The Validity of Low ACT Scores

YoungWoo Cho, PhD, Jeffrey Steedle, PhD, David Woodruff, PhD, and Dean Colton, PhD

Abstract

Different uses of the ACT Assessment emphasize different intervals on its 1–36 score scale. Its use by colleges for applicant selection and course placement emphasize the middle and upper intervals of its score scale, and these uses have been extensively validated (ACT, 2019a). Its uses by school districts and state departments of education to assess high school students and evaluate high school performance potentially utilize more of the ACT score scale. The purpose of this study was to investigate the meaningfulness of lower ACT scores to inform the appropriate use of the ACT as an every-student-test for all high school students, not just college bound high school students.

Predictive and concurrent criterion validity approaches were used in combination with nonparametric regression to examine the usefulness of low ACT scores. Predictive validity was demonstrated for the ACT Composite score using first-year college grade point average, and meaningful distinctions between ACT Composite scores were observed down to a score of 12. Concurrent validity between WorkKeys tests and the ACT math and reading tests was investigated using nonparametric regression. Meaningful distinctions among ACT math scores were observed down to a score of 11, and ACT reading score distinctions were observed down to a score of 7. Considering that nearly all students earn scores in the range where meaningful distinctions were observed, results indicate that nearly all high school students—not just college bound high school students—have meaningful ACT scores. Hence, it is appropriate to use the ACT as an every-student-test to assess students' achievement of the high school curriculum.

The Validity of Low ACT Scores

The ACT is an achievement test typically taken by junior and senior high school students (ACT, 2019a). Colleges and universities use ACT scores for applicant selection and course placement. In recent years, the ACT has been used increasingly by school districts and state departments of education to assess individual high school students and evaluate high schools. Uses of the ACT in secondary and postsecondary education are intimately related. The ACT is designed to measure knowledge and skills taught in high school, and its success in predicting academic performance in college depends on its success in measuring mastery of the high school curriculum. However, these different

uses of the ACT emphasize different intervals on its 1–36 score scale. The middle and upper intervals of the score scale are of greatest concern for college applicant selection (depending on the selectivity of the college). Assessment at the high school level potentially involves a broader range of the ACT score scale. The purpose of this study was to examine the validity of interpreting ACT scores by identifying the range of scores for which useful distinctions between examinees can be made—especially in the low score range. There are numerous ways to gather validity evidence for an assessment program. The focus of this study was predictive and concurrent criterion-related validity.

Predictive validity was investigated for the ACT composite score as a predictor of first-year college grade point average (FYGPA). The main battery of the ACT comprises four subject tests: English, math, reading, and science (writing is optional). The ACT composite score is the average of the four subject test scores. Concurrent criterion-related validity was investigated for the ACT mathematics and reading tests. The relationship between the ACT math test and the WorkKeys Applied Mathematics test was studied as was the relationship between the ACT reading test and the WorkKeys Reading for Information test. The WorkKeys tests are most often used by employers for job applicant selection and in K–12 education to assess high school students' workplace skills (ACT, 2019b).[1]

Predictive Criterion Validity

One method used to investigate the predictive validity of low ACT scores was to examine the relationship between ACT composite score and first-year college grade point average (FYGPA). If FYGPA varies in systematic and meaningful ways for students with low ACT scores, this would indicate predictive validity for low ACT scores. To investigate this notion, data from community colleges and four-year institutions in eight states were analyzed. These data were chosen because the institutions were less selective than many state and private institutions, so it was likely that more students with lower ACT scores were admitted and completed the first year at these institutions. From these eight states, data were obtained for a total of 192,962 students with both ACT scores and a reported FYGPA. The data were analyzed by means of a nonparametric regression of FYGPA on ACT composite score.

