ACT® WorkKeys®

WorkKeys
Graphic Literacy
Technical Manual

Version 2.0
Fall 2019

Table of Contents

List of Tables

List of Figures

Chapter 1

General Description of the Graphic Literacy Assessment

1.1 Graphic Literacy—Defined

ACT developed a new workplace Graphic Literacy assessment to replace the Locating Information assessment and as a part of the ACT WorkKeys® National Career Readiness Certificate® (NCRC). In developing the Graphic Literacy assessment, ACT utilized four critical sources of information: (1) knowledge gained through 23 years of administering the Locating Information assessment, (2) knowledge gained through ACT's job profiling services that included profiling thousands of jobs where the Locating Information assessment was identified as critical, (3) input and feedback from a panel of external Subject Matter Experts (SMEs) with experience in workforce development or graphics/visual literacy, and (4) professional literature published over the past 50 years related to how students and workers read, interpret, and use graphical information.

The workplace Graphic Literacy assessment, like all ACT WorkKeys assessments, is designed to measure relevant skills required for success in today's workplace. 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 achievements or deficits.

ACT defined the Graphic Literacy assessment as a measure of an essential 21st-century workplace domain 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. As such, the skills required by the assessment include the ability to comprehend information presented in graphical format and then to take that information and solve some type of problem. At the lower ability levels, the expectation is that test takers will be able to find and identify the correct information from the graphic representation. At the middle ability levels, expectations are that test takers will not only be able to find and identify the correct information, but they will be able to correctly identify trends and patterns, make interpretations, make comparisons, and derive reasonable conclusions. At the high ability levels, expectations are 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.

1.2 The Workplace Skills Gap and the WorkKeys Solution

Employers have long relied on America's schools to educate the workforce of the future. Over the past few decades, traditional credentials such as a high school diploma or a four-year college degree no longer assure employers that a worker has the required skills to participate in the fast-paced, high-performing workplace. Increasingly, employers find that workers often have serious gaps in many of the personal and foundational skills needed for success. As business and industry processes and practices become progressively more complex, they perceive that workers' skill levels have improved little in both behavioral skills (e.g., collaboration, conscientiousness, and timeliness) and foundational skills (e.g., reading, writing, mathematics, and critical thinking).

Over the past 25 years, requirements for nearly all jobs have changed in the developed economies, resulting in drastic changes in worker skill requirements. Work environments are technology-centered, problems are often poorly defined, people work in teams to deal with these problems, and employers seek innovative answers. These new ways of working require a different set of job skills from those found in the manufacturing/industrial economy. Autor, Levy, and Murnane (2003) analyzed job tasks, categorizing them as manual, routine, or abstract. From 1960 to 2002, they found that the percentage of abstract tasks performed in the workplace increased by approximately 25%, while the percentage of manual and routine tasks decreased by nearly 10%. Clearly, 21st-century workers must deal with a technology and information-rich work environment, where abstract thinking is a requirement, and fewer and fewer tasks require either manual labor or routine operations (Autor et al., 2003; Griffin, Care, & McGaw, 2012).

Economic and workforce leaders debate the significance of the skills gap and its influence on economic growth (Bessen, 2014; Cappelli, 2012; Krugman, 2014). The skills gap is a term used to describe a problem that employers and hiring managers frequently face. The skills gap occurs because many well-paying jobs exist; but, due to the shortage of qualified workers, employers are unable to find qualified workers to fill them. From the workers' perspective, the skills gap means that many willing workers are unable to find employment because they lack the required skills. From a business perspective, it means that jobs are not filled resulting in lost opportunities and unrealized economic gains. From an overall economic perspective, it means that unemployment is unacceptably high and that economic growth is stagnant or fails to reach its full potential.

ManpowerGroup® (2015) surveyed 41,700 global employers and found that 38% of employers state they experience problems finding qualified workers. Thirty-two percent of United States employers reported experiencing problems finding qualified workers. Goldin and Katz (2008) provide evidence demonstrating that, since 1970, educational achievement in the United States has increased only marginally while technological advances and job requirements have greatly increased. They analyze the race between education and technology, and conclude that many of the economic trends that have developed over the past 30 years are a result of educational advances not keeping up with the advances in technology and worker demands. As a result, a discrepancy exists between employer expectations and the skill sets that many workers have (Autor, 2015; Goldin & Katz, 2008).

Such perceived gaps in job skills reflect a dynamic redrawing of America's demographic profile. The fastest growing demographic groups in the United States are the least educated (Kirsch, Braun,

Yamamoto, & Sum, 2007). Changes in the nation's demographic profile will present challenges to both the educational system and employers seeking highly skilled workers. These challenges require new approaches to both schooling and hiring practices. Paradoxically, these challenges coincide with the growth of a knowledge-based economy in which most job growth will be in jobs that require some postsecondary education or training (Carnevale & Desroches, 2003).

ACT created the WorkKeys system to address the discrepancy between foundational skill levels and job requirements. Because of the discrepancy, the WorkKeys system provides a solution that is beneficial to both employers and workers. WorkKeys assessments provide both employers and test takers clear, evidence-based, objective information about job skills. WorkKeys job profiling services provide employers with clear information regarding the foundational skill demands required for success in specific jobs. The ACT KeyTrain® online curriculum program provides workers with the opportunity to improve their skills and achieve the required levels to qualify for jobs. The WorkKeys system provides opportunities for employers to hire the right person for the job, and it provides workers with the opportunity to qualify and demonstrate that they possess the foundational skills required for success.

