ACT® WorkKeys®

WorkKeys Graphic Literacy Technical Bulletin

Table of Contents

Graphic Literacy Technical Bulletin

Section 1: General Description of the Graphic Literacy Assessment

ACT developed the Graphic Literacy assessment to replace the Locating Information assessment and to serve as a part of the ACT WorkKeys National Career Readiness Certificate® (NCRC). In developing the Graphic Literacy assessment, ACT utilized four significant sources of information: (a) knowledge gained through 23 years of administering the Locating Information assessment, (b) knowledge gained through ACT's job profiling services that included thousands of jobs for which the Locating Information assessment was identified as relevant, (c) input and feedback from a panel of external Subject Matter Experts (SMEs) with experience in workforce development or graphics/visual literacy, and (d) professional literature published over the past 50 years analyzing how students and workers read, interpret, and use graphic information.

The Graphic Literacy assessment, like all ACT WorkKeys assessments, is designed to measure relevant skills required for success in today's job market. WorkKeys assessment data is based on skills, not scores, and has the capacity to demonstrate readiness to learn and succeed in the workplace, rather than identify achievements or deficits.

ACT organized a Design Team, composed of individuals from various areas within ACT including Test Development Content, Measurement and Research, Industrial/Organizational Psychology, and Assessment Design, to review the Locating Information assessment and develop its replacement, the Graphic Literacy assessment. The Design Team was assisted by a group of external SMEs who provided feedback and recommendations on the assessment construct, proposed test blueprint, and sample items.

After a review and discussion of the pertinent empirical and professional literature in consultation with the SMEs, the Design Team defined graphic literacy as the ability to locate, comprehend, analyze, and utilize information presented in graphic resources to complete work-related tasks and to make decisions. Graphic resources are defined broadly as visual representations of information developed to convey understanding through the use of symbols, shapes, lines, numbers, and pictures. Graphic resources

include, but are not limited to, the following: tables, graphs, charts, digital dashboards, flow charts, timelines, forms, maps, and blueprints.

ACT designed the Graphic Literacy assessment to measure foundational 21st century workplace skill of comprehending, analyzing, and evaluating graphic resources to complete tasks and make decisions in workplace settings. At the lower levels, the expectation is that test takers will be able to find and identify the correct information from simple graphic resources. At the middle levels, the expectation is that test takers will not only be able to find and identify the correct information, but they will also be able to correctly identify trends and patterns, make interpretations, comparisons, and derive reasonable conclusions based on moderately difficult graphic resources. At the higher levels, the expectation is that test takers will be able to utilize multiple data points to interpret and compare multiple trends, derive reasonable conclusions, justify these conclusions, and identify when data is being presented accurately and effectively in more difficult graphic resources.

1.1 Graphic Literacy as a Foundational Workplace Skill

ACT defines foundational workplace skills as “the skills that are essential for conveying and receiving information that is vital to work-related training and success” (ACT, 2014). Job skills are different from foundational skills. Job skills are the skills required to perform a specific job. For example, registered nurses must develop multiple skills, including giving an injection to a patient, in order to successfully fulfill their job tasks. Foundational skills are more general than job skills. They are the skills that enable a person to learn specific job skills.

Foundational skills are often referred to as basic or academic skills that may be learned through formal schooling or on one’s own. Foundational skills are often defined in terms of academic subjects, including reading, writing, mathematics, and science. These skills enable individuals to acquire job specific skills, communicate information with fellow workers, and engage in lifelong learning. Foundational skills are also described as “portable” in that, rather than being job specific, they can be applied across a wide variety of jobs and occupations.

The development of the personal computer in the late 1970s and the subsequent development of office software packages designed to improve workplace communication and productivity led to the development and use of more and more graphic representations (Few, 2012). This increase in the use of graphical representations in the workplace has been confirmed by ACT’s job profiling (ACT, 2017), and its importance has been confirmed through ACT’s interviews with outside workforce development professionals (ACT, 2016). As a result, ACT has concluded that the ability to comprehend and accurately interpret graphic resources in the workplace has become as foundationally important to worker success and learning as the ability to read written communications and solve mathematical problems.

The original assessment, Locating Information, measured examinees’ ability to locate, compare, summarize, and analyze information presented in graphical format (ACT, 2008). The assessment was developed through input and evaluation from employers, workforce development officials, and community college leaders and instructors. In many ways, the assessment was one of the first tests of workplace graphic literacy. Locating Information’s content was developed through ACT’s collaboration with individuals in the workplace. These individuals provided work-related documents and ideas about how

the information was used in the workplace. ACT used this content to develop realistic workplace graphics and questions for the assessment. Its relevancy to the workplace was confirmed through advisory panels, ACT's job profiling services, and the fact that 11 states have contracted with ACT to administer the assessment as a part of their K–12 evaluation of career readiness.

In addition to measuring all of the Locating Information assessment skills (the skills of locating, comparing, summarizing, and analyzing information presented in graphical resources), the Graphic Literacy assessment measures the ability to evaluate the accuracy and effectiveness of graphic resources. When Locating Information was designed and developed in the early 1990s, office software packages were becoming important but had not yet become ubiquitous. A few specialized graphic artists or administrative assistants understood the capability of the software packages, and they created the office graphics. Over the past 25 years, office software packages have been loaded on nearly all workplace computers and workers of varying levels of responsibility have access to use these packages. (Avgerinou, 2015; Few, 2012; Koomey, 2008). As a result, individuals with varying abilities are creating graphics to use throughout business. Many of the individuals who create these graphics do not recognize the essential power of graphics—to take complex information and present it in an easy to understand format (Few, 2012). As a result, many of the graphics circulating in offices today fail to enlighten or inform. For this reason, ACT determined that an essential component of the Graphic Literacy assessment was the evaluation of the accuracy and effectiveness of different graphic representations designed to convey information.

In evaluating the current Locating Information assessment, the Design Team concluded that the assessment's title was limiting. At the lower levels, the assessment was designed to assess a test taker's ability to find and locate information in graphics. At the higher levels, Locating Information required the test taker to interpret and analyze information, a much more difficult set of skills. To capture a more thorough description of the assessment and to be consistent with the professional literature, the Design Team concluded that the name Graphic Literacy more accurately captured the fullness of the construct.

Graphic Literacy is a subcomponent of multi-media literacy. Mayer (2009) defines multi-media learning as the presentation of both words and pictures to better facilitate learning and retention. Mayer's cognitive theory of multi-media learning assumes that an individual's information processing system includes dual channels for learning through pictorial/graphical stimulus and written/verbal stimulus. Further, each channel has a finite capacity for processing information. Active learning occurs when an individual attends to a stimulus, determines the importance and relevancy of different aspects of the stimulus, and relates the important and relevant aspects to past learning so as to create a coherent body of knowledge. Confining active learning to a single channel limits the amount of learning and comprehension that potentially may occur. Presenting material using a dual channel approach (pictures and words) provides the learner with increased opportunity to process and retain information. Although each channel has a finite capacity, integrating the information presented in both channels increases overall capacity and the likelihood of learning and retention (Mayer, 2009; Moreno & Valdez, 2007).

Few (2012) argues that graphical representation enables users to take complex quantitative information and present it in a manner that can be more easily comprehended and interpreted. He believes that graphs are tools that, when used effectively, provide people with greater access to understanding complex trends, patterns, and relationships. An improved understanding of data trends, patterns, and relationships, whether in business or education, should facilitate better decision making and increase the likelihood of success. As a result, Few (2012) perceives the development of effective graphical

presentations of quantitative information as one of the most significant learning advancements of the past 250 years.

The Graphic Literacy assessment uses words, pictorial shapes and symbols, and numbers as visual representations to communicate information and inform decision making. In ACT's definition, graphic resources are used to communicate both quantitative and qualitative information. In addition to measuring the basic skill of locating information, Graphic Literacy also measures examinees' ability to interpret and apply data, trends, patterns, and relationships. At advanced levels, it measures the examinee's ability to identify accurate and effective graphics, and requires examinees to justify their decision making. As a complete assessment, Graphic Literacy provides the test score user with information regarding an examinee's ability to use information presented in graphical format to learn, succeed, and advance in the workplace.

1.2 Graphic Literacy—Assessment Claims

To validate test score interpretations and/or decisions is to review and evaluate the plausibility of the claims made regarding the test and its scores. Kane (2013) maintained that an argument-based approach to validation requires that the score-based claims be clearly articulated along with their associated inferences and assumptions. Validation therefore is a scientific process designed to evaluate the degree to which the analytic and empirical evidence supports the assessment claims along with the inferences and assumptions required to build the connections from examinee task performance to score-based interpretations and uses.

Based on the proposed uses of the Graphic Literacy assessment, the Design Team defined three major claims focusing on workforce development concerns related to worker productivity and success at the individual level and business level. Each claim, and the underlying assumptions required for it to be plausible is outlined below.

Claim #1: U.S. examinees of high school or workforce age who demonstrate scores that reach at least a given level on the Graphic Literacy assessment are more likely to successfully perform in more and higher levels of U.S. jobs (in the ACT job taxonomy) than examinees whose scores do not reach that level.

Claim #1 Assumptions:

1. Graphic literacy is a component of foundational workplace skills, and it is required for success in a large number of jobs (based on ACT's job profile database).
2. ACT has developed a professionally valid and appropriate definition of the graphic literacy construct.
3. ACT's Graphic Literacy assessment provides reliable scores that reflect the construct. ACT's Graphic Literacy assessment elicits observable evidence of the construct.
4. ACT has defined workplace appropriate Graphic Literacy Performance Level Descriptors (PLDs), and ACT has established standards (e.g., cut points) aligned to the PLDs.
5. Cut scores used to delineate each performance level have sufficient classification accuracy.
6. Businesses and employers are able to validly measure worker performance.

7. Scores on the Graphic Literacy assessment are positively related to measures of employee performance, including productivity and turnover rates.
8. Examinees who score well on Graphic Literacy are more likely to receive higher performance ratings and are more likely to have greater job success (defined as job retention and performance evaluations) than lower scoring examinees.

Claim #2: U.S. companies that hire U.S. examinees of high school or workforce age who demonstrate scores that reach at least a given level on the Graphic Literacy assessment are more likely to achieve greater gains in productivity (for example, measured as increased output per day) than companies that hire examinees whose scores do not reach that level.

Claim #2 Assumptions:

1. Claim #1 Assumptions 1–7
2. Employees who possess higher foundational workplace skills (as defined by ACT) are more likely to be productive and effective workers (as defined by supervisor evaluations) than employees who possess lower foundational workplace skills.
3. Having more productive workers leads to a business that is more effective and productive.

Claim #3: U.S. companies that hire U.S. examinees of high school or workforce age who demonstrate Graphic Literacy scores that reach at least a given level are more likely to reduce turnover (retain those examinees for at least 6 months) than companies that hire examinees whose scores do not reach that level.