Table 1 provides relevant statistics for the nonparametric regression of FYGPA on ACT score. The primary statistics of interest are the FYGPA means for students with the same ACT composite score. Along with these conditional means, sample size (N), conditional FYGPA standard deviation (SD), and FYGPA standard error of the mean (SE) at each ACT Composite score are also included in Table 1. Note that the FYGPA standard deviations were fairly homogeneous (between 0.75 and 0.90) for ACT composite scores between 10 and 35. At the ACT composite score 36, the FYGPA standard deviation was only 0.60, but this was due to a ceiling effect caused by the 4.0 upper limit of FYGPA. The FYGPA standard errors of the mean were small for mid-range ACT Composite scores where sample sizes were large, but they increased at both ends of the score scale where sample sizes were smaller. The correlation between ACT composite score and FYGPA was 0.31.

[1]These analyses were conducted using data gathered prior to the WorkKeys updates that occurred in 2017. These updates were designed to make the WorkKeys tests (Applied Math, Workplace Documents, and Graphic Literacy) better assess contemporary workplace skills.

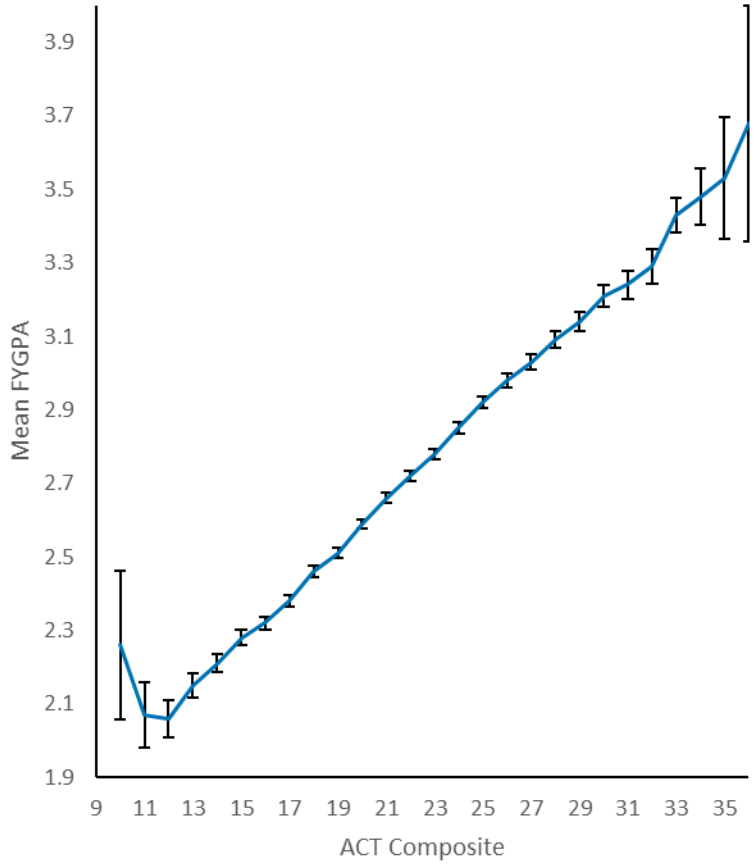
Table 1. FYGPA Statistics by ACT Composite Score

ACT Composite	N	FYGPA Mean	FYGPA SD	FYGPA SE
10	62	2.26	0.80	0.1016
11	357	2.07	0.84	0.0445
12	1166	2.06	0.86	0.0252
13	2716	2.15	0.88	0.0169
14	4732	2.21	0.86	0.0125
15	6952	2.28	0.84	0.0101
16	9291	2.32	0.84	0.0087
17	11694	2.38	0.83	0.0077
18	13779	2.46	0.83	0.0071
19	15124	2.51	0.84	0.0068
20	16974	2.59	0.83	0.0064
21	17130	2.66	0.85	0.0065
22	16360	2.72	0.86	0.0067
23	15079	2.78	0.88	0.0072
24	13044	2.85	0.89	0.0078
25	11417	2.92	0.89	0.0083
26	9837	2.98	0.88	0.0089
27	7842	3.03	0.90	0.0102
28	6161	3.09	0.89	0.0113
29	4435	3.14	0.90	0.0135
30	3529	3.21	0.86	0.0145
31	2303	3.24	0.90	0.0188
32	1473	3.29	0.87	0.0227
33	980	3.43	0.75	0.0240
34	415	3.48	0.79	0.0388
35	96	3.53	0.81	0.0827
36	14	3.68	0.60	0.1604