1.3 Graphic Literacy as a Foundational Workplace Skill

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 has led to the development and use of more and more graphical representations in the workplace (Few, 2012). The increase in the use of graphical representations in the workplace has been confirmed by ACT's job profiling[1] (ACT, 2017), and its importance has been confirmed through ACT's interviews with outside workplace development professionals (ACT, 2016). As a result, ACT has concluded that the ability to comprehend and accurately interpret graphic materials in the workplace has become as foundationally important to worker success 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 (Langenfeld, 2014). In many ways, the assessment was one of the first tests of workplace graphic literacy. Locating Information's content was developed through ACT's work with individuals in the workplace. These individuals contracted with ACT and provided workplace documents and ideas on how the information was used in the workplace. ACT utilized this content to develop realistic workplace scenarios and questions to build 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. More recently, ACT further confirmed the importance of graphical literacy skills through the findings of the ACT National Curriculum Survey (NCS) (ACT, 2016). The NCS found that employers identify the ability to analyze and interpret data in graphs and tables as an important workplace skill.

Where the original Locating Information assessment only assessed test takers' abilities to locate, compare, summarize, and analyze information presented in graphical format, the new Graphic Literacy assessment is designed to measure these skills plus others. 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).

> With the advent of modern computer tools, creating graphs from data involves trivial effort. In fact, it has probably become too easy. Graphs are often produced without thought for their main purpose: to enlighten and inform the reader (Koomey, 2008, p. 161).

In evaluating the current Locating Information assessment, ACT's 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 measured the ability to interpret and analyze information. When the team decided that it was appropriate to expand the construct by including the ability to create and evaluate the effectiveness of graphics (Avgerinou, 2015; Friel & Bright, 1996; Shah & Freedman, 2011), the title of Locating Information seemed incomplete and unsatisfactory. To capture a more thorough description of the construct and to be consistent with the professional literature (see references listed in Appendix on Graphic Literacy), the design team concluded that the name Graphic Literacy more accurately captured the essence of the assessment's construct.

Graphic Literacy is a subcomponent of multi-media literacy. Mayer (2009) defined 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 assumed that the individual's information processing system includes dual channels for learning through pictorial/graphical stimulus and written/verbal stimulus. Further, each channel had a finite capacity for processing information. Active learning occurred when an individual attended to a stimulus, determined the importance and relevancy of different aspects of the stimulus, and related the important and relevant aspects to past learning so as to create a coherent body of knowledge. Active learning, when confined to a single channel, limited the amount of learning and comprehension that potentially occurred. Presenting material using a dual channel approach (pictures and words) provided the learner with an increased opportunity to process and retain information. Although each channel has a finite capacity, integrating the information presented in both channels increased overall capacity and the likelihood of learning and retention (Mayer, 2009; Moreno & Valdez, 2007).

Few (2012) maintained 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 believed that graphs are tools that, when used effectively, provided 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) considered the development of effective graphical presentations of quantitative information to be one of the significant learning advancements of the past 250 years. In the first 220 years of graphic representations, the transfer of complex quantitative information into effective graphics required considerable time and expense. With the introduction of

personal computing, businesses and industries have witnessed a proliferation in the use of graphical representations. Increasingly, information is both analyzed and communicated with the assistance of graphics (Few, 2012).

Graphic Literacy utilizes words, pictorial shapes and symbols, and numbers as visual representations to communicate information and inform decision making. In ACT's definition, graphical representations are used to communicate both quantitative and qualitative information. The basic skill in comprehending graphical representations is locating information; however, the assessment measures more than just the basic skill. It 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.

1.4 Graphic Literacy—Assessment Claims

The three Graphic Literacy claims addressed workforce development issues including improving worker access to better jobs, improving worker productivity, and reducing worker turnover rates. The Graphic Literacy assessment was designed to measure specific skills as one part of a suite of assessments that assess (a) work and career readiness for high school students as a part of state accountability programs, (b) work and career readiness indicators for adults seeking state unemployment services, (c) job placement to assist businesses in identifying individuals who had the foundational skills needed to succeed, and (d) the requirements used to certify ACT Work Ready Regions and Communities.

ACT has defined the following three claims regarding Graphic Literacy score interpretation and usage.

> **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 #2:** U.S. companies who 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 gains in productivity (for example, measured as increased output per day) from new employees than if the company had hired examinees whose scores do not reach that level.

> **Claim #3:** U.S. companies who hire U.S. examinees of high school or work force 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 if the companies had hired examinees whose scores do not reach that level.

Note. For further elaboration on the Graphic Literacy assessment claims, including the assumptions associated with each claim, see Chapter 11 – Validity.

1.5 Test Users and Stakeholders

The critical stakeholders and intended test users 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 from high school age through the adult working lifetime. High school students take the assessment to gain an understanding of their level 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 separating themselves from other graduates. Working adults take the assessment to either 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 field.

Individuals who take the WorkKeys NCRC assessments late in their high school careers gain useful feedback in terms of whether they are college and career ready. Schultz and Stern (2015) found that a wide gap existed between students' perceptions of their college and career readiness and their skill levels as measured by the WorkKeys assessments. Their findings suggested that the WorkKeys assessments provided students valuable information, and students believed that the assessment scores were useful in helping them plan their future. These two findings indicated that WorkKeys assessment scores could be effectively used by schools, workforce development agencies, and others to assist young people in educational and career planning (Schultz & Stern, 2015).

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 WorkKeys Graphic Literacy assessment provides high schools and school districts with student data regarding the extent to which students have mastered the aspects 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 and 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 WorkKeys assessments in a variety of ways. Many community colleges use the program as part of the process for determining acceptance into Career and Technical Education

programs. Other community colleges use the assessments for program evaluation. Additionally, community colleges use the assessments as a means of assisting their graduates in obtaining employment.

The Graphics Literacy assessment has the flexibility to assist community colleges to improve their programs in different ways. It can assist a program in identifying students who have the foundational skills required to complete a specific program of study. In this way, it assists a program in achieving higher completion rates. In other cases, it can be used as a means of program evaluation allowing teachers to evaluate the extent to which students have mastered foundational skills. Lastly, because it is recognized by thousands of employers, it can help graduating students obtain employment.

State and local workforce development offices utilize the assessments 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 assessments, 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 appropriate WorkKeys assessments and then use their test scores as an additional piece of information to determine which candidates to interview.