Claim #3 Assumptions:

1. Claim #1 Assumptions 1–7
2. Employees with higher foundational skill levels are less likely to be terminated in the first 6 months of employment than employees with lower foundational skill levels.
3. Employees with higher foundational skill levels are less likely to quit in the first 6 months of employment than employees with lower foundational skill levels.
4. Businesses that utilize scores from the Graphic Literacy assessment as part of their hiring process will tend to experience less turnover than businesses who do not use the Graphic Literacy assessment as part of their hiring process.

The three Graphic Literacy claims address questions around employee job performance and success, improving worker productivity, and reducing employee turnover rates. Based on the claims, the critical stakeholders and intended test users are employers and hiring managers, state or regional workforce development officials, schools that develop students to take jobs in the state or region, and examinees who are or will be seeking employment and career advancement.

Included as a part of all three claims are issues related to graphic literacy as a component of foundational workplace skills, ACT's definition of graphic literacy, ACT's measure of graphic literacy skills, and the establishment of appropriate score standards. These are critical assumptions that must be plausible to support test score interpretations. (For information and data supporting the four underlying assumptions, please refer to Sections 7 and 8.)

1.3 Test Users and Stakeholders

The stakeholders are business employers, regional workforce development offices, schools that use the assessment as a measure of workforce readiness, and states or regions committed to developing their workforce. They are the individuals and groups who are invested in finding the right people for the right jobs.

Examinees. Individuals who take the Graphic Literacy assessment are students and workers interested in demonstrating their foundational skill level in order to qualify as career ready, receive specific skill-related training, or qualify for a specific job. The examinee group includes individuals in high school through adult working lifetime. High school students take the assessment to gain an understanding of their levels of career readiness in critical skill areas and/or as a part of state accountability programs. Community college students take the assessment to demonstrate that they possess foundational skills and are ready to move forward for advanced training. College graduates take the assessment to demonstrate their level of career readiness as a means of differentiating themselves from other graduates. Working adults take the assessment to qualify for a job or to demonstrate that they have the foundational skills needed for promotion or advanced training. In short, the examinee group includes high school students and adults who are either seeking employment or looking to advance in their fields.

Stakeholders. Stakeholder groups include high schools and local school districts, state departments of education, community colleges, state and local workforce development departments, and employers.

High schools and local school districts administer the WorkKeys assessments in order to evaluate whether their curricular programs are enabling students to become career ready. In doing this, they are also providing their students the opportunity to earn a career ready certificate. State departments of education use the WorkKeys assessments as an accountability measure for evaluating the effectiveness of high schools and school districts in assisting their students to become career ready.

More specifically, the Graphic Literacy assessment provides high schools and school districts with student data regarding the extent to which students have mastered facets of the curriculum related to the interpretation of data and information presented in graphic and visual formats. The application of graphic literacy skills to workplace scenarios differentiates the Graphic Literacy assessment from other standardized assessments of students' ability to read or to understand graphics. The assessment provides the students the opportunity to demonstrate their mastery of graphic literacy along with the application of these skills to real-world problems.

Community colleges utilize the Graphic Literacy assessment in a variety of ways. Many community colleges use it as part of the process for determining acceptance into Career and Technical Education programs. Other community colleges use the assessment for program evaluation. Additionally, some community colleges use it as a means of assisting their graduates in obtaining employment.

State and local workforce development offices utilize the assessment as a means of assisting unemployed or underemployed individuals in finding employment or better opportunities. The assessment provides a means for the workforce development office personnel to better understand the skill levels of individuals and to provide better guidance and assistance to them in finding employment.

Employers may use the Graphic Literacy assessment, when coupled with a job profile analysis, to assist them in screening job applicants and finding sufficiently-qualified employees. A WorkKeys Job Profile allows the employer to understand the level of skill needed by a newly hired employee to successfully meet job expectations. Following the profile process, the employer may have job applicants take the assessment and then use their test scores as an additional piece of information to evaluate applicants.

1.4 Alignment to ACT's Holistic Framework

Building on research conducted over the last 50 years, ACT has developed its Holistic Framework (Camara, O'Connor, Mattern, & Hanson, 2015), which provides a more complete description of college and career readiness. The framework is organized into four broad domains: core academic skills, cross-cutting capabilities, behavioral skills, and education and career navigation skills.

1. Core academic skills include the domain-specific knowledge and skills necessary to perform essential tasks in the core academic content areas of English language arts, mathematics, and science.
2. Cross-cutting capabilities include the general knowledge and skills necessary to perform essential tasks across academic content areas. This includes technology and information literacy, collaborative problem solving, thinking and metacognition, and studying and learning.
3. Behavioral skills include interpersonal, self-regulatory, and task-related behaviors important for adaptation to and successful performance in education and workplace settings.
4. Education and career navigation skills include the personal characteristics, processes, and knowledge that influence individuals as they navigate their educational and career paths (e.g., make informed, personally relevant decisions; develop actionable, achievable plans).

The Holistic Framework has been both broadened and deepened to have more specific, measurable strands against which to compare job-relevant skills. For graphic literacy, the alignment is as follows:

1. Finding information in graphics
2. Translating to a different form of graphic
3. Evaluating bias in the use of a graph

ACT translated and built out these three primary learning outcomes to form the primary cognitive skills utilized in Graphic Literacy.

Section 2: Test Development

The Graphic Literacy assessment is designed to assess foundational 21st-century workplace skills that employees use to find, summarize, compare, and analyze information to make decisions using graphic resources such as, but not limited to tables, graphs, charts, digital dashboards, flow charts, timelines, forms, maps, and blueprints. Through a review of the empirical and professional literature and through deliberations among team members, the team determined that the graphic literacy construct was defined through two variables: graphic complexity and cognitive skill.

Graphic literacy tasks are successfully completed when an individual identifies the relevant information communicated by the graphic resource and cognitively applies the required skill to complete the task. Consequently, tasks presented through test items are defined by the interaction of the complexity of the graphic resource and the complexity of the cognitive process. As a result, ACT bases its definition of item level on the *interaction* between *graphic complexity* and the *cognitive process* required to successfully complete the task elicited by the question.

2.1 Graphic Complexity

Graphic resources are visual representations of information designed to convey understanding to a potential user. Graphical materials may use words, but they also use symbols, shapes, lines, numbers, and pictures to enable understanding in ways that are more effective than using words alone. Graphic resources include all of the specified graphic types identified in Section 1, as well as other visual resources designed to communicate information, such as directions, data readings, data trends, variable relationships, and summaries.

Diagrams, maps, graphs, and other visuals have been used to communicate and convey information from the earliest times (Wainer, 1992). Cave drawings discovered in the south of France are the earliest known maps designed to depict geographical locations and human actions (Wolodtschenko & Forner, 2007). In the 17th century, René Descartes applied algebraic thinking to geometric problems using two- and three-dimensional graphs to illustrate mathematical solutions. In the 18th century, Scottish social scientist William Playfair applied the concept of the Cartesian coordinate graph to display and explain quantitative data about human phenomenon (Few, 2012).

All visuals and graphics are not created equal (Rogers & Scaife, 1998). Graphical representations range from the straightforward and easy to understand to extremely complex and specialized. Line graphs created by Playfair depicting the interest on the British national debt are relatively easy to interpret, while diagrams derived from the human genome project require years of training to understand. Although none of the graphic resources used in the Graphic Literacy assessment require specialized training, WorkKeys graphic resources do vary in terms of type, density of information, and presentation. The Design Team refers to these variations as graphic complexity. With the assistance of the external SMEs, the Design Team developed a table describing four categories of graphic complexity. Table 2.1 presents a description of the characteristics for each category.

Table 2.1: Characteristics of Simple, Low Moderate, High Moderate, and Difficult Graphics

Stimulus Characteristics	Simple	Low Moderate	High Moderate	Difficult
Number of Axes	One or two axes	One or two axes	One or two axes	One, two, or more axes
Levels of Data	One level of data	More than one level of data; no nesting	More than one level of data; nesting allowed	More than one level of data; nesting allowed
Number of Variables	Few variables (1 to 2)	Several variables (3 to 5)	Many variables	Many variables
Number of Representations of Data	No more than 20 data points/ fields	Moderate number of data points/ fields	Moderate number of data points/ fields	Densely presented data
Familiarity of Graphic Type	Common graphic types	Common graphic types	Less common graphic types	Less common graphic types, including composite graphics
Total Number of Graphics	One	One or two	May be one, two or more	May be one, two, or more

The characteristics of each category are based on a combination of the levels of data, number of variables, number of representations of data, familiarity with the graphic type, and the total number of graphics. Although classifying the complexity of a graphic resource into one of the four categories is somewhat subjective, using the defined characteristics permits the content development team to classify the graphic resources with a great deal of consistency.

Graphic Classification Consistency Study: ACT conducted two studies to evaluate the consistency of classification of different graphic resources into the four categories applying the principles described in Table 2.1.

Study No. 1: The first study asked four content specialists with substantial experience developing the Locating Information assessment to discuss how they classify graphics and to consider the merits of using a table similar to Table 2.1. Following the discussion, the four content specialists independently evaluated 31 graphics and classified them into the four categories. The 31 graphic resources represented a variety of graphic types including tables, bar charts, line graphs, forms, maps, flow charts, and multiple graphics.

ACT utilized Generalizability Theory (Brennan, 2001) to analyze the consistency of the content specialists' ratings. A graphics x rater design was modeled and used the GENOVA software program (Crick & Brennan, 2001) to analyze the ratings. The analysis provided a Generalizability Coefficient of 0.81, and a Phi Coefficient of 0.80. These consistency indices revealed that the four content specialists, using a table similar to Table 2.1 and their training, classified graphic resources in a relatively consistent manner.

Although relatively good consistency was demonstrated through the study, the content team believed that they could become more consistent. Through a series of meetings, they further refined the definitions and means of classification. The result was the development of Table 2.1, along with a stronger understanding of the need for consistent classification.

Study No. 2: The content team wanted to verify the gains achieved through their additional work in defining the four categories, and consequently the second study was organized. The same four content specialists along with two additional content specialists were asked to classify 25 different graphic resources using the recently refined and developed materials.

The second analysis provided a Generalizability Coefficient of 0.91 and a Phi Coefficient of 0.87. With these results, the Graphic Literacy content team concluded that, through the exercises, discussion, and refinement of classification criteria, the team had achieved a high level of graphic resource classification consistency.

2.2 Graphic Literacy—Cognitive Process Definitions

The cognitive processes required to solve a problem using graphical information also vary. The Design Team divided the Graphic Literacy assessment construct into four cognitive skills: Locate Information, Assess Trends/Patterns/Relationships, Make Inferences or Decisions, and Selecting the Graphic to Represent Information. Each skill was then divided into subskills. A total of 13 subskills define the cognitive processes used by examinees to solve the graphic literacy problems. Table 2.2 presents the Graphic Literacy assessment domain define through four skills with each skill divided into subskills.