Figure 1 shows a plot of the FYGPA conditional means at each ACT Composite score along with plus and minus two standard errors of the mean error bands. Note that sample size below an ACT composite score of 10 was too small to provide useful results. As shown in Figure 1, FYGPA means increased steadily in a nearly perfect linear fashion starting at an ACT composite score of 12. As note previously, the SEs were larger near the lower and upper ends of the ACT score scale. This reveals less precise estimation of the conditional means (i.e., more random sampling error), yet the linear increasing pattern was consistent at the high end of the scale. At the low end of the scale, there was little differentiation in mean FYGPA between ACT composite scores 11 and 12, and mean FYGPA decreased from ACT composite score 10 to 11. Thus,

differences between ACT composite scores below 12 did not indicate which students were more or less likely to earn a higher FYGPA.

Figure 1. Mean FYGPA versus ACT Composite Score with ±2 Standard Errors of the Mean Bands



The correlation of 0.31 between ACT composite score and FYGPA was modest in magnitude. This was due to the relatively constant but large conditional FYGPA standard deviations. Though the strength of the relationship between the two variables was modest, there was a systematic and meaningful relationship between the two variables in the range of ACT composite score from 12 to 36.

One state included in the study administered the ACT to all eleventh-grade high school students regardless of the students' plans for college. Analysis of the data from this state revealed very few high school eleventh graders with ACT composite scores below 12. Specifically, over 120,000 students took the ACT in that state, and only 2.6% had ACT composite scores below 12. This finding supports the appropriateness of using the ACT as a test of achievement related to the high school curriculum for all students. Under such usage conditions, nearly all students would obtain scores that meaningfully indicate who is more or less likely to exhibit academic achievement in the first year of postsecondary studies.

Concurrent Criterion Validity

Another method for investigating the validity of interpreting low ACT scores is to examine the relationship between an ACT test and another test of similar content administered at the same time. If a systematic and meaningful relationship between the pair of tests is observed—and if that relationship holds for low ACT scores—that provides evidence of concurrent validity for low ACT scores insofar as both tests already have been validated by other means. To investigate the validity of interpreting low ACT scores, this study included analyses to examine the relationship between ACT math scale scores and WorkKeys Applied Mathematics raw scores and the relationship between ACT reading scale scores and WorkKeys Reading for Information raw scores. All four of these tests have been administered broadly, and evidence for the valid interpretation and use of scores is documented in their respective technical manuals (ACT, 2019a, 2019b).

The data analyzed were from a spring administration of the ACT and the concurrent administration of the tests to all high school juniors attending public schools in a certain state. Nonparametric regression was applied to evaluate ACT scores as predictors of WorkKeys scores and to determine the range of ACT scores for which there was a meaningful relationship between ACT and WorkKeys scores. Statistics for the math regression are presented in Table 2, and statistics for the reading regression are presented in Table 3. Again, these tables contain the conditional means, sample sizes, conditional standard deviations, and standard errors of the mean.

A plot of the regression of WorkKeys Applied Mathematics on ACT math is presented in Figure 2. There was a strong correlation of 0.79 between these tests. Figure 3 provides a plot of the regression of WorkKeys Reading for Information on ACT reading. In this case, the test scores exhibited a correlation of 0.64. Both regression plots show plus and minus two standard errors of the mean bands about each conditional mean. Tables 2 and 3 reveal that the conditional standard deviations were not homogeneous but tended to decrease as ACT score increased, especially for the mathematics regression.

