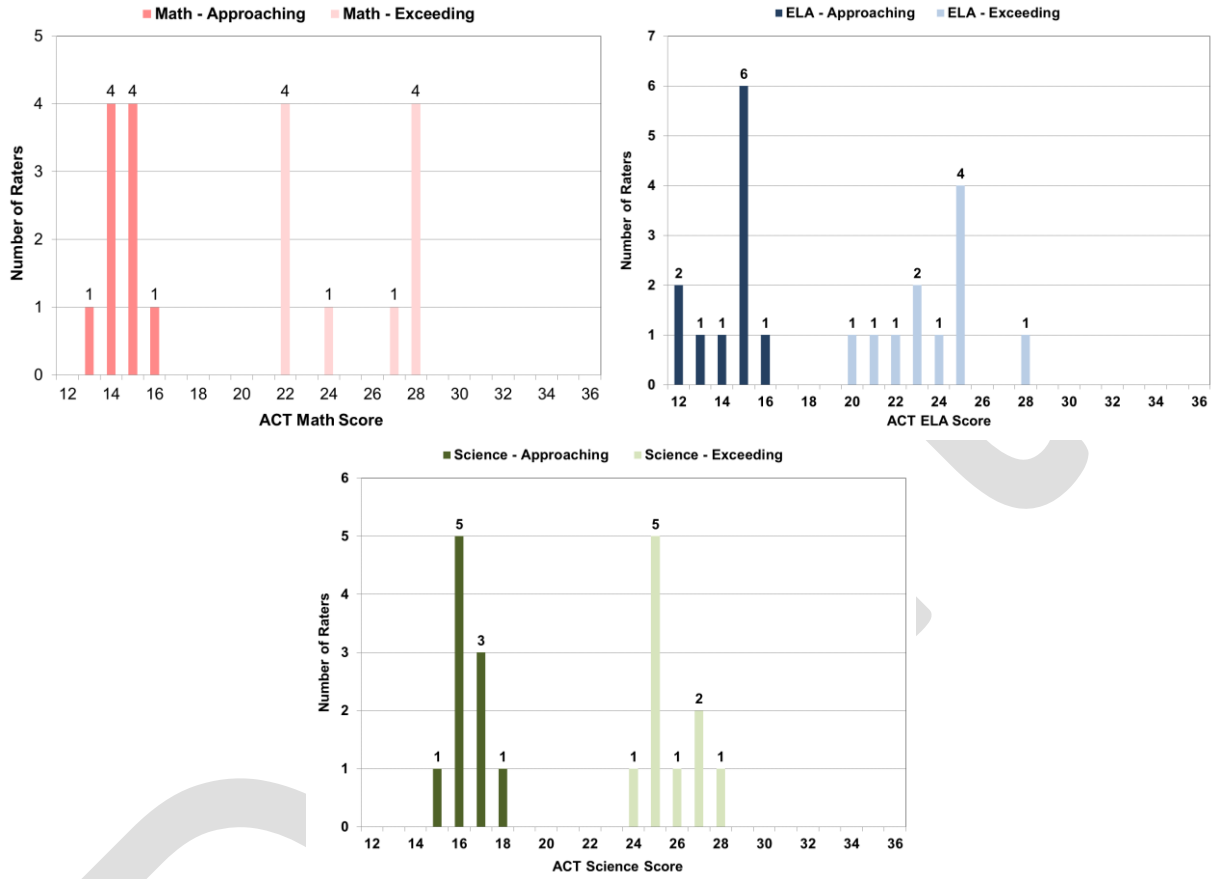

Figure 11. Round 3 Approaching and Exceeding ratings by ACT score.