Table 2.2: Graphic Literacy Cognitive Skills and Subskills

	Skills/Subskills
Find	Locate information
1.F.1	Locate information
1.F.2	Identify the next or missing step in an illustrated process
2.F.1	Compare two or more pieces of information
2.F.2	Locate information in a graphic using information found in another graphic
Trends	Assess trends/patterns/relationships
2.T.1	Identify a trend/pattern/relationship
3.T.1	Interpret a trend/pattern/relationship
3.T.2	Compare two or more trends/patterns/relationships
Decisions	Make inferences or decisions
2.D.1	Make an inference or decision
3.D.1	Make a reasonable inference or decision based on one graphic after finding information in another graphic
3.D.2	Justify an inference or decision based on information
Represent	Selecting the graphic to represent information
2.R.1	Identify the graphic that represents the data
3.R.1	Identify the most effective graphic given a defined purpose
3.R.2	Justify the most effective graphic given a defined purpose

An additional means of defining the cognitive processes used to solve problems or complete tasks using graphical information is based on the number of cognitive steps performed (Curcio, 1987; Friel, Curcio, & Bright, 2001; Wainer, 1992). Based on the item and the associated task, a test question might prompt the examinee to perform one cognitive step. For example, the item task could prompt the examinee to find a piece of information located within a table. In this case, the examinee is extracting the needed information from the graphic and is solving the task using a one-step cognitive process.

The task associated with a second item might prompt the examinee to locate a piece of information and then use the information in a second cognitive step. For example, the item task could prompt the examinee to locate two pieces of data within a table and then decide whether the data indicates that the standard is met. For this item, the task requires the examinee to extract the data (one cognitive step) and then process the information through a second step to reach a reasonable decision. The examinee is solving the task through a two-step cognitive process.

The task associated with a third item might prompt the examinee to extract information from the graphic and then use the information in a multi-step process to derive the solution. For this item, the task requires the examinee to locate the information (one cognitive step), process that information in a

specific manner (e.g., compare information, interpret information) and then find the final solution (e.g., determine if the bar graph is the best representation of the sales data). The examinee is solving the task through a three-step cognitive process.

Graphic Literacy—Skill Definitions

For the Graphic Literacy assessment, the Design Team defined three levels of skills based on the number of cognitive steps that must be performed to complete the task (Curcio, 1987; Friel, Curcio, & Bright, 2001; Shah & Freedman, 2011; Wainer, 1992).

One Step—Extracting Data: skills involve locating or filling in data in a graphic with no additional cognitive steps

Two Step—Read between the data: skills involve using one cognitive process after extracting the relevant data from the graphic

Three Step—Read beyond the data: skills involve using two or more cognitive processes beyond extracting the relevant data

For example, a manager needed to determine whether Group A or Group B had more sales in a given month using data presented in a bar graph. First, the manager would compare the length of each bar. Then, he or she would identify which bar was highest, and thus infer that Group A had made more sales. Finally, the manager might consider if the bar graph is the best possible representation of the data and if a chart that also included the profitability of both groups might be more effective.

The cognitive skills and subskills (Table 2.2) were integrated with the three-step cognitive skill model to derive the graphic literacy cognitive process model. Table 2.3 presents the Graphic Literacy Cognitive Process Model.

Table 2.3: Graphic Literacy—Cognitive Skill Model

	Cognitive Steps		
Cognitive skills	One-step cognitive process: extract	Two-step cognitive process: between	Three-step cognitive process: beyond
Locate information	• 1.F.1 Locate information • 1.F.2 Identify the next or missing step in an illustrated process	• 2.F.1 Compare two or more pieces of information • 2.F.2 Locate information in a graphic using information found in another graphic	
Assess trends/patterns/ relationships		• 2.T.1 Identify a trend/ pattern/ relationship	• 3.T.1 Interpret a trend/pattern/ relationship • 3.T.2 Compare two or more trends/patterns/ relationships
Make inferences or decisions		• 2.D.1 Make an inference or decision	• 3.D.1 Make a reasonable inference or decision based on one graphic after finding information in another graphic • 3.D.2 Justify an inference or decision based on information
Select the graphic to represent information		• 2.R.1 Identify the graphic that represents the data	• 3.R.1 Identify the most effective graphic given a defined purpose • 3.R.2 Justify the most effective graphic given a defined purpose

2.3 Graphic Literacy—Score-Level Definitions

Examinees may score at five different proficiency levels on the Graphic Literacy assessment—Level 3 to Level 7. (Examinees who demonstrate graphic literacy proficiency below a Level 3 do not receive a level score.) Graphic Literacy scores or performance levels are determined by the interaction of the graphic complexity categories with the cognitive skill processes. The Graphic Literacy performance levels defined through the interaction are presented in Table 2.4.

Table 2.4: Graphic Literacy Score or Performance Levels

	Graphic complexity			
Cognitive skill levels	Simple	Low moderate	High moderate	Difficult
1-Step	Score Level 3	Score Level 3	Score Level 4	Score Level 5
2-Step	Not Tested	Score Level 4	Score Level 5	Score Level 6
3-Step	Not Tested	Score Level 5	Score Level 6	Score Level 7

The interaction of the two facets, graphic complexity and cognitive skill, provides the overall performance level of the Graphic Literacy item. Performance levels are based on the concept that it is more difficult to apply the same skill to a graphic resource of higher complexity. For example, if a one-step process is applied to a graphic of low-moderate complexity, the performance level overall is defined as Level 3. However, if that same one-step process is applied to a graphic of high moderate complexity, the overall performance level is then defined as Level 4. In effect, when the same cognitive skill is applied to a more complex graphic, the task elicited by the item is at a higher performance level.

Likewise, when the task elicited by an item requires an examinee to apply a more difficult cognitive skill to a similarly complex graphic, the result is an increase in the performance level. As a result, an individual applying a two-step cognitive process to a graphic of high moderate complexity results in an overall performance Level 5. To further illustrate, if a three-step cognitive process is applied to the same high moderate complexity graphic, the performance level increases to Level 6.

2.4 Graphic Literacy—Performance Level Descriptors (PLDs)

The Graphic Literacy construct is defined through Tables 2.1, 2.2, 2.3, and 2.4. The tables collectively provide direction for item writers to develop items with tasks that elicit the skills aligned to each of the performance levels. By integrating this information, the Design Team defined the Graphic Literacy PLDs.

Examinees scoring at Level 3 have demonstrated the following abilities:

- Locate and find information or identify the next step in a simple graphic
- Locate and find information or identify the next step in a low moderate graphic

Examinees scoring at Level 4 have demonstrated all of the skills defined at Level 3 and have demonstrated the ability to find information or identify the next or missing step in a high moderate graphic. In addition, they have also demonstrated the following skills with low moderate graphics:

- Locate information in a graphic using information found in another graphic
- Compare two or more pieces of information
- Identify a trend/pattern/relationship
- Make an inference or decision
- Identify the graphic that accurately represents the data

Examinees scoring at Level 5 have demonstrated all of the skills defined at Levels 3 and 4 and have demonstrated the ability to locate and find information or identify the next or missing step in a difficult graphic. In addition, they have also demonstrated the following skills with a high moderate graphic:

- Locate information in a graphic using information found in another graphic
- Compare two or more pieces of information
- Identify a trend/pattern/relationship
- Make an inference or decision
- Identify the graphic that accurately represents the data

In addition, they have demonstrated the following skills with a low moderate graphic:

- Compare two or more trends/patterns/relationships
- Interpret a trend/pattern/relationship
- Make a reasonable inference or decision based on one graphic after finding information in another graphic
- Justify an inference or decision based on information
- Identify the most effective graphic given a defined purpose
- Justify the most effective graphic given a defined purpose

Examinees scoring at Level 6 have demonstrated all of the skills defined at Levels 3, 4, and 5 and have demonstrated the following additional skills with a difficult graphic:

- Locate information in a graphic using information found in another graphic
- Compare two or more pieces of information
- Identify a trend/pattern/relationship
- Make an inference or decision
- Identify the graphic that accurately represents the data

In addition, they have demonstrated the following skills with a high moderate graphic:

- Compare two or more trends/patterns/relationships
- Interpret a trend/pattern/relationship
- Make a reasonable inference or decision based on one graphic after finding information in another graphic
- Justify an inference or decision based on information
- Identify the most effective graphic given a defined purpose
- Justify the most effective graphic given a defined purpose

Examinees scoring at Level 7 have demonstrated all of the skills defined at Levels 3, 4, 5, and 6 and have also demonstrated the following additional skills with a difficult graphic:

- Compare two or more trends/patterns/relationships
- Interpret a trend/pattern/relationship

- Make a reasonable inference or decision based on one graphic after finding information in another graphic
- Justify an inference or decision based on information
- Identify the most effective graphic given a defined purpose
- Justify the most effective graphic given a defined purpose

2.5 Designing Items to Elicit Examinee Evidence of Graphic Literacy

Graphic Literacy uses multiple-choice items to measure examinees' proficiency in comprehending, analyzing, and evaluating graphic resources to complete tasks and make decisions in workplace settings. The graphic literacy skills measured by the assessment were defined by the Design Team and confirmed by external SMEs with backgrounds in business, industry, and education. To properly elicit evidence of the skills in the Graphic Literacy domain, ACT follows an item-design model aligned with both evidence-centered assessment design (Mislevy, Steinberg, & Almond, 1999) and the *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), & National Council for Measurement in Education (NCME), 2014).

Item Writing: Item writers qualify to write for the Graphic Literacy assessment by completing item-writing training modules. The modules cover numerous aspects of developing quality multiple-choice items including creating text that elicits evidence of the skill the item measures, writing effective distractors, employing realistic workplace contexts, and avoiding common item-writing errors. For Graphic Literacy, the training also provides explicit direction in terms of acceptable workplace graphical resources. Once an item writer has successfully completed all required training modules, he or she is given an item-writing assignment that details the number of items to be developed at specific levels and for specific subskills. Once an item writer has demonstrated the ability to write items, they receive materials explaining item task models.

The task models provide item writers with the following information: (a) skill name, (b) skill description, (c) evidence statement, (d) item components, and (e) item exemplars. Additional requirements related to the items include:

- All items are linked to a stimulus
- Stimulus materials are graphic resources designed to communicate information related to workplace phenomenon
- Stimulus materials may contain one graphic or multiple associated graphics
- Stimulus materials should use pictures, arrows, diagrams, or other visual representations to communicate information, and they should use as few of words as possible
- Lower-level stimuli will not include scientific terminology; for upper-level stimuli, scientific terminology is acceptable
- Multiple items will be developed for each stimulus

In the development of the task models, questions arose related to whether ancillary skills that may be required to respond to an item were construct relevant. More specifically, three questions were identified related to the construct relevance of ancillary skills:

1. In evaluating graphic effectiveness, is the identification of biased representations construct relevant?
2. Is the application of proportional reasoning skills construct relevant?
3. Is the application of mathematics skills construct relevant?

The Design Team asked the external SMEs to provide their thoughts on these questions as it related to the construct and the use of graphic literacy in the workplace.