Table 2. WorkKeys Applied Mathematics Score Statistics by ACT Math Score

ACT Math	N	WorkKeys Mean	WorkKeys SD	WorkKeys SE
9	31	12.03	4.57	0.8208
10	84	12.31	3.93	0.4288
11	496	12.73	3.92	0.1760
12	1622	13.65	4.06	0.1008
13	3478	14.28	4.03	0.0683
14	7719	15.58	4.00	0.0455
15	9218	17.20	3.81	0.0397
16	9254	18.65	3.49	0.0363
17	9135	20.06	3.15	0.0330
18	8920	21.14	3.03	0.0321
19	5460	21.98	2.94	0.0398
20	5442	22.77	2.85	0.0386
21	2551	23.26	2.93	0.0580
22	5111	23.71	2.84	0.0397
23	4783	24.32	2.73	0.0395
24	4428	24.93	2.69	0.0404
25	5974	25.68	2.58	0.0334
26	3580	26.41	2.45	0.0409
27	4981	26.96	2.28	0.0323
28	2909	27.60	2.02	0.0375
29	3753	28.03	1.91	0.0312
30	1075	28.42	1.67	0.0509
31	1895	28.58	1.52	0.0349
32	831	28.81	1.36	0.0472
33	1396	28.98	1.25	0.0335
34	580	29.21	1.12	0.0465
35	847	29.31	0.95	0.0326
36	181	29.50	0.83	0.0617

Table 3. WorkKeys Reading for Information Score Statistics by ACT Reading Score

ACT Reading	N	WorkKeys Mean	WorkKeys SD	WorkKeys SE
5	31	14.26	4.61	0.8280
7	107	14.67	4.05	0.3915
8	231	15.32	3.92	0.2579
10	450	15.64	3.67	0.1730
11	2223	16.13	3.80	0.0806
12	7906	17.08	3.59	0.0404
13	8137	18.12	3.41	0.0378
14	4912	18.75	3.30	0.0471
15	5073	19.24	3.23	0.0453
16	5174	19.67	3.15	0.0438
17	5276	20.09	3.04	0.0419
18	5276	20.52	2.98	0.0410
19	10547	21.01	2.93	0.0285
20	4869	21.56	2.82	0.0404
21	4763	21.88	2.76	0.0400
22	4438	22.19	2.73	0.0410
23	4188	22.49	2.67	0.0413
24	3927	22.79	2.66	0.0424
25	7166	23.17	2.60	0.0307
26	3145	23.63	2.51	0.0448
27	2906	23.79	2.42	0.0449
28	4915	24.15	2.38	0.0339
29	2158	24.47	2.34	0.0504
30	1786	24.88	2.23	0.0528
31	1577	25.10	2.19	0.0551
32	1327	25.29	2.22	0.0609
33	1081	25.62	2.10	0.0639
34	890	25.90	2.13	0.0714
35	673	26.30	1.91	0.0736
36	577	26.62	2.24	0.0933

Figure 2. WorkKeys Applied Mathematics Score versus ACT Math Score with ±2 Standard Errors of the Mean Bands