1.6 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 (HF Data Strand: Substrands Data Analysis; Gathering & Presenting Data)

2. Translating to a different form of graphic (HF Data Strand: Substrands Data Analysis and Gathering & Presenting Data)

3. Evaluating bias in the use of a graph (HF Data Strand: Substrands Data Analysis and Gathering & Presenting Data; Scientific Reasoning Strand: Substrand Scientific Reasoning & Argument; Modeling. Cross Cutting Science Concepts Using Patterns; Cause & Effect)

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

Notes

[1] ACT's job profiling services has found that, after profiling over 21,000 jobs throughout the United States primarily in business, manufacturing, and health care, 64% require skills related to comprehending and using information presented in graphic format.

Chapter 2

Test Development

2.1 Graphic Literacy—Overview

WorkKeys® Graphic Literacy is designed to assess an essential 21st-century workplace domain 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. The ability to find, summarize, compare, and analyze information found in workplace graphics is critical to workplace success. The Graphic Literacy assessment measures skills that individuals use when they read and comprehend graphical materials to solve work-related problems.

To ensure that the Graphic Literacy assessment would measure useful and relevant skills, a team composed of individuals from within ACT including Test Development Content, Psychometric Research, Industrial/Organizational Psychology, and Assessment Design was established to design the specifications for the Graphic Literacy assessment. The team pooled resources to define the Graphic Literacy construct, test specifications, and develop item prototypes. The design team's work was reviewed by external Subject Matter Experts (SMEs) who also provided feedback and recommendations, which were incorporated by the team.[1]

Through a review of the pertinent 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. These two variables and their interplay are central to the assessment's construct definition and are described below.

Tasks defined as constituting graphic literacy are successfully completed through the individual's cognitive interpretation and use of the relevant information communicated by the graphic. Consequently, tasks presented through test items are defined by the interaction of the graphic complexity and the complexity of the cognitive process. As a result, ACT bases its definition of overall item level on the *interaction* between *graphic complexity and the cognitive process* required to successfully complete the task elicited by the question.

2.2 Graphic Complexity

WorkKeys defines the concept of graphic materials broadly. Graphic materials 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. Graphical materials include all of the graphic resources identified in the graphic literacy definition, as well as other visual resources designed to communicate information, such as directions, data readings, data trends, variable relationships, and summaries.

The use of diagrams, maps, graphs, and other visuals to communicate information have been used by humans 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).

Humans appeared to have understood from the earliest points in history that learning through pictures or visual representations is an effective means of transmitting knowledge and information (Mayer, 2009). In the 21st century, with the prevalence of computer software packages, the use of graphical representations to convey information has grown exponentially (Griffin et al., 2012; Koomey, 2008).

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

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 (composite graphics)
Total Number of Graphics	One	May be two	May be multiple	May be multiple

The design team categorized graphic complex into four categories: Simple, Low Moderate, High Moderate, and Difficult. The characteristics of each of these was 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 stimuli into one of the four categories is somewhat subjective, using the defined characteristics permits the content development team to classify the graphic stimuli with a great deal of consistency.

2.2.1 Graphic Complexity Classification Evaluation

ACT conducted a study to evaluate the content specialists' ability to consistently classify different graphic materials into the four categories applying the principles described in Table 2.1.

Study No. 1: The first study asked four content specialists who regularly worked on the Graphic Literacy assessment to discuss how they classified graphics and to identify 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 graphics represented a variety of graphic types including line 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 graphics 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 their definitions and means of classification. The result was the development of Table 2.1 along with additional resource information that they would use to aid 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 graphics 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 classification consistency.

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

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 elicit the examinee to perform one cognitive step. For example, the item task could elicit that the examinee 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 single cognitive step.

The task associated with a second item might elicit that the examinee locate a piece of information and then use the information in a second cognitive step. For example, the item task could elicit that the examinee locate two pieces of data within a table and then make a decision as to whether the process is meeting the standard. 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 an appropriate decision. The examinee is solving the task through two cognitive steps.

The task associated with a third item might elicit that the examinee 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 three separate cognitive steps.

Graphic Literacy—Skill Definitions

For Graphic Literacy, 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 et al., 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 costing between the 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

2.4 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 little to no proficiency do not receive a level score.) Graphic Literacy score or performance levels are determined by the interaction of the graphic complexity categories with the cognitive skill processes. The Graphic Complexity performance levels defined through the interaction are presented in Table 2.4.

Table 2.4: Graphic Literacy Score or Performance Levels

Cognitive skill levels	Graphic complexity			
	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 task. Performance levels are based on the concept that it is more difficult to apply the same skill to a graphic of higher complexity. For example, if a 1-step process is applied to a graphic of low-moderate complexity, the performance level overall is defined as Level 3. However, if that same 1-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 the performance level increases. As a result, an individual applying a 2-step cognitive process to a graphic of high moderate complexity results in an overall performance Level 5. To further illustrate, if a 3-step cognitive process is applied to the same high moderate complexity graphic, the performance level increases to Level 6.

2.5 Graphic Literacy—Performance Level Descriptors

The Graphic Literacy construct is defined through Tables 2.1, 2.2, 2.3, and 2.4 which 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 Performance Level Descriptors.

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

2.6 Designing Items to Elicit Examinee Evidence of Graphic Literacy

Graphic Literacy uses multiple-choice items to measure examinees' proficiency in locating information and using information found in workplace graphical materials. The domain of graphic literacy skills measured by the assessment was defined by the design team and confirmed by external SMEs with backgrounds in business, industry, and education (see Table 2.5). 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).

2.6.1 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 materials. 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. Once an item writer has completed training and demonstrated the ability to write items, they receive materials explaining item task models.

The task models provide item writers with the following instruction: (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 or visual representations of a workplace phenomenon designed to communicate information

- Stimulus materials may contain one graphic or multiple associated graphics

- Stimulus materials should use as few words as possible; when possible, they should use pictures, arrows, diagrams, or other visual representations to communicate information

- 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 issues were identified related to the construct relevance of ancillary skills:

1. In evaluating graphic effectiveness, is the identification of biased presentations 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 interpretations 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 to answer. WorkKeys is extremely sensitive to this question due to the fact the program also includes an Applied Math 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 basic reading skills; likewise, to fully comprehend many graphs requires a 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 extent to which mathematics skills may be included to answer the Graphic Literacy items.