1. **Evaluating bias in graphic presentation**: Wainer (1992) presents several interesting examples of how a graphic developer might present quantitative information to bias the user's interpretations and conclusions. To demonstrate that the problem is more common than expected, he used examples from publications such as *Forbes.* Although graphic developers may manipulate a graphic presentation to unfairly present information, in the normal workplace this type of usage is either extremely rare or non-existent. As a result, the Design Team concluded that, although the identification of bias in graphic presentation is construct relevant, for the workplace it has limited applications. The final conclusion was that such items are acceptable, but the content team should not specifically focus on or encourage their development.
2. **Application of proportional reasoning skills:** The external SMEs believed that proportional reasoning is used in nearly all applications of graphic literacy. When a worker examines a bar graph, whether intentionally or unintentionally, the individual is comparing the heights of the different bars and drawing conclusions on how one bar relates to a second bar. When a worker studies a flow chart, he or she is identifying the tasks that come early in the process and the ones that come later. Because size and shape are fundamental to the interpretation and use of graphics, proportional reasoning skills are ubiquitous and an inherent part of graphic literacy. Thus, questions asking examinees to compare the size of one part of a graph to a second part to make conclusions about whether something is twice as large (or ¼ the size) are construct relevant.
3. **Application of mathematics skills:** The question of the use of mathematics skills in the graphic literacy assessment was the most difficult question. WorkKeys is extremely sensitive to this question due to the fact the program also includes an applied mathematics assessment. The Graphic Literacy assessment is a measure of an examinee's ability to find information and solutions applying his or her graphic literacy skills, thus performance should not depend on the examinee's mathematics ability. With that understanding, ACT recognizes that to fully comprehend a majority of graphs requires both basic reading skills and basic understanding of quantitative reasoning and mathematics. Few (2012) maintains that one of the primary purposes of graphs is to display quantitative information in an easy to understand format. As a result, as one external SME commented, "it is difficult to completely separate out mathematics from graphic literacy."

With that understanding, WorkKeys developed a set of guidelines defining the limited extent to which mathematics skills may be included to answer the Graphic Literacy items.

Item Review Process: After items have been developed, edited, and tentatively finalized by the Content Assessment team, they are submitted to external consultants with experience in creating and using workplace graphic resources.

In the content review, they evaluate

- the content, including concerns about whether the item is appropriately aligned to the construct
- whether the context and the solution method are workplace relevant
- whether there is one and only one correct response

Reviewers also evaluate items and graphic resources on the basis of fairness and cultural bias. The reviewer is asked to evaluate the item and stimulus in terms of how members of different demographic groups would respond to them. (ACT asks the item reviewer to evaluate the item from the perspective of men and women examinees, and from the perspective of African-American, Hispanic-American, and Asian-American examinees.) The reviewer is asked to comment on whether there is anything within the item that any group might find offensive. Also, the reviewer is asked to evaluate if each demographic group has equal access to, and opportunity to learn, the information and skills assessed.

Item reviewers include representation from various facets of our multicultural society. Reviewers are recruited to achieve a balance of gender and a wide representation of ethnicity, geographic region, and urbanity. All test reviewers are recruited in part for their alertness to cultural diversity factors and for their sensitivity to issues of cultural diversity and fairness. Reviewers' performance is regularly evaluated by ACT staff.

For both the content and fairness reviews, item reviewers complete a questionnaire either approving the item as written or identifying specific concerns. The content team gathers the information from the reviewers and determines how to appropriately address any concerns. Items are not classified as ready for pretesting until all relevant issues are resolved.

2.6 Item Pretesting

All Graphic Literacy items are pretested before they become operational. Newly developed or recently revised items are embedded in current forms of the Graphic Literacy assessment. As a result, examinees respond to the pretest items as a part of their responses to the operational assessment.

ACT conducts statistical analyses to determine if each pretest item meets required statistical criteria. ACT analyzes the items using both classical and item response theory (IRT) statistics to evaluate their psychometric properties, including item difficulty and discrimination. If the pretest item meets the statistical criteria, it has passed pretesting. If it fails to meet the criteria, the Graphic Literacy content team reviews it and considers whether it should be edited, modified, or removed from the pool. When an item is edited or modified, it receives a new item identifier and is pretested a second time.

To ensure item fairness, ACT compares item difficulty values based on group membership (item analysis is conducted comparing difficulty levels by gender and ethnic status) and performs Differential Item Functioning (DIF) evaluations (Holland & Wainer, 1993). Items that are flagged through the DIF evaluations are sent to the Graphic Literacy content team for review. The content team determines whether the flagged item should remain as it currently is, be revised and returned to pretesting, or be removed from the pool.

Section 3: Test Specifications

Test specifications must be carefully defined to ensure that the assessment tasks are construct relevant and representative of the domain purported to be measured (Messick, 1989; Mislevy, Steinberg, & Almond, 1999). In the context of Graphic Literacy, construct relevance requires not only that the examinee demonstrate the ability to comprehend and interpret workplace graphics, but that he or she also demonstrates the ability to apply the information conveyed by the graphic to complete a job task. Because WorkKeys assessments are designed to measure skills that are widely applicable to a large number of jobs, construct representativeness refers to a range of graphic materials and the various graphic skills required in the workplace. To illustrate, graphic materials must represent the full range of job sectors, from manufacturing to construction to office work and beyond. The graphics must also represent appropriate ranges of difficulty, from straightforward frequently used graphics, to more complicated and nuanced graphics, to the newly emerging graphics.

The Graphic Literacy specifications were created by first developing the assessment's claims and score interpretations, followed by articulating the set of behaviors that need to be elicited through the test content to provide evidence to support the claims. In articulating the set of behaviors, the team evaluated the degree to which examinee responses to the item content provided support for the assessment's claims and score interpretations. Item and test content must elicit examinee behaviors that are aligned to the Graphic Literacy construct and provide evidence supporting score interpretations (Kane, 2013; Messick, 1989).

The Graphic Literacy Design Team utilized a variety of reputable source materials to identify relevant content that would constitute graphic literacy. Over the past 25 years, through its job profiling services, ACT has gathered information related to workplace graphical materials, tasks, and skills from the manufacturing, health care, construction, transportation, financial, and sales sectors. The Graphic Literacy team reviewed these findings and used the information to determine what types of graphical materials should be included and which skills were most frequently required. To further support content-related decisions, the team reviewed professional literature around workplace graphic literacy (Binkley, Erstad, Herman, Ripley, Miller-Ricci, & Rumble, 2012; Brumberger, 2011; Few, 2012) and workplace competency models (National Network of Business and Industry Associations (NNBIA), 2014). Lastly, the team consulted with a group of external SMEs to obtain their perspective on workplace graphics and related skills.

Based on the findings from the review of these resources, ACT formulated the Graphic Literacy test specifications. Using the findings in conjunction with the assessment's purpose, claims, and score interpretations, the team defined the critical content facets and weighted the skills based on their importance and frequency.

3.1 Graphic Literacy—Test Blueprint

ACT developed detailed blueprints defining the content attributes of each test item. The content specifications were developed by clearly specifying the complexity attributes of a graphic for each of four levels (See Section 2). They were further defined by specifying the workplace graphic literacy skill and subskill. Within the test specifications table, each subskill was evaluated and aligned to a level. Following

the alignment of subskills, weights were determined based on the overall importance of the subskill to the construct of graphic literacy (Allen & Yen, 2002).

The Graphic Literacy construct was based on three critical facets:

- Graphic Complexity Category of the Stimulus
- Graphic Skill elicited by the item
- Interaction of the Graphic Complexity of the Stimulus with the Graphic Skill of the Item

The Graphic Complexity Category was defined by the stimuli's number of variables, data levels, number of axes, graphic type, and the total number of graphics (see Table 2.1). ACT content specialists evaluated each stimulus and, based on these characteristics, determined its category.

Graphic Literacy skills were divided into four primary skills: locate information, assess trends/patterns/relationships, make inferences or decisions, and select the graphic to represent information. These four skills were defined based on an analysis of the professional literature on graphic literacy (Curcio, 1987; Friel, Curcio, & Bright, 2001; Shah & Freedman, 2011; Wainer, 1992).

The team divided each of the skills into separate subskills that further refined the graphic literacy domain. Using data from job profiling along with feedback from the SMEs, the team weighted the skills and subskills based on their importance to the construct of Graphic Literacy and on the frequency of use in the workplace.

Tables 3.1, 3.2, and 3.3 present the Graphic Literacy test specifications. The content specifications provide a blueprint for form development and also represent the relative importance of the graphic literacy skills and subskills in the workplace.

Table 3.1: Interaction of Graphic Complexity Level with Cognitive Skill Levels with the Overall Graphic Level Definitions

	Graphic Complexity			
Cognitive Skill Levels	Simple	Low Moderate	High Moderate	Difficult
1-Step Cognitive Tasks	Level 3	Level 3	Level 4	Level 5
2-Step Cognitive Tasks	Not Tested	Level 4	Level 5	Level 6
3- or more Step Cognitive Tasks	Not Tested	Level 5	Level 6	Level 7

Table 3.2: Number of Items by Graphic Complexity and Overall Graphic Literacy Level

	Overall Graphic Literacy Level					
Graphic Complexity Categories	Level 3	Level 4	Level 5	Level 6	Level 7	Total
Simple	3	0	0	0	0	3
Low Moderate	1	5–6	1–2	0	0	7–9
High Moderate	0	1	6–7	1–2	0	7–9
Difficult	0	0	0–1	6–7	5	11–13
Total	4	6–7	8–9	7–9	5	32

Table 3.3: Graphic Literacy Skill Distribution by Level

Skill Domain	Level 3	Level 4	Level 5	Level 6	Level 7	Total
Locate Information	4	1–4	1–3	1–2	0	10–13
Assess Trends, Patterns, and Relationships	0	1–3	1–3	1–3	1–2	6–11
Make Inferences or Decisions	0	1–3	1–3	1–3	1–2	6–11
Select the Graphic to Represent Information	0	0–1	0–1	1–2	1–2	3–4
Total	4	6–7	8–9	7–9	5	32

Table 3.4: Cognitive Skill—Number of Items by Graphic Complexity

			Cognitive Skill			
Graphic Complexity Categories	Total Graphic Sets	Items per Graphic	One-Step Extract Items	Two-Step Between Items	Three-step Beyond Items	Total Items by Graphic Complexity
Simple	3	1	3	0	0	3
Low Moderate	4	2	1	5–6	1–2	8
High Moderate	3	3	1	6–7	1–2	9
Difficult	4	3	0–1	6–7	5	12
Total	14	NA	5–6	18–20	7–8	32

Each form of the Graphic Literacy assessment conforms to the test specifications defined in Tables 3.1, 3.2, 3.3, and 3.4. ACT's test development and psychometric staff members thoroughly review each form to ensure that it meets the specifications and that each form' content is parallel to all other Graphic Literacy forms.