Table 12
Descriptive Statistics for Round 3 Approaching and Exceeding Ratings

	Approaching			Exceeding		
Round 3	**ACT Mathematics Score**	**ACT ELA score**	**ACT Science score**	**ACT Mathematics score**	**ACT ELA score**	**ACT Science score**
Range	13–16	12–16	15–18	22–28	20–28	24–28
Median	15	15	16	26	24	25
Mode	--	15	16	--	25	25
SD	0.85	1.35	0.84	2.92	2.24	1.25

Table 13
Probabilities and Impact Data Associated with Round 3 Approaching, On Track, and Exceeding Ratings

	ACT score	Probability			Percentage at or above			
		A	**B or higher**	C or higher	**NE Juniors Census 2017**	**NE Juniors 2016**	**Census States Juniors 2016**	**National Juniors 2016**
Mathematics								
Exceeding	27	0.45	0.73	0.85	10	15	8	17
On Track	17	0.07	0.26	0.56	62	78	59	71
Approaching	15	0.04	0.19	0.46	88	95	87	92
ELA								
Exceeding	24	0.32	0.66	0.84	18	42	16	30
On Track	18	0.14	0.43	0.71	54	78	47	66
Approaching	15	0.09	0.31	0.61	74	91	67	81
Science								
Exceeding	25	0.25	0.60	0.84	16	23	14	23
On Track	21	0.12	0.41	0.71	42	57	39	52
Approaching	16	0.04	0.19	0.52	76	90	78	85

Impact evidence corresponding to the Approaching and Exceeding ratings was presented by subject area to panelists. The panelists also reviewed expected college enrollment and completion rates for students who just met the Approaching and Exceeding cut scores recommended in Round 3. Each panel then discussed how their ratings compared within and across subject areas, how the impact data may have affected their ratings, and which data were most influential in making their ratings.

Final Round: Approaching and Exceeding Levels

Before the final round of ratings, the panelists reviewed the impact data based on the Round 2 and Round 3 ratings. In this case, the data were presented as the percentages of spring 2017 examinees that would be classified as Developing, Approaching, On Track, and Exceeding. Presenting the data as in Figure 12 allows for straightforward comparisons across subject areas. For example, after Round 3, the percentage of On Track students in Math (52%) was twice the percentage of On Track students in Science (26%). Such results could reflect different perceptions of minimally On Track (and minimally Exceeding) students as well as different perceptions about the academic strengths and weaknesses of students in Nebraska.

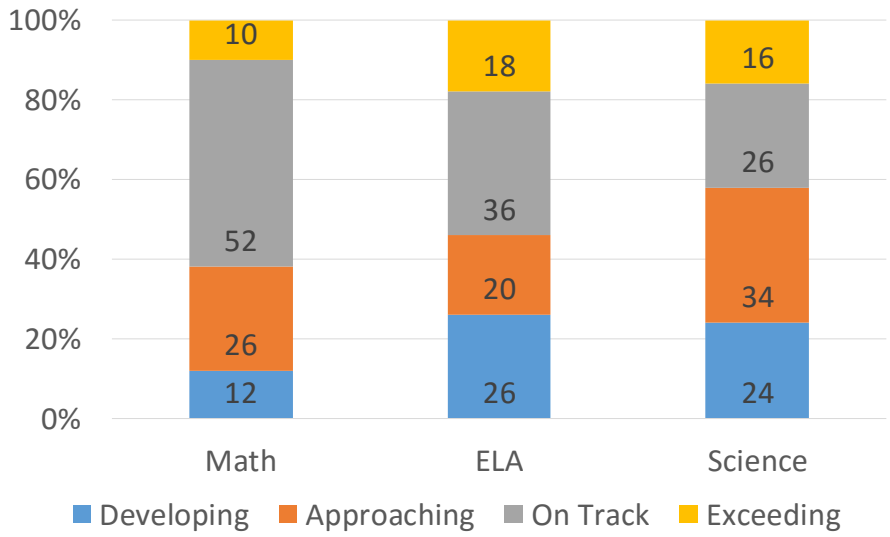

Figure 12. Impact data after Rounds 2 and 3 ratings.

After the impact data were presented, panelists were given time for final discussion, followed by final ratings for the Approaching, On Track, and Exceeding cut scores. The rating sheets for the Final Round were the same format as those used for Rounds 2 and 3 (Tables M7, ELA7, and S7, respectively, in Appendices E, F, and G). Panelists were instructed to highlight three rows—one for Approaching, one for On Track, and one for Exceeding—on the rating sheet for their subject area. After the final ratings, panelists completed an evaluation form before adjourning (see Appendix D for a copy of the evaluation form and summary of results).

