Scale Score Interpretation Guide

In response to client need for finer-grained score reporting options for the ACT® WorkKeys® assessments, ACT created a Scale Score for the assessments. This document helps in understanding the WorkKeys Scale Score by explaining what the score is, how it can be used, and how it was developed.

Applied Math	
Scale Score	Level Score
65-71	< 3
72-75	3
76-79	4
80-82	5
83-85	6
86-90	7

Graphic Literacy	
Scale Score	Level Score
65-71	< 3
72-75	3
76-77	4
78-81	5
82-85	6
86-90	7

Workplace Documents	
Scale Score	Level Score
65-71	< 3
72-76	3
77-80	4
81-82	5
83-85	6
86-90	7

Background

WorkKeys was originally developed to be used in conjunction with a job profiling process for employee selection and promotion decisions. The score used for this purpose is the Level Score, which yields a broadband score range of 5 points across the performance range. Only Level Scores should be used for selection, promotion, or other individual high-stakes purposes that are based on WorkKeys profiles. The profiles are aligned to Levels and not to a more granular score. Further, the Level Scores, validated by profiling, have greater stability than more fine-grained scores.

Types of WorkKeys Scores

Level Scores:

Use for selection, promotion, or other individual high-stakes purposes when additional work such as a job profile has been conducted.

Scale Scores:

Use to provide finer grain score distinctions for analyzing growth over time, evaluate group comparisons on outcome measures, and provide evidence of benefit from educational and training programs.

Uses for WorkKeys Scale Scores

The rationale for developing the Scale Scores was to provide users with more detailed information for use in program evaluation and outcome measurement. Therefore, the Scale Scores make finer distinctions than can be made with the Level Score scale.

The most typical user case scenario may be when educators and trainers assess achievement by administering a pretest and posttest in selected subject areas. In order to determine improvement, these clients need a scale that is sensitive to instructions and reports subtle score changes. So an individual may score a Level 4 at both pretest and posttest, but an examination of the Scale Score could show growth within that Level Score.

The Scale Score Defined

To develop the Scale Score conversions, ACT identified a base form for each test based on an evaluation of technical qualities. ACT then applied an Item Response Theory (IRT) method combined with an arcsine transformation method to develop the Raw-to-Scale Score conversion for each base form. The score scale was set to range from 65 to 90 and is the same range as the original WorkKeys assessments. The new score scale also has an approximately equal standard errors of measurement (about 2.0 or less Scale Score points) for each test. These base form conversions will be used for future equating of new forms.

Technical Information About Scale Scores

The Data. The data used to determine the Level Scores and Scale Scores comes from the scaling study which was the second of three field studies conducted as part of the process of updating the three assessments associated with the ACT® WorkKeys® National Career Readiness Certificate® (ACT® WorkKeys® NCRC®). Fifty-one test sites were recruited to participate in the study; 40 sites actually provided test data. These sites included 13 high schools and 27 adult testing centers across 22 states in different regions of the country.

Scaling Study	
Test Takers	Percentages
Male	44%
Female	53%
Adults	40%
High School	60%
White	61%
Black/African American	18%
Hispanic	6%

Percentage of test takers for each group (Liu, Zhu, Chen, Wang, Lin, and Gao, 2017).

Over 2,500 individuals participated in the study with approximately 80% taking all three tests; all test centers had been instructed to administer all three tests to each test taker. After data cleaning, the sample sizes ranged from 1,096 to 1,196 for individual forms. Approximately 920 test takers took all three forms. To maximize the available data, analyses were done with the individual test forms (Liu, Zhu, Chen, Wang, Lin, and Gao, 2017). In order to determine the association of Scale Scores with Level Scores, Scale Scores were calculated for a large number of items from each assessment.