Section 4: Test Administration

The *ACT WorkKeys Administration Manual—Paper Testing* and *ACT WorkKeys Administration Manual—Online Testing* provide direction in the administration of the WorkKeys assessments including timing instructions. It is important that all staff involved in the administration of WorkKeys assessments follow the instructions as provided by ACT to appropriately measure the skills and abilities of the individuals completing the assessments.

4.1 Standard Test Administrative Procedures

Included in the two manuals are detailed directions for securing materials and administering the assessments in a standardized manner. The following actions violate ACT policies and procedures for delivering WorkKeys assessments:

- accessing or obtaining a test booklet or test questions prior to the test for any reason (An exception is provided for American Sign Language and Signing Exact English interpreters assisting examinees)
- photocopying, making an electronic copy, or keeping a personal copy of the test or of any test items
- taking notes about test questions or any paraphrase of test questions to aid in preparing examinees for testing
- aiding or assisting an examinee with a response or answer to a secure test item, including providing formulas
- rephrasing test questions for examinees
- creating an answer key or “crib sheet” of answers to test questions
- editing or changing examinee answers after completion of the test, with or without the examinee’s permission
- allowing examinees to test in an unsupervised setting
- leaving test materials in an unsecured place or unattended
- failing to properly report and document incidents of prohibited behavior involving examinees, staff, or others
- allowing examinees to test longer than the permitted time
- failing to return and account for all testing materials after the testing session has ended

4.2 Test Administration Personnel

ACT identifies a Test Center Coordinator to select and reserve appropriate rooms required for testing, select and train qualified personnel to administer the assessments, order and ensure the security of all test materials, and coordinate all activities on test days. The Test Center Coordinator should have a back-up coordinator who is responsible for carrying out activities should the Test Coordinator not be able to do so.

ACT provides copies of the administrative manuals, which every staff member is expected to read, understand, and follow. The manual is proprietary information and is copyrighted by ACT. It is to be used only for the purpose of administering the ACT WorkKeys assessments and is not to be copied or shared for any other purpose.

The Test Center Coordinator trains all staff members in administrative and security procedures. Each testing staff member is to be provided with a complete copy of the administrative manual before the training session. The Test Center Coordinator works with the testing staff to ensure that the WorkKeys assessments are administered in a standardized manner in rooms that meet WorkKeys requirements.

Section 5: Accessibility

The Graphic Literacy assessment uses a variety of levels of accessibility supports including default embedded tools, open access tools, and full accommodations to allow all examinees, including those with disabilities, to participate in testing.

5.1 Assessment Support System

The Graphic Literacy assessment contains a system of supports for effective communication that spans a continuum from the most simple, common accessibility tools used by everyone, to the most intensive accessibility supports that require the user to have specific qualifications and expertise. The levels of this continuum provide for an assessment system that meets the needs of all populations tested and provides a fair communication and performance pathway for all learners.

“Accessibility is the degree to which the items or tasks on a test enable as many test takers as possible to demonstrate their standing on the target construct without being impeded by characteristics of the item that are irrelevant to the construct being measured.” (AERA et al., 2014, p. 215). The Graphic Literacy assessment support continuum is an inclusive concept that recognizes that the need for personalized communication supports is not restricted to any one group of examinees. It describes needs we all have, regardless of whether or not we have an official diagnostic label. It encompasses the needs of the entire testing population, including those with disabilities, those who are English Learners, as well as all the rest who have no diagnostic label at all. All of these individuals have a shared need to be able to fairly and effectively communicate what they know and can do when they take a test.

To provide a fair performance pathway for all learners, including populations with diverse needs, the development of the Graphic Literacy assessment followed a theory of action known as Access by Design (Fedorchak, 2013), which incorporates elements of Universal Design for Learning (UDL) described by

the Center for Applied Special Technologies (CAST, 2011), and Evidence Centered Design (Mislevy, Almond, & Lukas, 2004; Mislevy & Haertel, 2006).

In September 2015, in anticipation of the development of this assessment, a week-long workshop related to accessible test development was held with leadership and content developers of WorkKeys NCRC assessments. The topic of the workshop focused on methods of mapping the characteristics and accessibility needs of learner populations to the content models intended to be measured by the NCRC assessments. During this training, accessibility consultants provided feedback with respect to accessible definitions of constructs to be tested and a plan was established for ongoing accessibility consultation and advisement during test development. In later production follow-up, ACT's Test Services developed an accessible color palette and conducted an external user review of graphic rendering.

The Graphic Literacy assessment accessibility supports are structured along a continuum of increasingly intensive supports designed to meet the needs of all participating learner populations. Three levels of accessibility supports are offered: (a) Embedded Tools, (b) Open Access Tools, and (c) Accommodations. Embedded tools are commonly used by many people, available to all examinees, and do not need to be requested in advance. Open Access Tools are used by fewer people, are also available to anyone, but their use must be identified and planned for locally in advance. Accommodation-level supports and tools are the most intensive levels of support. Accommodations are available to those who are qualified to use them. Currently, certain supports are only available with the paper form of the test.

Beginning in 2018, several new accessibility supports will be added to the Graphic Literacy assessment for both paper and online forms. These additions will fill out the planned continuum of accessibility supports and will provide many options for unique personalization of experience for each examinee.

5.2 Test Administration and Accessibility Levels of Support

Over the last decade, the educational assessment profession has come to understand that all examinees have tools they need and use every day to engage in the classroom and to communicate effectively what they have learned and can do. There are different levels of support that examinees may need in order to demonstrate what they know and can do on academic tests. The Graphic Literacy assessment makes several possible levels of support available. All these levels of support taken together are called accessibility supports. These accessibility supports:

- allow all examinees to gain access to effective means of communication that in turn allow them to demonstrate what they know without providing an advantage over other examinees
- enable effective and appropriate engagement, interaction, and communication of examinee knowledge and skills
- honor and measure academic content as the test developers originally intended
- remove unnecessary barriers to examinees demonstrating the content, knowledge, and skills being measured on the Graphic Literacy assessment

In short, accessibility supports do nothing for the examinee academically that he or she should be doing independently; they just make interaction and communication possible and fair for each examinee.

The Graphic Literacy assessment accessibility system defines four levels of support that range from minor support (default embedded system tools) to extreme support (modifications). Figure 5.1 shows the architectural structure of WorkKeys assessments accessibility supports. The width of the triangle shows the proportionate number of students who use that set of accessibility tools.

The Graphic Literacy assessment permits the use of only those accessibility supports that validly preserve the skills and knowledge that the assessment claims to measure, while removing needless, construct-irrelevant barriers to examinee performance. The four levels of support in the assessment accessibility system represent a continuum of supports, from least intensive to most intensive, and assumes all users have communication needs that fall somewhere on this continuum. The continuum of supports permitted in the Graphic Literacy assessment results in every examinee having a personalized performance opportunity.

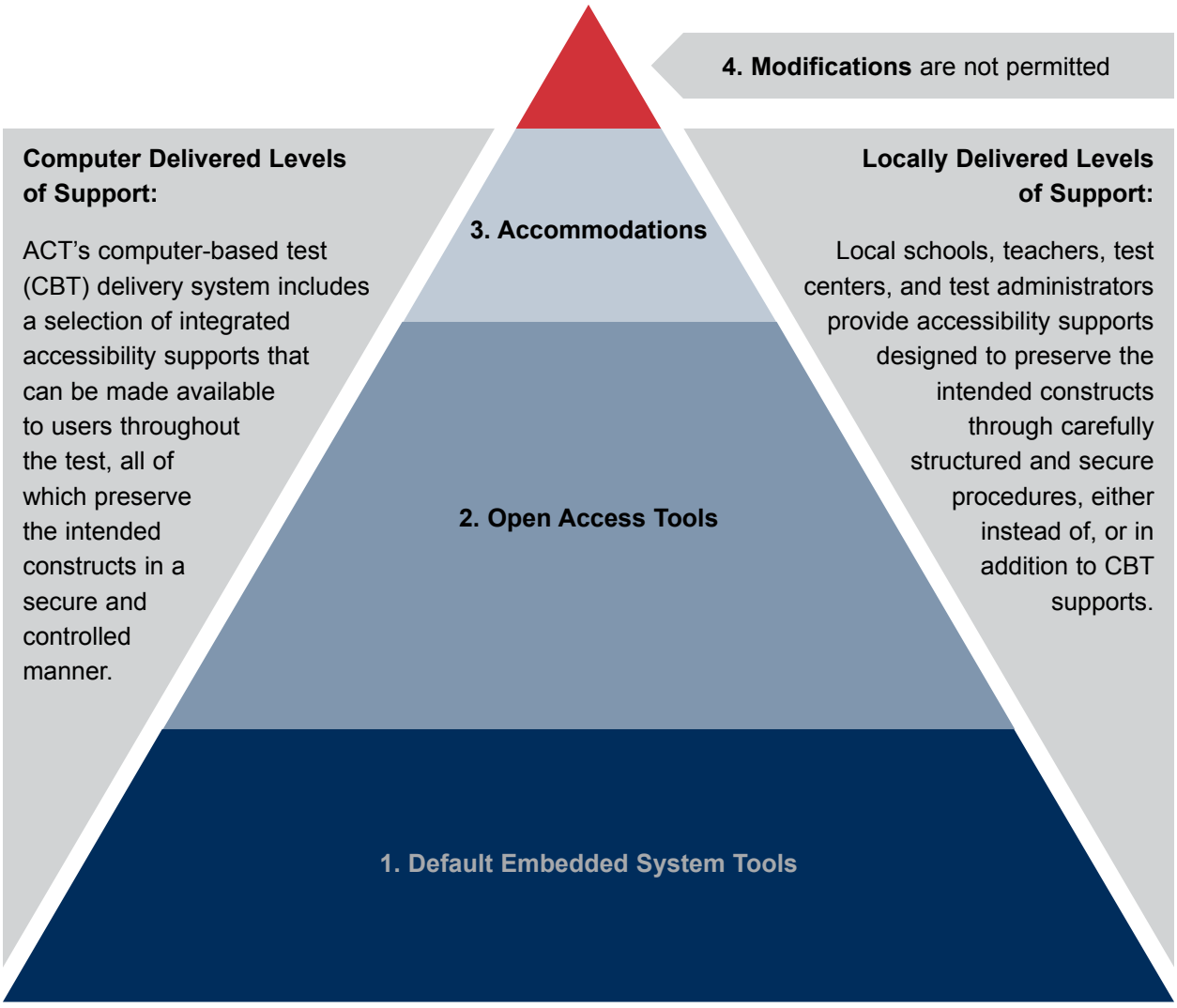

Figure 5.1 Architectural Structure of Accessibility Supports

Support Level 1: Default Embedded System Tools

The first level of supports is called the Default Embedded System Tools. (See the first level of the pyramid in Figure 5.1.) These tools are automatically available to a default user whose accessibility needs are sufficiently met through the basic test administration experience.