2.6.2 Guidelines for the Use of Mathematics Skills in Graphic Literacy Items

Graphic Literacy items may have a limited amount of numeracy skills involving basic math.

Given that the information and data in a Graphic Literacy item are presented through graphs and tables, it is acceptable to require examinees to find or interpret data from a graphic and then apply basic numeracy to solve the problem. For example, calculating or recognizing that 50% of a pie chart is twice as much as 25% of a pie chart, calculating or recognizing from a gauge that we need to increase the pressure by 10 psi to meet specifications, or calculating from a graph the volume of water associated with 1 part cleaner and 4 parts degreaser.

The following types of numeracy problem solving are allowed. (The permissible numeracy skills are basic computations using addition, subtraction, multiplication, and division with whole numbers, common fractions, or common percentages.)

- Single-digit addition of at most five or subtraction of at most two whole numbers. (2 + 5 + 3; 9–8)
- Single-digit multiplication or division of at most two whole numbers, without remainders. (2 x 4, 6 ÷ 3)
  ◦ Includes an operation to double, triple, or halve a whole number within 100. (There were twice as many apples as bananas.)
- Multiple-digit addition, subtraction, multiplication, or division of at most two whole numbers (without remainders) within 100. (35 + 7, 85–3, 12 x 3, 16 ÷ 4)
  ◦ For realism, multiple-digit addition and subtraction of dollar amounts including cents is permissible.
- Multiple-digit addition, subtraction, multiplication, or division by 10s or 100s, including percentages. (100 + 200; 10 x 100; 900–500, 70% + 10%)
- Use of fractions less than or equal to one; limit the denominator to 2, 3, 4, or 10. (1/2, 2/3, 3/4)
- Use of simple ratios, described in parts. (1 part cleaner to 4 parts water)

Graphic Literacy items cannot involve setting up equations; solving for unknown variables; adding, subtracting, multiplying, or dividing of uneven amounts such as decimals, fractions, or ratios; or use of advanced operations/calculations. Calculators are not allowed for the Graphic Literacy test.

Examples of problem-solving not allowed would include calculating area based on a diagram, performing a math operation using data presented in a spreadsheet format, or determining the average of a set of numbers. These problems involve equations and/or more than one operation.

2.6.3 Item Review Process

After items have been developed, edited, and tentatively finalized by the Content Assessment team, they are submitted to external consultants with backgrounds in workplace graphical materials for review. They review the item in terms of

> • the content, including concerns about whether the item is appropriately aligned to the construct;

> • whether the context and the solution method are workplace relevant; and

> • whether there is one and only one correct response.

The reviewer is also required to evaluate the item and the stimulus 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 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 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 after the content specialists conclude that all relevant issues are resolved.

2.6.4 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 the required statistical criteria. ACT analyzes the items using both classical and item response theory (IRT) statistics to evaluate the psychometric properties. Items must meet criteria based on overall difficulty and discrimination. If the pretest item meets the statistical criteria, it has passed pretesting. If it fails to meet the criteria, the Graphics Literacy content team reviews it and considers whether it should be edited, modified, or removed from the pool. When items are edited or modified, the item 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. 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. (For detailed information on the evaluation of items for fairness, please refer to Chapter 12.)

Note

[1] Thirteen external SMEs reviewed the Graphic Literacy test development documentation and provided feedback. The SMEs were provided definitions of graphic complexity, cognitive skill domains and subdomains, sample items, and related questions. The SMEs reviewed the information and then participated in small group, two-hour interviews (between three and four SMEs participated in each interview). Following the interviews, the SMEs were asked to return comments to ACT. Based on this feedback, the design team made modifications to all related materials. With these changes, the design team prepared a second draft of the documentation. SMEs reviewed the documentation, participated in a small group, two-hour interview, and returned comments. With these comments, the design team moved forward to provide a preliminary blueprint and first form of the assessment. The individuals who served as external SMEs are provided in the table below along with their affiliations.

Table 2.5: Graphic Literacy—External Subject Matter Experts

Name	Institution	Qualifications
Beverly Deal	S.B. Phillips	Workforce Readiness Director
Ana Gilbertson	Kirkwood Community College	Advanced Manufacturing Department Coordinator
Les Harrison	Retired	ACT Job Profiler; Industrial Engineer
Julia Holdridge	Sedgwick Industries	Director, Colleague Resources
Randy Lane	Eastman Chemical	Supervisor and ACT Job Profiler
Chris Manheim	Manheim Solutions	President and ACT Job Profiler
Angela Mosley	Kirkwood Community College	Career Development Coordinator
Scott Oppler	Society for Human Resource Management (SHRM) – VP of Psychometric and Test Development	Psychometrician; developed multiple assessments for certification and licensing programs
Wayne Rollins	Mid-East Commission of North Carolina	ACT Job Profiler; community college vocational-technical advisor
Priti Shah	University of Michigan	Professor of Cognition and Cognitive Neuroscience and Educational Psychology research and publications on graphic literacy
Andrew Stull	University of California Santa Barbara	Scientist studying the cognitive and perceptual effects of concrete and virtual reality manipulatives
Charles Wayne	State of Pennsylvania Department of Education	State Assessment Programs; former middle school and high school math instructor
Eric Vincent	VIO Consulting (Independent Consultant)	Former ACT employee in I/O Psychology; currently working as independent consultant to business and industry in Phoenix area

Chapter 3

Test Specifications

3.1 WorkKeys Graphic Literacy Specifications—Overview

The purpose of the WorkKeys assessment program is to assist workers, students, employers, and workforce development leaders by providing a system to measure and improve individuals' skills. Chapter 1 of the Technical Manual provided evidence demonstrating that the ability to comprehend, interpret, apply, and construct information conveyed through graphics was a foundational skill required for success in the modern economy.