Final Round Results

Figure 13 shows the distributions of the final ratings. Agreement within each subject area was strong, with notable improvements in agreement for the Math and ELA Exceeding ratings. Several of the median ratings changed in the Final Round (Table 14). The Math On Track median rating increased from 17 to 18, with a notable shift in the range of ratings, such that the Round 2 maximum rating of 18 was equal to the final round minimum rating, and the maximum rating in the final round increased by four points to 22. The Exceeding median rating in Math decreased from 27 to 26. In ELA, the On Track median stayed the same, but the Approaching median changed from 15 to 14, and the Exceeding median increased from 24 to 25. With an increase from 25 to 26, the Exceeding median was the only Science cut score that changed in the Final Round.

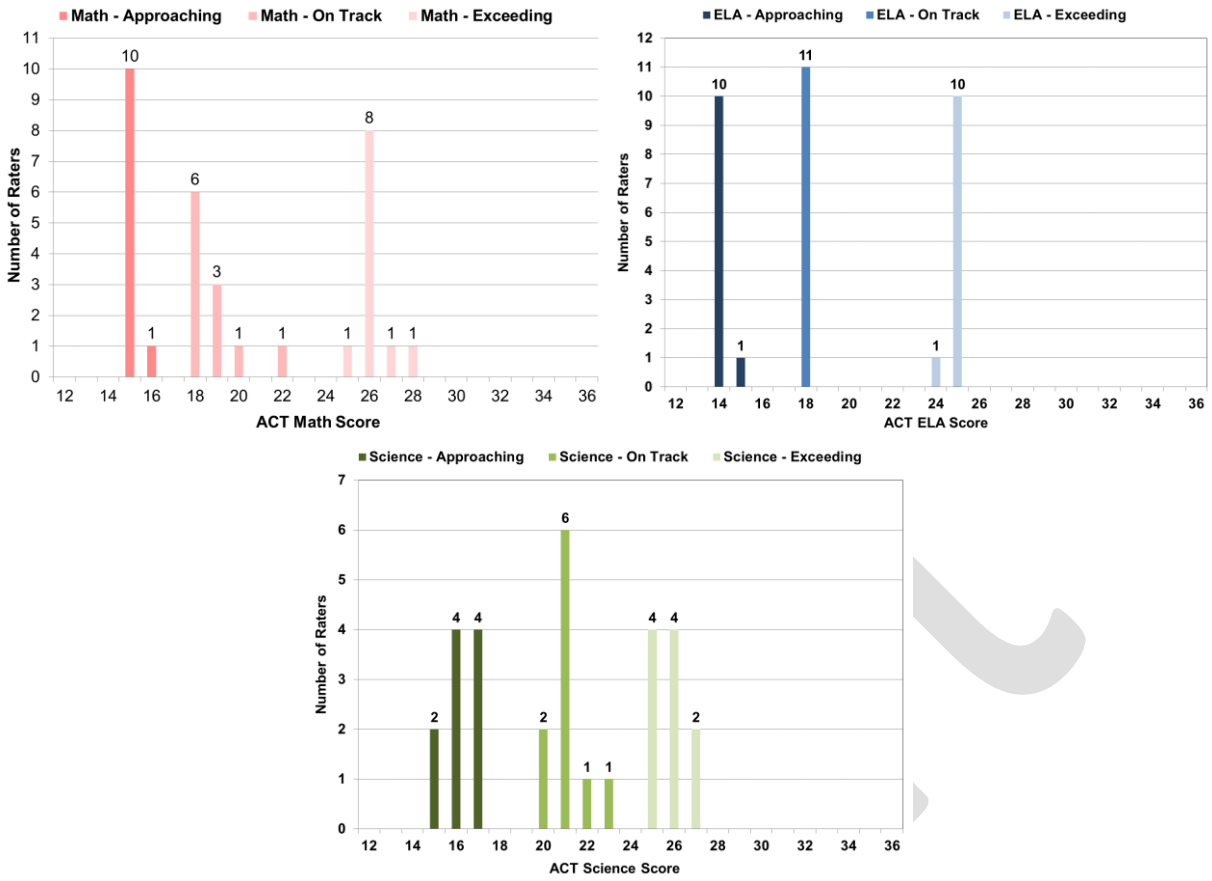

Figure 13. Final Approaching, On Track, and Exceeding ratings by ACT score.