Default embedded system tools meet the common, routine accessibility needs of the most typical test takers. All examinees are provided these tools as appropriate, even examinees who have no documented support plan. Default embedded system tools include but are not limited to the following examples in online and paper tests:

- Magnifier Tool (online and paper)
- Browser Zoom Magnification (online)
- Answer Eliminator (online and paper)
- Test Directions Available on Demand (online and paper)
- Highlighter (online and paper)
- Keyboard Navigation (online)
- Scratch Paper (online and paper)
- Mark Item for Review (online and paper)

Default embedded system tools are common supports made available to all users upon launch or start of the test; they are the accessibility tools that nearly everyone uses routinely and assumes will be made available although we seldom think of them in this way. These tools are either embedded in the basic computer test delivery platform, or locally provided as needed. No advance request is needed for these supports.

Support Level 2: Open Access Tools

Open Access tools (the second level of the pyramid in Figure 5.1) are available to all users, but must be identified in advance, planned for, and then selected from the menu inside the test to be activated (online), or else provided locally.

Many examinees' unique sensory and communication accessibility needs are predictable and can be met through a set of accessibility features designed into the underlying structure and delivery format of test items. Rather than overwhelm the user with all the possible tools, Open Access tools provide just the tools needed by individual users, allowing true personalization of the test experience.

Open Access tools are slightly more intensive than default embedded system tools but can be delivered in a fully standardized manner that is valid, appropriate, and personalized to the specific access needs identified for an individual examinee. Some of these require the use of tool-specific administration procedures. In the Graphic Literacy assessment, Open Access tools include, but are not limited to the following examples:

- Color Contrast (online and paper)
- Line Reader (online and paper)
- Translated Verbal: Directions Only (online and paper)—locally provided

- Signed Exact English (SEE) for Directions Only (paper)—locally provided
- Answer Masking (online and paper)
- Dictate Responses (online and paper)
- Respond in Test Booklet or on separate paper (online and paper)
- Audio Indicator of Time Remaining (online and paper)
- Individual Administration (online and paper)
- Special Seating/Grouping (online and paper)

Open Access tools should be chosen carefully and specifically to prevent the examinee from becoming overwhelmed or distracted during testing. Room supervisors must follow required procedures. Prior to the testing experience, examinees need to have an opportunity to practice and become familiar and comfortable using these types of tools as well as using them in combination with other tools.

Support Level 3: Accommodations

Accommodations are high-level accessibility tools needed by relatively few examinees. (See the third level of the pyramid in Figure 5.1.) The Graphic Literacy assessment system requires accommodation-level supports to be requested by education personnel on behalf of an examinee. The accommodations must be identified in advance, planned, and selected from the menu inside the test to activate them (online), or else provided locally. Accommodations often require advance ordering of specialized paper materials from ACT. The advance planning process allows any needed resources to be assigned appropriately and documented for the examinee.

Typically, examinees who receive this high level of support have a formally documented need and have therefore been identified as qualifying for resources or specialized supports that require expertise, special training, and/or extensive monitoring to select, to administer, and even to use the support effectively and securely. These can include but are not limited to the following examples:

- Braille EBAE, contracted, includes tactile graphics (paper)
- Braille UEB with Nemeth contracted, includes tactile graphics (paper)
- Cued Speech (online and paper)
- Word-to-Word Bilingual Dictionary, ACT approved (online and paper)
- English Audio DVD (designed for user with blindness) (paper)
- English Audio Reader Script (designed for user with blindness) (paper)
- Signed Exact English (SEE): Test Items
- Abacus
- Extra Time

Decisions about accommodation-level supports are typically made by an educational team on behalf of and including the examinee. Accommodation decisions are normally based on a formal, documented evaluation of specialized need and require the examinee to have personal familiarization and successful prior experience with the tools so they may be used fluidly and effectively during the test experience. Accommodation supports require substantial additional local resources or highly specialized, expert knowledge to deliver successfully and securely.

Accommodations are available to users who have been qualified by the local governing school or employment authority to use them, (e.g., a school district, a work training agency, an employer, or a branch of military or other government service). Official determination of qualification for accommodation-level support by a governing school or workforce authority is usually documented in writing in the form of an accommodation plan, or such qualification may have been routinely recognized and permitted for this examinee by that governing authority. NCRC assessments, including the Graphic Literacy assessment, require that examinees who use accommodation-level supports have a formally documented need, as well as relevant knowledge and familiarity with these tools. Accommodations must be requested through the local test site according to WorkKeys assessment procedures, as defined in the administration manual. Appropriate documentation of accommodation need, as specified in the manual, must be provided prior to testing by the examinee, or by a local governing educational authority on behalf of the examinee.

Support Level 4: Modifications

Modifications are supports that are sometimes used during instruction, but when used in a testing situation, they alter the construct that the test is designed to measure. While they may provide an individual with the experience of taking 'a test,' modifications provide so much support that they actually prevent the examinee from having meaningful access to performance of the construct being tested. (See the top level of the pyramid in Figure 5.1.) Because modifications violate the construct being tested, they invalidate performance results and communicate low expectations of examinee achievement. Modifications are not permitted during Graphic Literacy testing and, if used, invalidate the resulting test score.

5.3 Allowable Embedded Tools, Open Access, and Accommodations

In our commitment to provide a fair testing experience for all examinees, ACT WorkKeys NCRC assessments provide an integrated system of accessibility supports that include accommodations as well as other forms (less intensive levels) of accessibility support. There are times when supports provided for those who test using the online format are combined with other types of locally provided or paper-format supports. The reverse is also true, as examinees using the paper format sometimes also take advantage of certain online options. Regardless of test format, all examinees who use accommodation-level accessibility features must have this use documented by appropriate school (or test site) personnel.

5.4 Valid Test Scores and Equal Benefit for All Examinees

ACT aims to ensure that all examinees may benefit equally from the Graphic Literacy assessment. Accommodations and other accessibility supports administered under these standardized conditions result in a valid and fully reportable NCRC score. Use of any accessibility supports that are not specified by ACT or not properly administered violate what the test is designed to measure and result in a score that is invalid and non-comparable for the stated purposes of the assessment.

Section 6: Test and Information Security

6.1 Test Security

In order to ensure the validity of the ACT WorkKeys Graphic Literacy test scores, test takers, individuals that have a role in administering the tests, and those who are otherwise involved in facilitating the testing process, must strictly observe ACT's standardized testing policies, including the Test Security Principles and test security requirements. Those requirements are set forth in the *ACT WorkKeys Administration Manual—Paper Testing* and the *ACT WorkKeys Administration Manual—Online Testing* and may be supplemented by ACT from time to time with additional communications to test takers and testing staff.

ACT's test security requirements are designed to ensure that examinees have an equal opportunity to demonstrate their academic achievement and skills, that examinees who do their own work are not unfairly disadvantaged by examinees who do not, and that scores reported for each examinee are valid. Strict observation of the test security requirements is required to safeguard the validity of the results.

Testing staff must protect the confidentiality of the WorkKeys test items and responses. Testing staff should be competent and aware of their roles, including understanding ACT's test administration policies and procedures and acknowledging and avoiding conflicts of interest in their roles as test administrators for WorkKeys.

6.2 Information Security

ACT's Information Security program framework is based on the widely recognized ISO/IEC 27000 standard (International Organization for Standardization, 2017). This framework was selected because it covers a range of information security categories that comprehensively matches the broad perspective that ACT takes in safeguarding information assets.

ACT has developed well defined procedures and processes for the daily handling and safeguarding of secure information, as well as procedures for safeguarding secure information in the event of a disaster or adverse event. The procedures and processes are overseen by the Information Security Officer and all ACT personnel are required to participate in security training. Access to secure information by ACT personnel is limited on a "need to know basis."

ACT's Information Security Incident Response Plan (ISIRP) brings needed resources together in an organized manner to deal with an incident (classified as an adverse event) related to the safety and security of ACT networks, computer systems, and data resources.

The adverse event could come in a variety of forms: technical attacks (e.g., denial of service attack, malicious code attack, exploitation of a vulnerability), unauthorized behavior (e.g., unauthorized access to ACT systems, inappropriate usage of data, loss of physical assets containing Confidential or Confidential Restricted data), or a combination of activities. The purpose of the plan is to outline specific steps to take in the event of any information security incident.

The Information Security Incident Response Plan charters an ACT Security Incident Response Team (ISIRT) with providing an around-the-clock (i.e., 24/7) coordinated security incident response throughout ACT. Information Security management has the responsibility and authority to manage the Information Security Incident Response Team and implement necessary ISIRP actions and decisions during an incident.

Section 7: Preliminary Field Test Data Analyses and Findings

ACT conducted a series of field test studies to better understand the psychometric properties of the Graphic Literacy assessment and to make decisions related to test administration. Specifically, the studies were designed to access the following questions:

- What is the appropriate test time for the Graphic Literacy assessment?
- What is the reliability (internal consistency) of the Graphic Literacy assessment?
- What is the factor structure of the Graphic Literacy assessment?
- Do examinee responses to the assessment items indicate that the IRT assumptions are met?
- How was the primary scale score established and what methods will be used for equating to ensure that scale scores have the same meaning regardless of form?

For each of the field test studies, ACT attempted to recruit samples representative of the regular WorkKeys testing population. Although variability existed in terms of demographic groups amongst the field test studies, each sample consisted of approximately 60% high school students and 40% adult test takers, 55% female and 45% male test takers, and 60% Caucasian, 16% African American, and 8% Hispanic test takers.

7.1 Testing Time Limit

ACT designed the Graphic Literacy assessment to be a power test where examinees are provided adequate time to complete the assessment (Miller, Linn, & Gronlund, 2013). As such, testing time should allow sufficient time for at least 90% of examinees to complete the assessment. In the first field test study, with examinees taking assessments both on paper and online, ACT allowed examinees either 55 or 60 minutes to complete the assessment. With these time conditions, ACT surveyed examinees to learn if they believed that they had sufficient time to complete the assessment.

In the first study, ACT found that the mean testing time was slightly less than 30 minutes. ACT also learned that 95% of examinees completed the assessment in less than 50 minutes for both the paper and online formats. On the post-administrative surveys, 97% of examinees who were given 55 minutes Agreed or Strongly Agreed with the statement, I had a sufficient amount of time to answer each test question. Based on these findings, ACT concluded that a 55-minute time limit would provide sufficient time to test.

For the second field test study, ACT continued to evaluate the amount of time that examinees required for testing. In the second field test study, all examinees took two forms of the assessment online. The findings of the second field test study confirmed that 55 minutes was a reasonable time limit. Although mean testing time increased slightly (Form A mean testing time = 32.18 minutes, and Form B mean testing time = 34.08 minutes), over 95% of examinees completed all items on the forms. On the post-administrative survey, 85% of the examinees Agreed or Strongly Agreed with the statement, I had a sufficient amount of time to answer each test question.

Based on the examinee response patterns, completion rates, and survey responses, ACT concluded that allowing examinees 55 minutes to complete the assessment was reasonable and that it would result in the test not being speeded.