In this chapter, the Graphic Literacy test specifications are provided. An assessment's test specifications are 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 in support of 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 should constitute a measure of 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 et al., 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 Subject Matter Experts (SMEs) to obtain their perspective on workplace graphics and related skills. (See the list of participating SMEs in the Chapter 2 Note.)

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.

3.2 Content Relevance and Representativeness

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 et al., 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.

A second purpose of the test specifications involves the development of alternate forms. The size of the WorkKeys testing population combined with the need for security and fairness necessitates the construction of alternate forms of Graphic Literacy. In developing alternate forms, ACT believes that all forms must meet Lord's (1980) equity property. Lord's equity property states, from the test taker's perspective, it must be a matter of score indifference whether he or she is administered Form A or Form B. To achieve alternate forms that meet the equity property, the content representativeness of each form must be identical (Kolen & Brennan, 2014).

As a result, by carefully defining the test specifications, ACT accomplishes two critical assessment goals:

1. Content is construct relevant and representative.
2. Content representation is identical across alternate forms.

3.3 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 Chapter 2: Test Development). 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. Through analyzing the professional literature on workplace graphic literacy and data from ACT's job profiling, ACT learned that graphics are used to communicate information, interpret trends and patterns, and make decisions; at higher job levels, individuals are expected to be able to develop graphics to communicate information.

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 (1F1, 1F2)	Level 3	Level 3	Level 4	Level 5
2-Step (2F1, 2F2, 2D1, 2R1,2T1)	Not Tested	Level 4	Level 5	Level 6
3-Step (3D1, 3D2, 3R1, 3R2, 3T1, 3T2)	Not Tested	Level 5	Level 6	Level 7

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

Graphic Complexity Categories	Overall Graphic Literacy Level					
	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

Graphic Complexity Categories	Total Graphic Sets	Items per Graphic	Cognitive Skill			Total Items by Graphic Complexity
			One-Step Extract Items	Two-Step Between Items	Three-step Beyond Items	
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 is built to conform 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 is parallel in terms of content to all other Graphic Literacy forms.

C h a p t e r   4

Test Administration

The *ACT WorkKeys® Administration Manual—Paper Testing* and *ACT WorkKeys® Administration Manual—Online Testing* manuals contain the instructions for administering the ACT WorkKeys assessments. Staff members associated with approved sites are responsible for the secure administration of the WorkKeys assessments.

In addition to the testing manuals, ACT WorkKeys has additional resources available online.[1] (The online resources are available through the ACT website. See the Note at the end of the chapter for the link to the online resources.)

4.1 Policies and Procedures

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.

4.1.1 Standardized 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

4.1.2 Selecting Testing Staff

Test Coordinators are responsible for selecting their testing staff. The Test Coordinator provides the continuity and administrative uniformity necessary to ensure that all examinees are tested under the same conditions, and to ensure the security of the test. Relatives and guardians of individuals taking the WorkKeys assessments are not allowed to participate in the delivery of WorkKeys assessments.

The school or organization should strive to ensure that all individuals administering the assessment are of sound ethical standing. Room supervisors and proctors may be current or retired faculty members, school administrative or clerical employees, substitute teachers, student teachers, or paraprofessionals.

The following individuals may not act as testing staff:

- High school examinees, volunteers, and lower-division undergraduates

- Anyone who intends to take ACT WorkKeys tests within the next 12 months

- Anyone involved in ACT WorkKeys test preparation activities at any time during the current testing year (September 1 through August 31), due to potential conflict of interest. (**Note:** ACT recognizes that the normal duties of a counselor or teacher may involve some responsibilities for test preparation. These activities are not a conflict of interest, provided they are part of job responsibilities specifically defined by one's employer and the employer is not a commercial enterprise.)

In addition, if any relative or ward will test at your site or any school in the state during the testing window:

- You *may not* serve as test coordinator for the administration of any of the tests. You must delegate all supervisory responsibilities—including the receipt and return of test materials—to a qualified colleague.

- You *may not* have access to the secure test materials prior to test day.

- You *may* serve as a room supervisor or proctor, provided that the examinee is not assigned to test in a room where you are working. You must not have access to the examinee's answer document or test materials.

- Relatives and wards include children, stepchildren, grandchildren, nieces, nephews, siblings, in-laws, spouses, and persons under your guardianship.

Scores for an examinee will be cancelled if any of these policies are violated

4.2 Test Administration Personnel and their Responsibilities

4.2.1 Test Coordinator

The Test Coordinator ensures that examinees test under the same conditions as examinees at every other site. The Test Coordinator can serve at only one test site.

Table 4.1: Responsibilities of the Test Coordinator

Category	Responsibility
Facilities and Staffing	• Selecting and reserving test rooms and preparing them for test day according to ACT guidelines • Selecting and training qualified testing staff
Before Testing	• Reading the testing manuals and ensuring compliance with its policies and procedures • Viewing and participating in training provided by ACT • Ordering standard time materials for the initial test date • Ordering alternate testing formats for examinees needing accommodations • Receiving, checking-in, and securely storing test materials • Arranging for the application of barcode labels on the answer documents by testing staff if required • Arranging for examinees to complete the non-test portions of their answer documents • Preparing rosters and organizing test materials • Notifying examinees of the test date(s), location, and materials needed
On Test Day	• Conducting a briefing session for testing staff • Counting and distributing test materials to staff • Ensuring that testing begins at the same time in all rooms • Supervising and assisting staff during testing • Arranging for transfer of test responses to answer documents for examinees approved by ACT for alternate response modes, or approved locally to mark answers in the test booklet • Serving as room supervisor as needed

4.2.2 Back-up Test Coordinator

The Test Coordinator should have a qualified Back-up Test Coordinator available if the Test Coordinator becomes ill or is otherwise unable to be present on test day. The Back-up Test Coordinator is encouraged to assist the Test Coordinator prior to, during, and after testing.