Table 14
Descriptive Statistics for Final Approaching and Exceeding Ratings

Round 3	Approaching		
	ACT Mathematics score	**ACT ELA score**	**ACT Science score**
Range	15-16	14-15	15-17
Median	15	14	16
Mode	15	14	--
SD	0.30	0.30	0.79

Round 3	On Track		
	ACT Mathematics score	**ACT ELA score**	**ACT Science score**
Range	18–22	18	20–23
Median	18	18	21
Mode	18	18	21
SD	1.25	0	0.88

Round 3	Exceeding		
	ACT Mathematics score	**ACT ELA score**	**ACT Science score**
Range	25–28	24–25	25–27
Median	26	25	26
Mode	26	25	--
SD	0.75	0.30	0.79

Table 15 shows the probabilities of earning A, B or higher, and C or higher as well as the percentages of students at or above the recommended cut scores based on the Final Round of ratings. As shown in Figure 14, the changes in median ratings that occurred in the Final Round resulted in somewhat greater consistency across subject areas in terms of the percentages of students classified in each of the four performance levels.

The PowerPoint slides in Appendix H contain additional outcome information such as college enrollment and college completion rates for each of the cut scores.

Table 15
Probabilities and Impact Data Associated with Final Approaching, On Track, and Exceeding Ratings

	ACT score	Probability			Percentage at or above			
		A	**B or higher**	C or higher	**NE Juniors Census 2017**	**NE Juniors 2016**	**Census States Juniors 2016**	**National Juniors 2016**
Mathematics								
Exceeding	26	0.39	0.69	0.83	14	21	12	22
On Track	18	0.09	0.30	0.60	51	68	49	62
Approaching	15	0.04	0.19	0.46	88	95	87	92
Exceeding	25	0.36	0.70	0.86	14	37	12	25
On Track	18	0.14	0.43	0.71	54	78	47	66
Approaching	14	0.08	0.28	0.58	80	94	74	85
Science								
Exceeding	26	0.30	0.64	0.86	11	17	9	17
On Track	21	0.12	0.41	0.71	42	57	39	52
Approaching	16	0.04	0.19	0.52	76	90	78	85

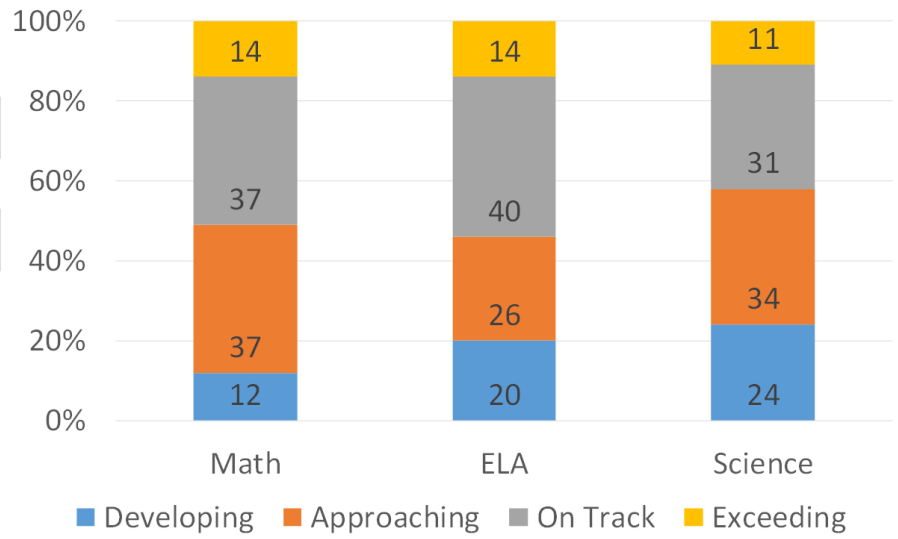

Figure 14. Impact data after Final Round ratings.