7.2 Score Reliability

ACT used Coefficient Alpha to estimate the overall reliability of the Graphic Literacy assessment. Coefficient Alpha is a measure of the internal consistency of the items constituting the assessment. Prior to administration, ACT determined that 0.80 would be a minimal acceptable level of internal consistency. Following the calculation of Coefficient Alpha, ACT calculated the Standard Error of Measurement (SEM). SEM summarizes the amount of error or inconsistency in scores on a test. Scale score SEMs were estimated using a four-parameter beta compound binomial model (Kolen, Hanson, & Brennan, 1992). If the distribution of measurement error is approximated by a normal distribution, true score scores for about two-thirds of the examinees are within plus or minus one SEM from their reported scale score. Table 7.1 presents the reliability estimates and SEM based on data from the second field test study.

Table 7.1: Graphic Literacy—Reliability Estimates and Standard Error of Measurement of Form A and Form B

		Raw Score		Scale Score	
Form	N	Coefficient Alpha	SEM	Coefficient Alpha	SEM
Form A	1,170	0.85	2.34	0.85	1.66
Form B	1,096	0.85	2.35	0.86	1.73

ACT also estimated the reliability of the forms applying a multivariate generalizability framework (Brennan, 2001). Generalizability theory provides estimates of universe score variance, absolute score variance, and absolute error variance. The analysis allows for the calculation of a G-coefficient, which provides an alternative estimate of internal consistency. The G-coefficients estimated for the two forms of the Graphic Literacy assessment were 0.85 and 0.86, respectively. Based on the reliability and generalizability analyses, ACT concluded that the Graphic Literacy met the requirements for score reliability.

7.3 Factor Structure

For both the second and third field test studies, ACT used Exploratory factor analysis (EFA) assess dimensionality for the Graphic Literacy assessment. EFA uses an inter-item correlation matrix to identify the underlying factors accounting for the observed variance in the items. In these analyses, four criteria are evaluated to determine dimensionality. A scree plot of eigenvalues can be used to determine test dimensionality, with unidimensionality indicated by having only one eigenvalue above the “elbow” in the scree plot. Hatcher (1994) suggested that a factor should be retained if it accounted for at least 10% of the total variance. Reckase (1979) suggested that if the first factor explains 20% of the variance of a set of items, the item set should be considered unidimensional. Hattie (1985) suggested that the first factor is relatively strong if the factor difference ratio index (FDRI) (Johnson, Yamashiro, & Yu, 2003) is larger

than 3. FDRI is defined as the ratio of the difference between the eigenvalue of the first factor and the second factor to the difference between the eigenvalue of second and the third factor.

Table 7.2 summarizes the eigenvalues and FDRI for the Graphic Literacy forms. Evaluation of the scree plots for all three forms indicated a single factor as the "elbow" appeared immediately following the first eigenvalue. Applying the rules of Hatcher (1994), Reckase (1979), and Hattie (1985) also supported the conclusion that each Graphic Literacy form was unidimensional. The data presented in Table 7.2 indicate that more than 10% of the variance on all three forms could be attributed to the first factor and no other factor met the 10% threshold; data indicate that the variance attributed to the first factor exceeded 20% on all forms; and the FDRI was larger than 3 for all forms with Form C—online being the smallest (FDRI = 10.50). Collectively, the findings provide strong evidence that the Graphic Literacy assessment is a unidimensional measure.

Table 7.2: Graphic Literacy—Summary of Eigenvalues and Factor Difference Ratio Index (FDRI)

	First Factor	Second Factor	Third Factor
Form A—Online			
Eigenvalue	9.53 (29.8%)	1.78 (5.6%)	1.36 (4.2%)
Difference		7.75	0.42
FDRI			18.32
Form B—Online			
Eigenvalue	9.75 (30.5%)	1.77 (5.5%)	1.38 (4.3%)
Difference		7.99	0.38
FDRI			20.96
Form C—paper			
Eigenvalue	11.58 (37.4%)	2.02 (6.5%)	1.39 (4.5%)
Difference		9.56	0.63
FDRI			15.20
Form C—online			
Eigenvalue	11.72 (37.8%)	2.10 (6.8%)	1.18 (3.8%)
Difference		9.62	0.92
FDRI			10.50

Note. The percentage in the parenthesis is the percentage of total variance accounted for by that factor.

7.4 3-PL IRT Model

With evidence indicating that the Graphic Literacy assessment was unidimensional, ACT checked the IRT assumption of Local Independence using the Q_3 method (Yen, 1984). For an item pair, Q_3 is the correlation of item residuals where the residual is the difference between the observed item responses and the responses predicted for each item by an IRT model (3-PL in this study). Items that do not share

a common graphic are considered locally independent. The Q_3 indices for all items that are not in a common set were computed and were interpreted as the baseline. Following that step, Q_3 indices for items that share a graphic were computed and compared to the baseline to evaluate whether the items in a set were more dependent than the items that were not within a set. The 95th percentile of the baseline was defined as the cut point. If the Q_3 for a pair of items within a set was larger than the cut point, the item pair met the assumption of Local Independence.

The Q_3 analysis for the Graphic Literacy forms indicated that the items within a set do not show higher correlation compared to those items between sets. Evaluating the matrices led to the conclusion that no strong evidence of Local Item Dependence existed among items for all forms of the Graphic Literacy assessment.

With both IRT assumptions met, ACT analyzed item data using a 3-PL IRT model for Forms A and B for the second field test study. Table 7.3 presents the summary IRT parameter estimates for the two forms.

Table 7.3: Summary of IRT Item Parameter Estimates

	IRT a				IRT b				IRT c			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Form A	1.08	0.32	0.55	1.86	0.39	0.90	-1.36	1.98	0.19	0.06	0.09	0.30
Form B	1.01	0.30	0.53	1.74	0.42	0.99	2.00	2.23	0.18	0.07	0.08	0.31

Figure 7.1 shows the Test Characteristic Curve (TCC) and the Test Information Function (TIF) for the base form administered in the second field test study. Although the Graphic Literacy TCC and TIF indicate an assessment that provides a good deal of measurement information for examinees who are within two standard deviations of the mean, it appears that the assessment has greater precision for examinees above the mean as compared to those below the mean. Test information appears to peak at approximately one standard deviation above the mean.

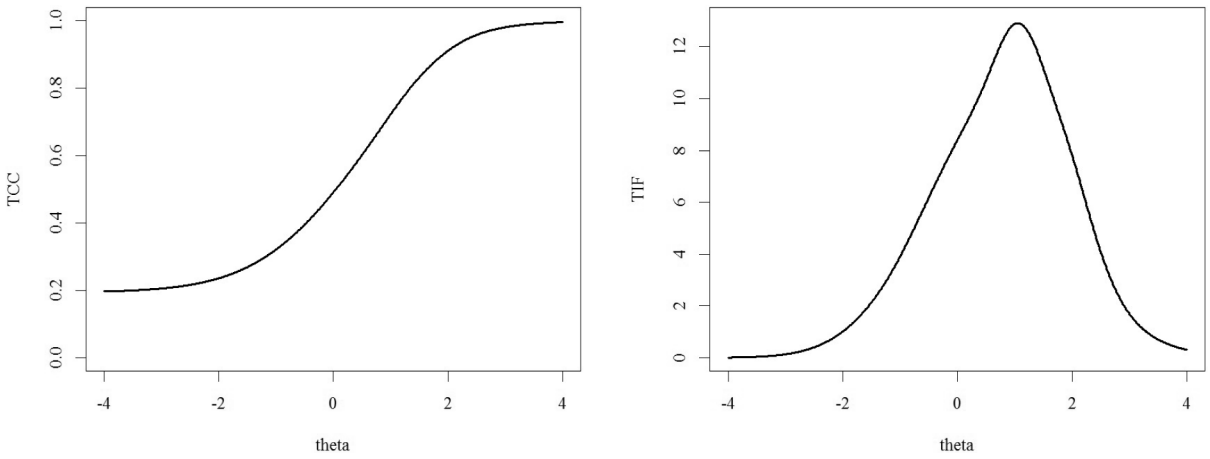

Figure 7.1: Graphic Literacy—Test Characteristic Curve and Test Information Function of Base Form

7.5 Graphic Literacy Score Scale and Form Equating

The base form score scale was established on the Graphic Literacy assessment using examinee response data from the second field test study. The primary purpose was to develop a score scale so that examinee responses to items on all Graphic Literacy forms could be converted from raw scores to scale scores, and then analyzed and compared on a common score metric. For the scaling, Form A was defined as the Graphic Literacy base form. The scaling process used IRT scaling (Ban & Lee, 2007) and applied the arcsine transformations (Kolen, 1988; Kolen & Brennan, 2004), resulting in the base form raw-to-scale score conversions. By applying the arcsine transformations, the Conditional Standard Errors of Measurement (CSEMs) were relatively equal across the score scale.

ACT set the scale scores to range from 65 to 90, which was the traditional scale score range used for the NCRC assessments. Applying the scale score average and CSEMs of the current WorkKeys 1.0 assessments, the target mean and target CSEM of the arcsine transformation were set to 77.9 and 1.7. Figure 7.2 presents the CSEMs across the full range of scale scores for the base form.

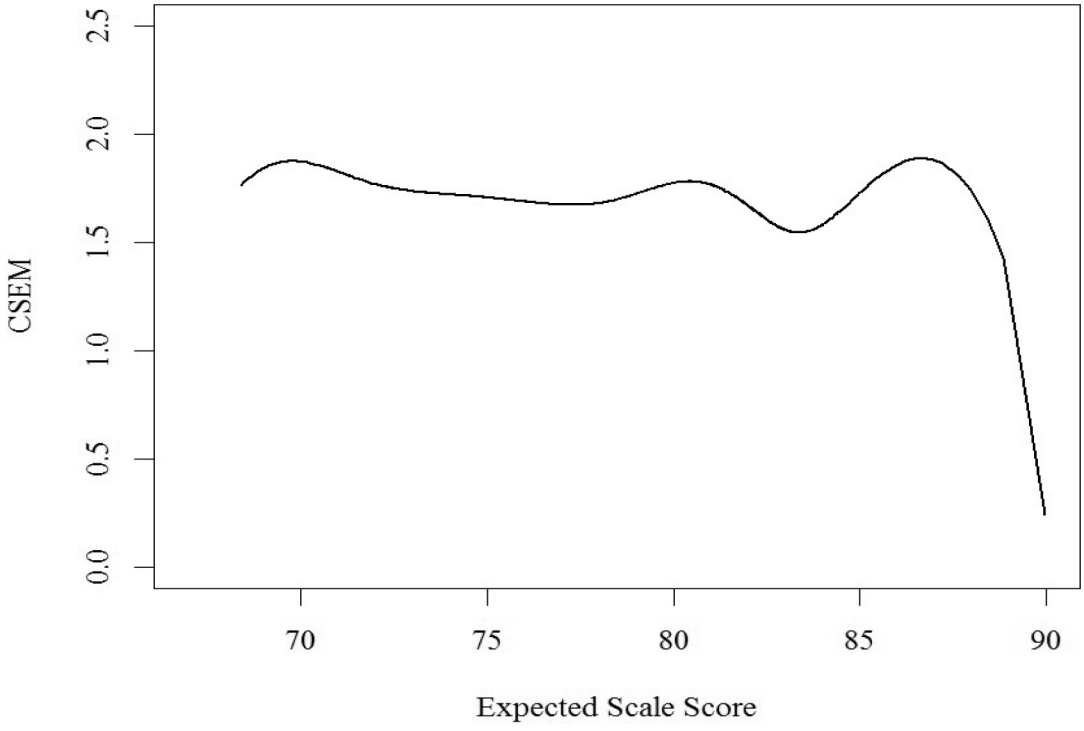

Figure 7.2: Graphic Literacy—Conditional Standard Errors of Measurement for Scale Scores

ACT equated scores from Form B to the base form (Form A) using IRT true score equating (Kolen & Brennan, 2004). Table 7.4 provides the summary scale score statistics for Forms A and B following equating. Based on equated scores, only trivial differences existed in mean scores for examinees who took Form A compared to those who took Form B. Also, examinees who took Form B appeared to have slightly greater variability in their scores.