He or she is also expected to participate in training conducted by ACT (if previously untrained by ACT) prior to the test date. The Back-up Test Coordinator can serve at only one test site. If the Test

Coordinator is not able to supervise the administration, the Back-up Test Coordinator must complete and submit a profile change form online by going to the web address listed on your *Checklist of Dates*.

4.2.3 Test Accommodations Coordinator

The Test Coordinator must name a qualified Test Accommodations Coordinator. The Test Accommodations Coordinator is responsible for the following:

- Assisting the Test Coordinator in his or her responsibilities as needed
- Reading the testing manuals and complying with its policies and procedures
- Evaluating and approving requests for ACT WorkKeys accommodations
- Notifying the Test Coordinator of any examinees needing alternate format test materials from ACT
- Viewing and participating in accommodations training provided by ACT
- If the Test Accommodations Coordinator is no longer able to serve in his or her role, the Test Coordinator must contact ACT at 800.553.6244, ext. 1788, to designate a replacement

4.2.4 Room Supervisor

Each room is required to have a Room Supervisor who must serve for the entire session. The Test Coordinator or Test Accommodations Coordinator may serve as room supervisor if only one room is used.

Specific responsibilities include:

- Reading the testing manuals and complying with the policies and procedures it describes
- Attending both the training and briefing sessions conducted locally by the Test Coordinator
- Being responsible for the test room and providing an environment conducive to testing
- Checking ID or personally recognizing and admitting examinees
- Marking attendance/ID on the roster
- Directing examinees to seats
- Counting test booklets upon receipt from the Test Coordinator
- Distributing test materials and keeping test booklets in sequential serial number order
- Reading verbal instructions to examinees exactly as they are written
- Properly timing tests and recording the start, 5-minutes-remaining, and stop times in the manual using two timepieces
- Completing all information on the Seating Diagram and Test Booklet Count Form as found in the *Administration Manual for Paper and Pencil Testing*.
- Being attentive to examinees and materials at all times (Proctor may assist with this activity)

- Walking around the test room during testing to be sure examinees are working on the correct sections of the test booklet and answer document (Proctor may assist with this activity)

- Paying strict attention to monitoring examinees during the entire test session to detect and discourage prohibited behavior (Proctor may assist with this activity)

- Collecting and accounting for all answer documents and test booklets before dismissing examinees (Proctor may assist with this activity)

- Completing detailed documentation of any irregularities and, as required, voiding examinees' tests

- Returning all test materials and forms to the Test Coordinator immediately after testing

4.2.5 Proctor

A Proctor may be used to assist a Room Supervisor or the Test Coordinator if fewer than 10 examinees are testing. A Proctor *is required* (in addition to the Room Supervisor) for every 10 examinees (or portion thereof) after the first 10 in the room. For example, if there are 30 examinees, three proctors are required.

The Proctor's responsibilities include:

- Reading the testing manuals and complying with the policies and procedures it describes

- Attending both the training and briefing sessions conducted locally by the Test Coordinator

- Helping admit examinees and marking attendance/ID on the roster

- Directing examinees to seats

- Helping distribute test materials and keeping test booklets in sequential serial number order

- Verifying the timing of the tests using a different timepiece than the room supervisor

- Being attentive to examinees and materials at all times

- Walking around the room during testing to replace defective materials, to be sure all examinees are working on the correct test, and to observe examinee behavior

- Reporting any irregularities to the room supervisor immediately

- Accompanying examinees to the restroom if more than one is allowed to leave during the timed tests

- Paying strict attention to monitoring examinees during the entire test session to discourage and detect prohibited behavior

- Helping collect and account for all answer documents and test booklets

4.3 Training Testing Staff

For testing to occur successfully, staff members must understand their responsibilities. It is critical that the standardized test administration procedures are followed by every test center.

4.3.1 Training Session

Test Coordinators are required to hold a training session **before** test day to prepare staff for test day activities and to stimulate discussion. In addition, on each test day morning, Test Coordinators are required to hold a briefing session to discuss any last-minute issues that arise as well as concerns staff members may have.

4.3.2 *Administration Manual*

ACT provides the *Administration Manual*, which every staff member is expected to read and communicate its expectations. 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.

Each testing staff member is to be provided with a complete copy of this manual before the training session. It is especially important that Room Supervisors read and understand the policies, procedures, and directions.

4.4 Test Administration Room Requirements

Test administration rooms must be set up according to the requirements defined below. If these requirements are not met, scores may be cancelled.

- **All examinees in the test room must face the same direction,** regardless of the number of examinees in the room or the distance between them.
- There must be **at least three feet of space between examinees** (side-to-side measured shoulder-to-shoulder, and front-to-back measured head-to-head).
- In a room with multiple-level seating, examinees must be **at least five feet apart** front-to-back.
- There must be sufficient aisle space for staff to get to every seat during testing without disturbing examinees.
- Seat examinees in straight rows and columns, directly in line with each other.
- If a clock is in the room, seat examinees facing the clock whenever possible so they can see it without looking around.
- The room supervisor must be stationed in the room facing the examinees. Staff must be able to see every examinee clearly. Seating with dividers or partitions, such as study carrels, partitioned tables, or booths, is not acceptable because it obstructs staff's view of examinees.

Note

[1] ACT WorkKeys provides test administrators multiple support materials. The support materials can be found at http://www.act.org/content/act/en/products-and-services/workforce-solutions/act-workkeys/administer.html#techspecs.

C h a p t e r   5

Accessibility

The ACT WorkKeys® 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 ACT WorkKeys Graphic Literacy Assessment Support System

ACT has established for the Graphic Literacy assessment a continuum of supports for effective communication that spans 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. To build an assessment system that meets the needs of all populations tested and provides a fair communication and performance pathway for all learners, more than one level of support is needed.

"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 all test takers have, regardless of whether or not they 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) into its conceptual structure.