Empirical Performance Level Predictors

Performance level descriptors (PLDs) are statements about the knowledge and skills that a student who meets the performance standard are expected to know and be able to demonstrate. When a traditional content-based standard setting approach is used to set performance standards,

the PLDs are typically developed in advance of the standard setting. In this approach, the PLDs are used to select exemplar items that will be used during the standard setting workshop. Exemplar items are selected to reflect the knowledge and skills expected of students at a given performance level, and the standard setting task is to select items that a student at that performance level has a reasonable probability of answering correctly.

Because an empirical standard setting focuses on probabilities of success in college courses rather than the specific content measured by the assessment, the standard setting occurs prior to creating PLDs. In this approach, the cut scores are used to select exemplar items that will be used during the PLD workshop. Exemplar items are selected that students scoring at or near that cut score have a reasonable probability of answering correctly, and the PLD task is to create statements about the levels of knowledge and skills that the exemplar items measure.

For Nebraska, the PLDs will be developed after the final cut scores are approved. However, empirical PLD statements can still be made about the performance of students scoring at each proposed cut score. These statements reflect the probabilities of postsecondary success rather than test content. Provided below are empirical PLD statements resulting from this standard setting. Note that these are based on the recommended scores; the final cut scores have not been approved at the time this report was prepared. If adjustments are made to the proposed cut scores, the PLD statements can be adjusted accordingly to reflect the probabilities of success associated with the approved cut scores.

Math Empirical PLDs

Students meeting the On Track standard in math (18) have a 30% chance of earning a B or higher in a first-year credit-bearing college algebra course, and a 60% chance of earning a C or higher in the course.

Students meeting the Approaching standard in math (15) have a 19% chance of earning a B or higher in a first-year credit-bearing college algebra course, and a 46% chance of earning a C or higher in the course.

Students meeting the Exceeding standard in math (26) have a 69% chance of earning a B or higher in a first-year credit-bearing college algebra course, and an 83% chance of earning a C or higher in the course.

ELA Empirical PLDs

Students meeting the On Track standard in ELA (18) have a 43% chance of earning a B or higher in a first-year credit-bearing college course in English composition or social science, and a 71% chance of earning a C or higher in the course.

Students meeting the Approaching standard in ELA (14) have a 28% chance of earning a B or higher in a first-year credit-bearing college course in English composition or social science, and a 58% chance of earning a C or higher in the course.

Students meeting the Exceeding standard in ELA (25) have a 70% chance of earning a B or higher in a first-year credit-bearing college course in English composition or social science, and an 86% chance of earning a C or higher in the course.

Science Empirical PLDs

Students meeting the On Track standard in science (21) have a 41% chance of earning a B or higher in a first-year credit-bearing college biology course, and a 71% chance of earning a C or higher in the course.

Students meeting the Approaching standard in science (16) have a 19% chance of earning a B or higher in a first-year credit-bearing college biology course, and a 52% chance of earning a C or higher in the course.

Students meeting the Exceeding standard in science (26) have a 64% chance of earning a B or higher in a first-year credit-bearing college biology course, and an 86% chance of earning a C or higher in the course.

Reliability of Ratings

Unlike a content-based standard setting where panelists provide ratings of many items, each panelist in the empirical standard setting only provided an overall rating for each cut score (three rounds of ratings for On Track and two rounds for Approaching and Exceeding). Additionally, panelists were expected to adjust their ratings as they considered additional evidence, so inter-rater reliability is not a meaningful measure of reliability in this context. Therefore, reliability of cut scores is presented as within-group consensus. Table 16 contains the percentages of panelists who agreed upon a single rating within subject area for each round of ratings. This is a coarse measure of agreement because it ignores ratings that were very close (e.g., ±1 point) from the modal rating. For example, in Round 1, 5 out of 10 math panelists (50%) chose a B or higher success probability of .26, but there were 2 panelists who selected a success probability of .24.

For the math panel, the agreement percentage was consistent for the On Track cut score through three rounds of ratings. Agreement on the Approaching and Exceeding cut scores increased substantially from Round 3 to the Final Round. Agreement within the ELA panel increased consistently from round to round, and Final Round agreement with very strong. Science panel agreement for the On Track cut score improved through three rounds of ratings. Approaching and Exceeding agreement levels were fairly consistent in Round 3 and the Final Round.