Table 7.4: Graphic Literacy—Equated Scale Score Distribution Summary Statistics

	N	Mean	Median	SD	Min Score	Max Score
Form A	1,170	78.02	78	4.45	65	90
Form B	1,096	77.99	78	4.83	65	90

Section 8: Establishing Level Score Standards

All WorkKeys assessments report to examinees their performance in terms of a scale score and a level score. The interpretation and use of the scale score is appropriate when the assessment is used to evaluate achievement or effectiveness of an educational or training program. The scale scores provide the user with finer performance distinctions than the level score. As such, they are more sensitive to gains achieved through educational or training programs.

WorkKeys level scores should be interpreted and used for selection, promotion, or other high-stakes purposes. WorkKeys assessments were originally developed to be used in conjunction with a job profiling process for employee selection and promotion. The job profiles are aligned to the assessment score levels and not to the more granular scale score.

Section 7 provided information regarding the method ACT applied to develop the Graphic Literacy score scale. Following the development of the Graphic Literacy score scale, ACT conducted a standard setting in order to identify the points on the scale score where examinees who have scored at or above that point have demonstrated the ability to earn a specified level score.

8.1 Standard Setting Method

ACT conducted a standard setting study to establish the minimum scores required to achieve each of the five Graphic Literacy levels. To establish the minimum score, ACT assembled a panel of experts consisting of educators and business people, some of whom are current WorkKeys customers. Because the Graphic Literacy assessment is a criterion-referenced test, scores on the assessment are aligned to a set of skills that an examinee demonstrated. (Please refer to Section 2.4 for detailed PLDs associated with each level score.) The goal of the standard setting process was 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 the study (Schulz & Mitzel, 2005). It is a variation of the popular Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012). The primary difference between Mapmark and Bookmark is the Item Map. The Ordered Item Booklet (OIB), used in both procedures, has a sample of items from the item pool ordered from easiest to hardest. The Mapmark procedure also includes an Item Map, on which the difficulty of an item is

indicated by actual scale value. The Item Map, therefore, shows “how much” more difficult one item is than another item.

Mapmark with Whole Booklet Feedback is a three-round process. In Round 1, the panelists (a) took the Graphic Literacy assessment, (b) reviewed the PLDs, (c) reviewed test items and their associated Scale Score, (d) linked test items to the PLDs, and (e) 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 items that they felt were easy enough for a minimally qualified examinee in the skill level to have mastered, and those items that were too difficult for a minimally qualified examinee to have mastered. In this context, mastery was defined as having a 2-in-3 chance of success (or a response probability of .67) on the item. This was done to establish the initial cut scores for the five levels (e.g., Levels 3–7).

For Round 2, the panelists received feedback regarding their bookmark placements relative to recommended Scale Scores on the item map scale and to the group’s median cut score. The group was then provided with Whole Booklet Feedback. Specifically, they were provided with data showing how fifteen examinees answered each of the items on the base form. Item data was provided for three examinees who scored at or near the Round 1 cut score for each skill 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 each PLD. Using all of this information, panelists were asked to repeat the process of placing bookmarks in the OIB for each level.

For Round 3, panelists received feedback regarding their bookmark placement from Round 2. The feedback consisted of impact or consequential data of their Round 2 placements, which provided the percentage of examinees who achieved at or above the cut scores set for each skill level. ACT emphasized to the panelists that the PLDs should take precedence since the assessments are criterion-referenced. With that, they again placed bookmarks in the OIB.

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 reviewed a Cut Score Distribution Chart showing the distribution of panelists’ Round 3 cut scores across all skill levels. Finally, the panelists discussed impact data based on the final cut scores. The panelists approved the final median cut scores to define the five performance levels.

The Design Team reviewed the work of the Standard Setting. Design Team members evaluated whether the work of the panelists achieved the desired result of a criterion-referenced assessment with level scores aligned to the PLDs. After reviewing the panelists’ work and recommendations, the Design Team approved the recommended cut scores for the five score levels of the Graphic Literacy assessment.

Section 9: Final Note Regarding On-going Studies

ACT plans to continually collect and analyze data from the Graphic Literacy assessment. These analyses will range from studies of score reliability, factor structure of the assessment, fairness evaluations, and the relationship of assessment scores to significant outcome variables (i.e., job performance ratings, training completion rates and scores, program grades). ACT will periodically publish papers providing the findings of these studies. Additionally, ACT will provide information and analyses in the electronic Technical Manual in an ongoing manner as studies are completed. The information, data, and analyses are designed to provide understanding and insight regarding score interpretations and usage.

References

ACT. (2008). *WorkKeys Locating Information technical manual.* Iowa City, IA: ACT, Inc.

ACT. (2014). *Foundational skills: What makes a skill "foundational"?* Retrieved from https://www.act.org/content/act/en/search.html?searchkey=Foundational%20Skills&submit=Search.

ACT. (2016). *ACT national curriculum survey 2016.* Iowa City, IA: ACT, Inc.

ACT. (2017). *Job profiling database*. Iowa City, IA: Iowa City, IA: ACT, Inc.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council for Measurement in Education (NCME). (2014). *Standards for educational and psychological testing* (2014 Edition). Washington: AERA Publications.

Allen, M. J., & Yen, W. M. (2002). *Introduction to measurement theory.* Long Grove, IL: Waveland Press.

Avgerinou, M. D. (2015). Toward a cohesive theory of visual literacy. Retrieved from https://www.researchgate.net/publication/267988880_Toward_a_Cohesive_Theory_of_Visual_Literacy.

Ban, J., & Lee, W. (2007). *Defining a score scale in relation to measurement error for mixed format tests.* (CASMA Research Report No. 24). Iowa City, IA: University of Iowa.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.) *Assessment and teaching of 21st century skills* (pp. 17–66). Springer: New York.

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.

Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19.

Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). Beyond academics: A holistic framework for enhancing education and workplace success. ACT Research Report Series 2015 (4). Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACT_RR2015-4.pdf.

Center for Applied Special Technologies (CAST). (2011). Universal design for learning guidelines version 2.0. Wakefield, MA: Author. Retrieved from www.udlcenter.org/aboutudl/udlguidelines.

Crick, J. E., & Brennan, R. L. (2001). *GENOVA A general purpose analysis of variance system version 3.1*. A Fortran 77 program for analysis of variance and generalizability analyses with balanced designs.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*(5), 382–393.

Fedorchak, G. (2013). Access by Design—Implications for equity and excellence in education. Internal whitepaper prepared on behalf of the NH Department of Education for the Smarter Balanced Assessment Consortium. (Internal Document—no public link).

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd Ed.). Analytics Press: Burlingame, CA.

Friel, S. N., & Bright, G. W. (1996). Building a theory of graphicacy: How do students read graphs? Retrieved from http://eric.ed.gov/?id=ED395277.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 124–158.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139–164.

Holland, P. W., & Wainer, H. (1993). *Differential item analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

International Organization for Standardization. (2017). ISO/IEC 27000 family-information security management systems. Retrieved from https://www.iso.org/isoiec-27001-information-security.html.

Johnson, J. S., Yamashiro, A., & Yu, J. (2003). *ECPC annual report: 2002*. Ann Arbor, MI: English Language Institute, University of Michigan.

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement, 50*, 1–73.

Kolen, M. J. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement, 25*, 97–110.

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices.* New York: Springer Science and Business Media.

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, *29*, 285–307.

Koomey, J. G. (2008). *Turning numbers into knowledge: Mastering the art of problem solving* (2nd Ed.). Oakland, CA: Analytics Press.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd Edition) (pp. 225–253). New York: Routledge.

Mayer, R. E. (2009). *Multi-media learning* (2nd Ed.). Cambridge University Press: New York.

Messick, S. (1989). Validity. In American Council on Education Series on Higher Education. Robert L. Linn (Ed.), *Educational measurement: 3rd edition* (pp. 13–103). New York: Macmillan.

Miller, M. D., Linn, R. L. & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th edition). New York: Pearson.

Mislevy, R. J., Almond, R. G., & Lukas J. (2004). A brief introduction to evidence centered design (CSE Report 632). Los Angeles, CA: University of California-Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from files.eric.ed.gov/fulltext/ED483399.pdf.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*(4), 6–20.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design.* Princeton, NJ: Educational Testing Service.

Moreno, R., & Valdez, A. (2007). Immediate and delayed effects of using a classroom case exemplar in teacher education: The role of presentation format. *Journal of Educational Psychology, 99*, 194–206.

National Network of Business and Industry Associations (NNBIA) (2014). *Common employability skills: A foundation for success in the workplace: The skills all employees need, no matter where they work.* Retrieved from http://nationalnetwork.org/resources/common-employability-skills-3/.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics, 4*(3), 207–230.

Rogers, Y., & Scaife, M. (1998). How can interactive multimedia facilitate learning? In J. Lee (Ed.) *Intelligence and multimodality in multimedia interfaces: Research and applications* (pp. 1–25). AAAI Press: Menlo Park, CA.

Schulz, E. M, & Mitzel, H. C. (2005, April). *The Mapmark standard setting method.* Paper presented at the annual meeting of the National Council on Measurement and Education, Montreal, Canada.

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*(3), 560–578. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1756–8765.2009.01066.x/full.

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher, 21*(1), 14–23.

Wolodtschenko, A., & Forner, T. (2007). Prehistoric and early history maps in Europe: Conception of Cd-atlas. *E-Perimetron, 2*(2), 114–116. Retrieved September 20, 2016 from http://www.e-perimetron.org/Vol_2_2/Wolodchenko_Forner.pdf.

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three parameter logistic model. *Applied Psychological Measurement, 8*, 125–145.