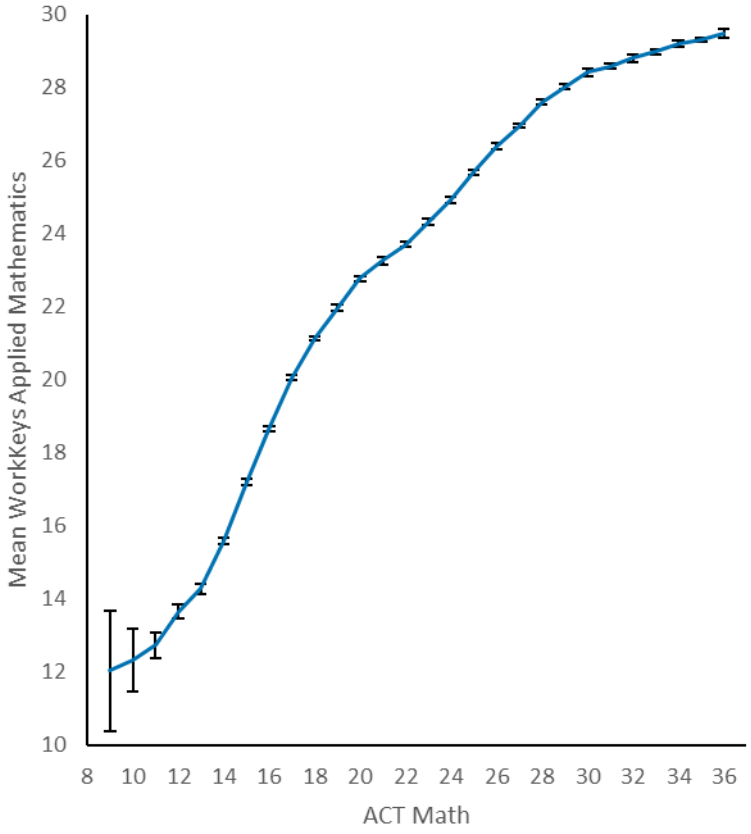
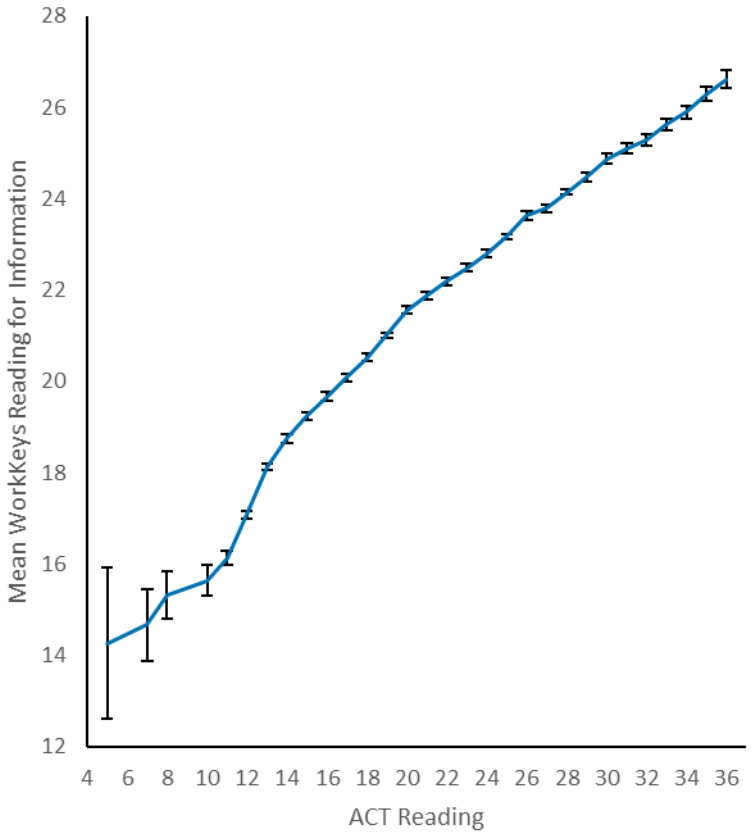


Figure 3. WorkKeys Reading for Information Score versus ACT Reading Score with ±2 Standard Errors of the Mean Bands



The mathematics regression in Figure 2 is S-shaped and extends from an ACT math score of 9 to 36. There was insufficient sample size below nine for useful results. The regression flattened out below an ACT score of 11 and above an ACT score of 30, suggesting that there was little differentiation below and above these scores. To understand this result, it is important to consider the WorkKeys Applied Mathematics conditional standard deviations at each ACT math score. As indicated by Table 2, the standard deviation dramatically decreased as ACT score increased. This was caused by a ceiling effect on the WorkKeys Applied Mathematics test. That is, students who earned an ACT math score of 30 or higher were all likely to answer nearly every item correct on the WorkKeys test. Thus, concurrent validity would be expected to suffer in this high score range. At the low end of score scale, the regression and the SE bands indicated that ACT math scores below 11 provided little information about who was likely to score higher or lower on the WorkKeys test. Of all students tested, less than 1.0% received ACT Mathematics scores less than 11. Thus, even when the ACT was taken by all high school students—not just college-bound students—nearly all students received scores in a range with evidence supporting concurrent validity.

Results for the reading tests are presented in Table 3 and Figure 3. Table 3 contains the usual sample statistics for the nonparametric regression of WorkKeys Reading for Information raw scores on ACT reading scale scores. Figure 3 shows a plot of the reading regression along with SE error bands about each mean. The regression of WorkKeys Reading for Information on ACT Reading was fairly linear except for some

curvature and flattening out at the low end of the ACT scale. The WorkKeys conditional standard deviations decreased as ACT score increases, but not as dramatically as they did for Applied Mathematics. Sample size allowed stable estimation of the regression down to an ACT score of 5. Even considering the SDs and SEs, which were larger near the bottom of the ACT score scale, the regression indicated that ACT reading scores as low as 7 provided meaningful distinctions in terms of which students were likely to perform better or worse on WorkKeys. Of all students tested, less than 1.0% received scores below 7. As in the mathematics analysis, nearly all high school students who took the ACT test earned scores in a range with evidence supporting concurrent validity.