In September 2015, in anticipation of the development of this assessment, a week-long accessibility test development workshop was held with leadership and content developers of ACT WorkKeys National Career Readiness Certificate (NCRC®) Assessments. The topic of this workshop focused on methods of mapping the characteristics and accessibility needs of learner populations to the content models intended to be measured by the ACT WorkKeys 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 mapping process presented in Figure 5.1 provides an evidence-based structure to determine accessible communication and performance pathways as well as accessibility support options to be allowed for the ACT WorkKeys NCRC assessments.
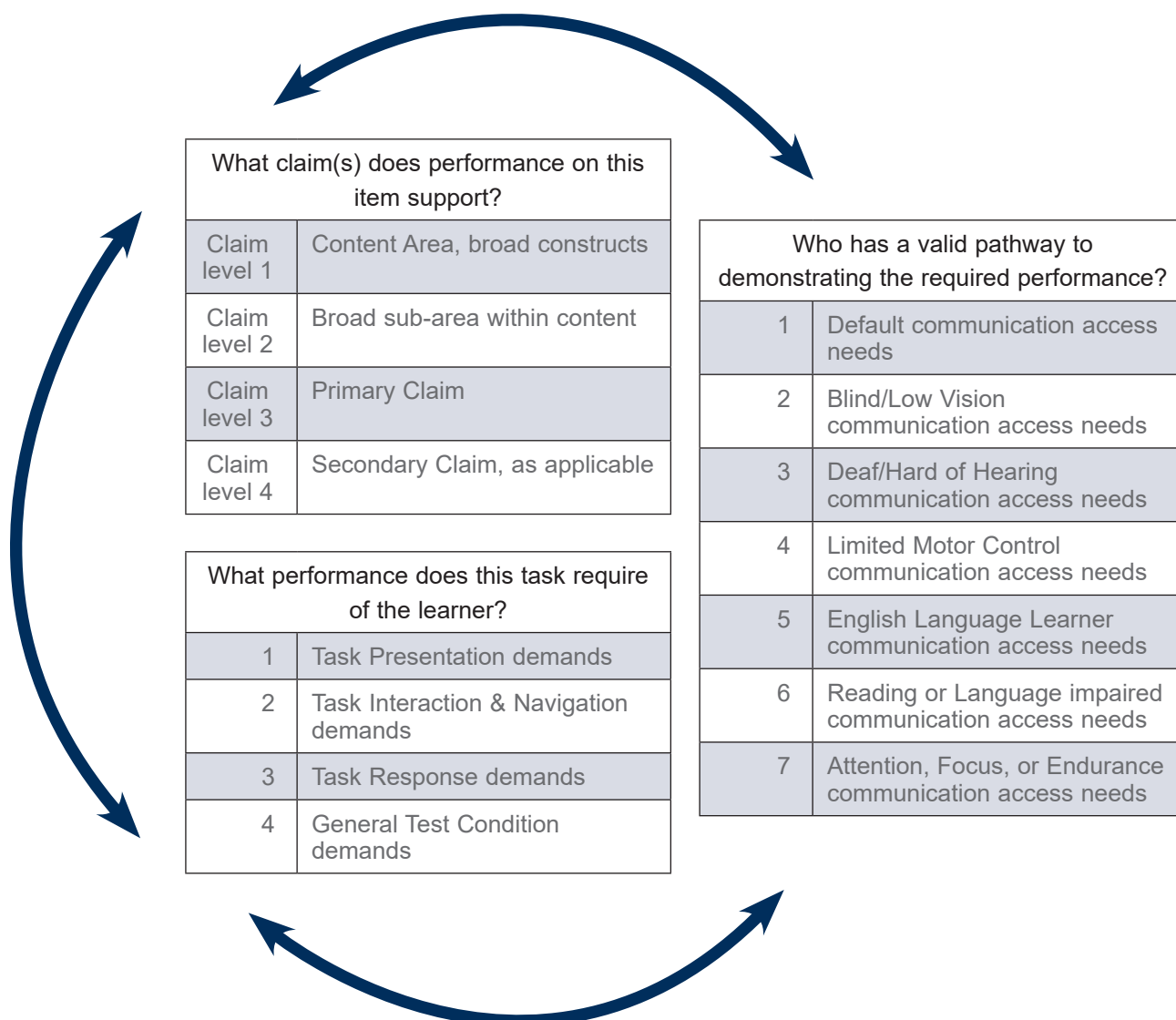
Figure 5.1: Accessibility Feature Mapping Process

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: 1) Embedded Tools, 2) Open Access Tools, and 3) 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. These are outlined later in this chapter. 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

Educational researchers and practitioners have learned over the last decade 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 any other examinee;

- 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 5.2 shows the architectural structure of ACT WorkKeys assessments accessibility supports.