Figure 15 illustrates between-group agreement in terms of chance of success associated with median ratings across rounds. For the On Track ratings, the spread of the B or higher median ratings decreased between Round 1 (panelist ratings ranged between 26–46% chance of B or higher) and Round 2 (26–43% chance) and again between Round 2 and Round 4 (30–43%). There were only two rounds of ratings for the Approaching and Exceeding cut scores. Between-group agreement for the Approaching cut score increased from Round 3 (19–31% chance) to Round 4 (19–28%). Likewise, between-group agreement for the Exceeding cut score increased between Round 3 (60–73%) and Round 4 (64–70%). Results for the C or higher criterion showed similar trends for On Track and Approaching. For Exceeding, the between-group agreement was very high in both rounds.

Table 16
Percent of Agreement for Each Round by Subject Area

	Percent of Agreement		
	Mathematics	ELA	Science
	On Track		
Round 1	50%	36%	33%
Round 2	60%	82%	50%
Final Round	55%	100%	60%
	Approaching		
Round 3	40%	55%	50%
Final Round	91%	91%	40%
	Exceeding		
Round 3	40%	36%	50%
Final Round	73%	91%	40%

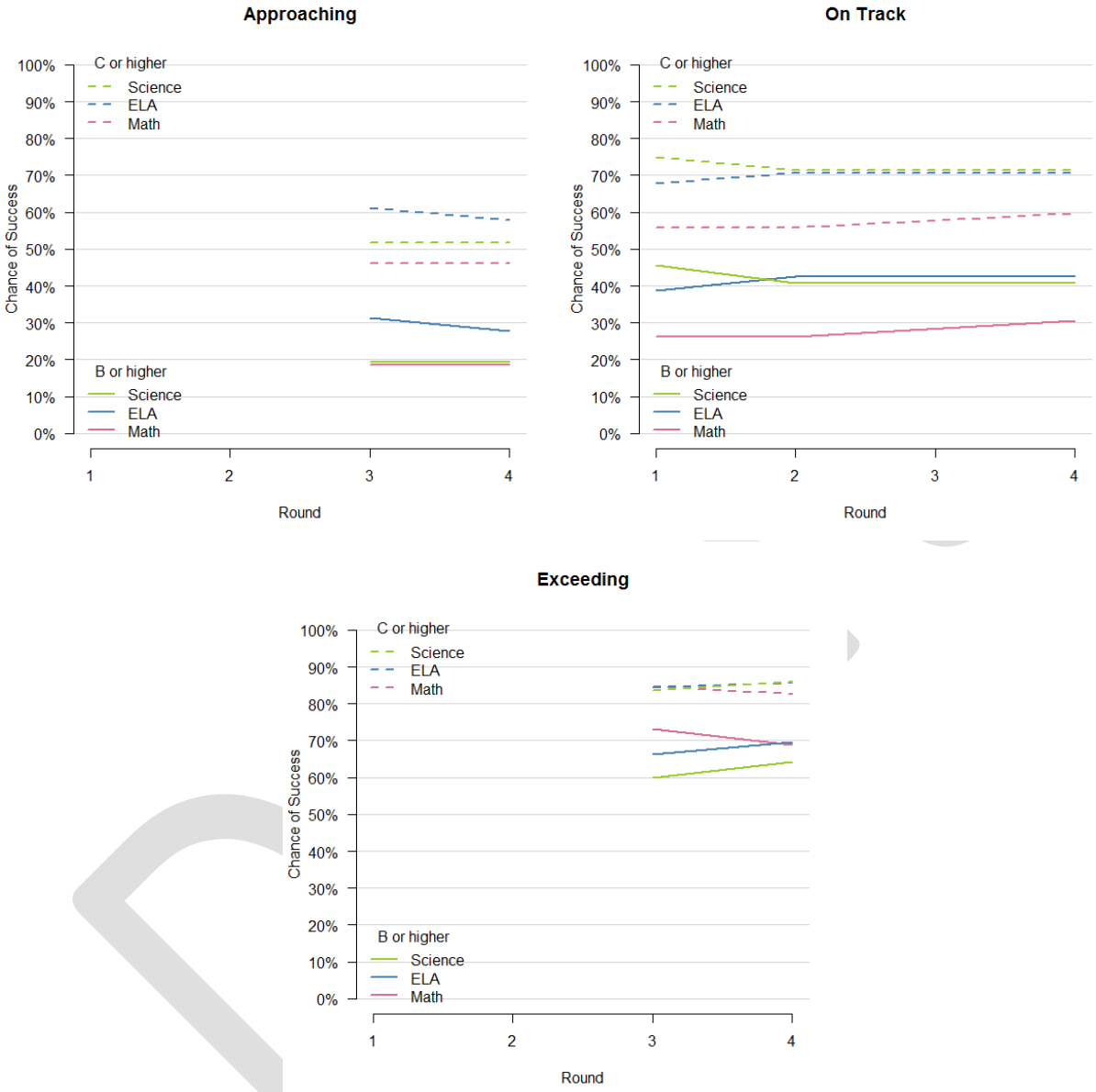

Figure 15. B or higher and C or higher probabilities associated with median ratings.

Summary of Discussion and Rationale for Recommended Cut Scores

Several themes emerged during discussions around the rationale panelists had for the cut scores that they set, particularly for setting the On Track performance level. When defining On Track, one panelist expressed a primary concern with access to credit-bearing courses rather than success in specific college courses. Some panelists agreed with this way of thinking. For example, another panelist suggested that the ACT Benchmarks were too rigorous, and that cut scores should be based on the scores needed for placement into the lowest level credit-bearing courses, rather than the most common first-year credit-bearing courses. Other panelists, however,

countered that access and placement are not sufficient to define college readiness because many students who are just barely eligible for credit-bearing courses are underprepared and ultimately drop out. One panelist suggested that the Exceeding cut score should be set at the ACT score needed to eligible for a certain scholarship. Some panelists were concerned because they did not want the cut scores to send conflicting messages; for example, what would be the reaction if a student is considered college ready by Nebraska's definition but not by ACT's definition? Some panelists wanted to set cut scores below the ACT Benchmarks because they believed that students' performance would improve between their junior and senior year. Other panelists cited the evidence provided by ACT staff that score gains from junior to senior year are typically around one point, and did not want to set the cut scores too low. Some panelists seemed overly concerned with comparing impact to results from the NeSA exam and were reluctant to recommend cut scores that could significantly decrease the percentages of students meeting each cut score compared to past results.