Standard Setting. ACT staff conducted a standard setting study for each assessment with a panel of experts consisting of educators and business people, some of whom are current WorkKeys customers. The purpose of the standard setting process is to gather data to assist ACT in establishing the standards for achieving a defined performance level on each of the NCRC assessments. The three skill assessments are criterion referenced measures. Because of this, scores on the assessment are aligned to a set of skills that a test taker has demonstrated. The goal of the standard setting process is to identify a point on the score scale where test takers who score at or above the point have demonstrated the ability to perform the skills, and test takers who score below the point have not demonstrated the ability to perform the skills.

The Mapmark with Whole Booklet Feedback standard setting method was used in this study. It is a variation of the popular Bookmark procedure. The primary difference between Mapmark and Bookmark is the Item Map. The Ordered Item Booklet (OIB) has a sample of items from the item pool ordered from easiest to hardest, but on the item map, the difficulty of an item is mapped to an actual scale value. The item map, therefore, shows "how much" more difficult one item is than another. In other words, the item map provides additional information on item difficulty.

Mapmark with Whole Booklet Feedback is a three round process. This means that the panelists set cut scores three times. In Round 1, the panelists 1) took each of the updated assessments, 2) reviewed the performance level descriptors (PLDs) for each assessment (PLDs indicate what individuals can do at each score level), 3) reviewed test items and their associated Scale Score, 4) linked test items to the PLDs, and 5) placed bookmarks in the OIB for each level. Specifically, the panelists were asked to divide the items for each WorkKeys Skill Level into two groups—those that you feel are easy enough for a minimally qualified examinee in the skill level to have mastered and those too difficult for this expectation, where mastery is defined as having a 2-in-3 chance of success (or a response probability of .67) on the item. This was done for the cut score between Below Level 3 and Level 3, Levels 3 and 4, Levels 4 and 5, Levels 5 and 6, Levels 6 and 7, and Level 7 and Above Level 7.

In Round 2, the panelists received feedback regarding their bookmark placement in terms of how it translated to a recommended Scale Score on the item map scale and how it compared to the group's median cut score.

The group was then provided with Whole Booklet Feedback. Specifically, they were provided with data showing how sixteen examinees answered each of the items on the Scaling Study Form. Data was provided for two examinees that scored at or near the Round 1 cut score for each WorkKeys Skill Level and data for a borderline examinee at each level. The purpose was to help the panelists understand what examinees at the Round 1 cut scores "can" do and consider whether this is what examinees "should" be able to do according to the Performance Level Descriptions for each WorkKeys Skill Level.

In Round 3, the panelists received feedback regarding their bookmark placement in Round 2. They were then provided with consequence or impact data. This data shows the percentage of examinees performing at or above the cut scores set for each WorkKeys Skill Level. The panelists were reminded that the WorkKeys Performance Level Descriptions should take precedence since the assessments are criterion-referenced and then they set their third bookmark.

During the final meeting, the panelists reviewed the Item Map with lines representing the Round 3 median cut scores drawn on the map. Next, they received instructions for recording the Round 3 cut scores in their Ordered Item Booklet, and reviewed a Cut Score Distribution Chart showing the distribution of panelists' Round 3 cut scores across all WorkKeys Skill Levels. Finally, the panelists discussed consequences data based on the final cut scores. The panelists final median cut scores were used to define each performance level on each of the NCRC assessments. As stated above, the three foundational skill assessments are criterion-referenced measures. Because of this, scores on the assessment are aligned to a set of skills that a test taker has demonstrated.

Additionally, the Scale Score range corresponding to each Level Score was held consistent across the forms of the test. For example, on all Workplace Documents forms, Scale Scores of 77-80 are associated with Level 4. Lastly, although a common score scale with 25 points was selected for the assessments; the Scale Score on one test does not necessarily need to convert to the same level on another test.

Reference

Liu, C., Zhu, R., Chen, H., Wang, M., Lin, H., Gao, X. (2017). WorkKeys Scaling Study. Iowa City, IA: ACT, Inc.

For more in-depth, detailed information, see the related Technical Manuals.

- WorkKeys Workplace Documents Technical Manual
- WorkKeys Graphic Literacy Technical Manual
- WorkKeys Applied Math Technical Manual