Interpretation of Results

In this study, meaningful distinctions between ACT scores were observed down to a composite score of 12, a math score of 11, and a reading score of 7. This result is, in part, a reflection of the methods applied when developing the ACT scales. Specifically, the 1–36 ACT scale was developed to have a constant conditional standard error of measurement (CSEM) of 2 points (ACT, 2019a). In general, a standard error of measurement indicates the amount of imprecision in test scores reflecting the fact that students would not necessarily earn the same scores if testing was repeated. Since the CSEM for students who score 14 or 22 or 30 is approximately the same, their scores are similarly reliable. This point is particularly important for states that set their own performance standards when using the ACT to meet accountability testing requirements. For example, a state might have to set three performance standards (or “cut scores”) to divide scores into four performance levels (e.g., Developing, Approaching, On Track, and Exceeding). Even the score dividing the two lowest performance levels should reliably distinguish between students of lower and higher ability. This will be true if performance standards are set in a range of ACT scores with supporting validity evidence, which is the case for all states who have set standards on the math and reading tests. Typically, lower math performance standards range from 16–18, and lower reading performance standards range from 17–18 (most states set an ELA or English + Reading standard rather than reading alone).

Note that ACT scores below the minimums identified in this study are approximately in the range of scores that would be expected from random guessing on the entire test or leaving many item responses blank. A relationship between ACT composite and FYGPA would not be expected in this score range because those scores indicate little about students’ academic achievement levels and therefore little about their likelihood of academic success in postsecondary studies. Although it would be desirable for scores along the entire 1–36 scale to be meaningfully interpretable, this situation is unavoidable for a multiple-choice test on which extremely low ability or unmotivated students can guess answers at random. For example, a random guesser might answer 10 out of 40 items correct on the ACT science test and receive a score of 8 on the 1–36 scale. In that example, the 1–7 score range (approximately) would be reserved for students with even fewer items correct (e.g., by leaving many items blank).

Summary

This study investigated the predictive validity of the ACT composite score using first-year college grade point average as the criterion. It also evaluated the concurrent validity of the ACT reading and math tests using the WorkKeys Reading for Information and Applied Mathematics score as criteria. In these validity evaluations of the ACT, particular interest was given to the meaningfulness of interpreting low ACT scores. These investigations revealed evidence of predictive and concurrent validity over a wide range of the ACT score scale and that nearly all students received scores in this range—even when the ACT was administered to all high school students regardless of college plans. The results support the use of the ACT as an every-student test to assess high school students' knowledge and skills taught in the high school curriculum.

References

ACT. (2019a). *The ACT technical manual*. Iowa City, IA: ACT. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACT_Technical_Manual.pdf

ACT. (2019b). *ACT WorkKeys technical manual.* Iowa City, IA: ACT. Retrieved from https://www.act.org/content/act/en/research/reports/technical-manuals-and-fairness-reports.html

YoungWoo Cho, PhD

YoungWoo Cho is a principal psychometrician in Assessment Transformation at ACT specializing in scaling and equating, IRT, and classical test theory.

Jeffrey Steedle, PhD

Jeffrey Steedle is a director in Assessment Transformation leading the team responsible for statistical analyses for the ACT test and guiding research studies related to maintaining measurement quality while making changes to the assessment program. Jeff holds advanced degrees in education, statistics, and educational psychology, and his research interests include assessment validation and motivation on achievement tests.

David Woodruff

David Woodruff was a senior psychometrician in Measurement Research Department at ACT specializing in statistical analyses and classical test theory.

Dean Colton

Dean Colton was a principal psychometrician in Measurement Research Department at ACT specializing in scaling and equating for large scale tests.