Process Evaluation Questionnaire Results

At the conclusion of the standard setting meeting, all panelists completed a Process Evaluation Questionnaire. This questionnaire was intended to gauge the level of understanding of panelists, to evaluate the standard setting process, and gather feedback that can be used to improve the process in future studies. Panelists responded to the questionnaire items on a 1–5 scale. This section summarizes typical responses, but full results can be found in Appendix D. Overall, the panelists indicated that advance communications to prepare them for the meeting were adequate, and they understood the purpose of the meeting very well. The panelists felt that instructions were moderately or very clear, and they understood their assigned tasks very well. Consistent with that finding, the panelists were very comfortable using the concept of "borderline" performance to make their ratings, and they were very confident in their own cut score recommendations.

When reviewing results, the panelists understood the median cut scores very well. They understood first-year credit-bearing courses and probabilities of success very well or extremely well. The panelists understood the difference between success probabilities and impact data very well or extremely well, and they were very comfortable using the impact data to evaluate the reasonableness of their recommended cut scores. The panel described the performance level setting method as moderately or very effective. In general, panelists felt that they were given an appropriate amount of time for explanation and discussion. The panelists felt that their input was very valued by their peers, and they did not feel pressured by their peers to make their cut score recommendations agree. Most panelists felt no pressure from staff or other panelists to make their recommendations higher or lower or stay the same, but some panelists indicated that they felt slightly pressured. In general, panelists felt that the process allowed them to use their best judgement moderately well or very well. Finally, the panelists reported that the final recommended cut scores were moderately to very defensible and moderately to very reasonable.

Optional open-ended comments revealed mostly positive sentiments about the standard setting process and appreciation for the opportunity to be included in the process. Several panelists expressed concerns that there was not a consensus on the definitions of "On Track" or "college and career ready." Two panelists expressed concerns that the On Track score for Math was too low, and others suggested improvements to the process in terms of the amount of time spent reviewing data versus time spent in discussion.

Adoption of Cut Scores

The final cut scores have not been approved at the time that this report was written.

Summary and Conclusions

Using the ACT for accountability purposes allows Nebraska to affirm their students' readiness for college and career using an assessment that is a strong predictor of college success. Panelists set Approaching, On Track, and Exceeding cut scores for Mathematics, ELA, and Science relative to students' readiness for college and career. These cut scores, if approved, represent higher standards than those previously set on the NeSA; the percentages of students scoring at or above On Track will likely be smaller than in previous years. However, this change corresponds to the increased rigor and higher standards that are now required of Nebraska high schools. Moreover, the reduction of the gaps between high school achievement and college and career readiness, and the greater articulation between them, will likely smooth the way for Nebraska high school students as they transition out of high school.

References

ACT. (2016). The condition of college and career readiness 2016: Nebraska key findings. Iowa City, IA: Author. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/state28_Nebraska_Web_Secured.pdf.

Allen, J., & Sconing, J. (2005). Using ACT Assessment scores to set benchmarks for college readiness. (ACT Research Report No. 2005-3). Iowa City, IA: ACT, Inc. Retrieved from http://www.act.org/content/act/en/research/pdfs/using-act-assessmentscorestosetbenchmarksforcollegereadiness.html

Allen, J. (2013). Updating the ACT College Readiness Benchmarks. (ACT Research Report No. 2013-6). Iowa City, IA: ACT, Inc. Retrieved from http://www.act.org/content/act/en/research/pdfs/updating-the-actcollegereadinessbenchmarks.html

Bandeira de Mello, V., Bohrnstedt, G., Blankenship, C., and Sherman, D. (2015). Mapping State Proficiency Standards Onto NAEP Scales: Results From the 2013 NAEP Reading and Mathematics Assessments (NCES 2015-046). U.S. Department of Education, Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/nationsreportcard/subject/publications/studies/pdf/2015046.pdf.

Camara, W. J., & Allen, J. (2017). Does testing date impact student scores on the ACT? (ACT Technical Brief, June 2017). Iowa City, IA: ACT, Inc. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/R1643-test-date-impact-on-act-scores-2017-06.pdf.

Fields, R., & Parsad, B. (2012). *Tests and Cut Scores Used for Student Placement in Postsecondary Education: Fall 2011*. Washington, DC: National Assessment Governing Board. Retrieved from https://www.nagb.org/content/nagb/assets/documents/commission/researchandresources/test-and-cut-scores-used-for-student-placement-in-postsecondary-education-fall-2011.pdf.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., Gebrekristos, S., Zhang, J., Rathbun, A., Barmer, A., Bullock Mann, F., and Hinz, S. (2017). The Condition of Education 2017 (NCES 2017-144). U.S. Department of Education.

Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubs2017/2017144.pdf.

Noble, J., & Radunzel, J. (2012). Predicting long-term college success through degree completion using ACT composite score, ACT benchmarks, and high school grade point average. (ACT Research Report No. 2012-5). Iowa City, IA: ACT, Inc.

Noble, J., & Sawyer, R. (2013). A study of the effectiveness of developmental courses for improving success in college. (ACT Research Report No. 2013-1). Iowa City, IA: ACT, Inc.

Radunzel, J., Westrick, P., Bassiri, D., & Li, D. (2017). Development and validation of a preliminary ELA readiness benchmark based on the ACT ELA score. (ACT Research Report No. 2017-9). Iowa City, IA: ACT, Inc.

U.S. Department of Education, National Center for Education Statistics. (2004). The Condition of Education 2004 (NCES 2004–077). Washington, DC: U.S. Government Printing Office. Retrieved from https://nces.ed.gov/pubs2004/2004077.pdf.

Appendices

Appendix A: Annotated Agenda
Appendix B: List of Panelists
Appendix C: Background Questionnaire
Appendix D: Evaluation Questionnaires
Appendix E: Mathematics Data Book
Appendix F: ELA Data Book
Appendix G: Science Data Book
Appendix H: Presentation Slides