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 Graphic Literacy 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.
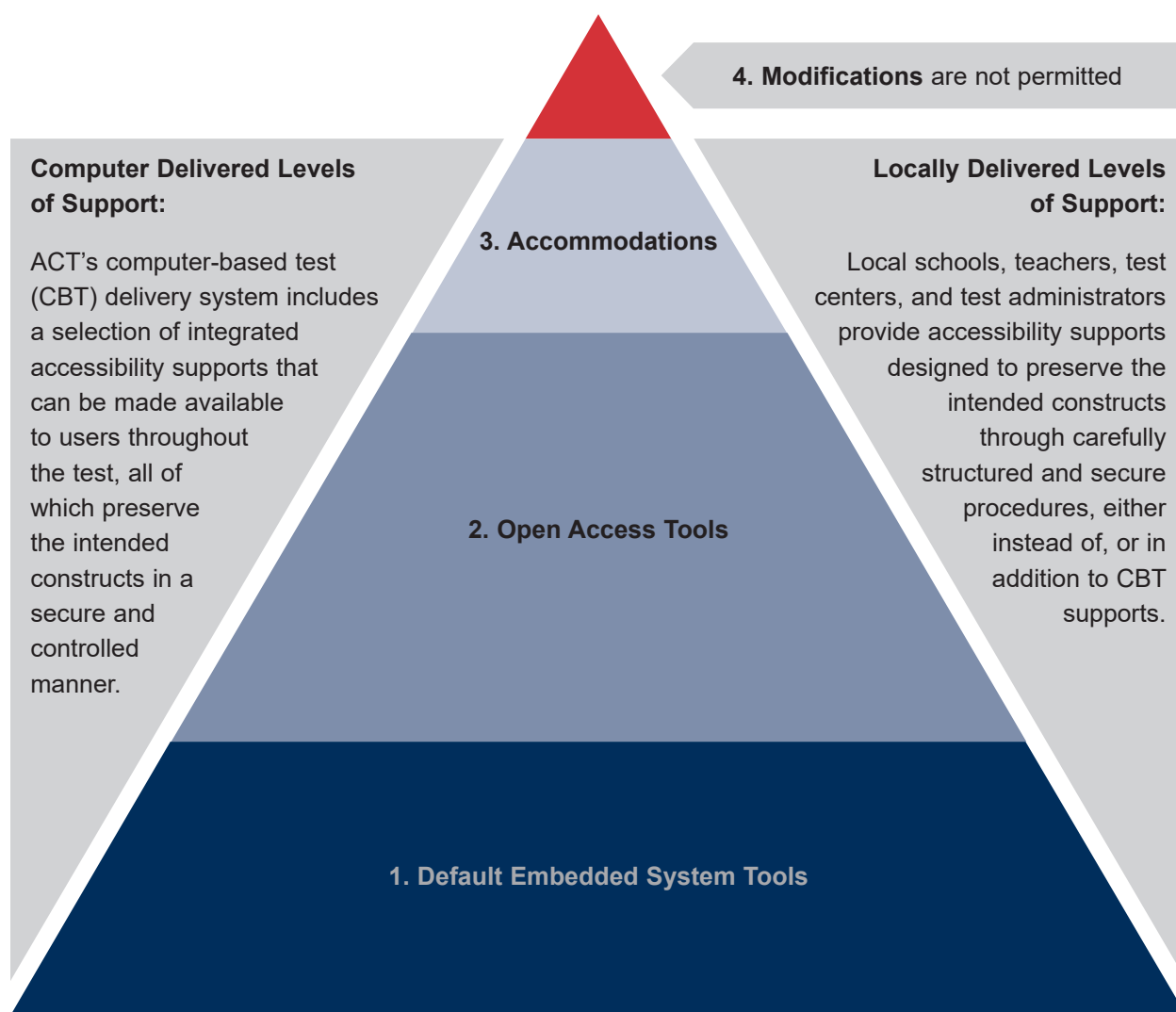
Figure 5.2: Architectural Structure of Accessibility Supports

Note. Width of the triangle above shows the proportion of examinees who use that set of assessment tools.

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 5.2.) They 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 they are seldom thought of in this way. These tools are either embedded in the basic computer test delivery platform, or they may be locally provided as needed. No advance request is needed for these supports.

Support Level 2: Open Access Tools

Open Access tools (see the second level of the pyramid in Figure 5.2) 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 – locally provided (paper)

- 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 5.2.) 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 use often requires 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, 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. ACT WorkKeys NCRC Assessments 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 ACT WorkKeys NCRC Assessments 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 5.2.) 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. For this reason, we have provided the general description of ACT WorkKeys NCRC Assessments Accessibility Supports here in one section. Full procedural requirements and instructions for using permitted supports during test administration are provided in the ACT WorkKeys NCRC Assessments Administration Manual.

5.4 Valid Test Scores and Equal Benefit for All Examinees

ACT aims to ensure that all examinees may benefit equally from the WorkKeys 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.

Table 5.1: NCRC Accessibility Supports Permissible by Assessment—Paper and Online Testing

Paper Testing

Presentation Supports	Support Level	Graphic Literacy
Test Directions Available on Demand (Printed)	Embedded	Yes
Magnifier Tool	Embedded	Yes
Full Page Magnification	Embedded	Yes
Line Reader	Open Access	Yes
Color Contrast (Color Overlays)	Open Access	Yes
Large Print Test Booklet, Printed	Open Access	Yes
Translated Verbal: Directions only (locally provided)	Open Access	Yes
American Sign Language (ASL) Directions Only	Open Access	Yes
Signed Exact English (SEE): Directions Only	Open Access	Yes
Signed Exact English (SEE): Test Items	Accommodation	Yes
Cued Speech	Accommodation	Yes
English Audio DVD (designed for user with blindness)	Accommodation	Yes
English Audio Reader Script (designed for user with blindness)	Accommodation	Yes
Word-to-Word Bilingual Dictionary, ACT approved	Accommodation	Yes
Braille EBAE, contracted, includes tactile graphics	Accommodation	Yes
Braille UEB with Nemeth, contracted, includes tactile graphics	Accommodation	Yes

Paper Testing (continued)

Interaction & Navigation Supports	Support Level	Graphic Literacy
Answer Eliminator	Embedded	Yes
Highlighter (Online-use highlighter tool. In paper form, for this purpose, a standard pencil must be used instead of a pen)	Embedded	Yes
Scratch Paper	Embedded	Yes
Calculator (includes accessible) See: http://www.act.org/content/dam/act/unsecured/documents/ACT-calculator-policy.pdf for allowable devices	Embedded	Yes
Answer Masking	Open Access	Yes
Custom Masking	Open Access	Yes
Abacus	Accommodation	Yes

Response Supports	Support Level	Graphic Literacy
Mark Item for Later Review (requires examinee mark to be erased thoroughly)	Embedded	Yes
Dictate Responses	Open Access	Yes
Respond in Test Booklet or on separate paper	Open Access	Yes
Accessible Keyboard or AAC Device, with local print-out	Accommodation	Yes
Large Print Answer Document	Accommodation	Yes

General Test Conditions	Support Level	Graphic Literacy
Proctor ability to add Extra Time (in event of test administration incident)	Embedded	Yes
Audio Indicator of Time Remaining	Open Access	Yes
Audio Indicator: 5-minute Warning	Open Access	Yes
Break: Supervised within each day (stop the clock)	Open Access	Yes
Individual Administration (not home)	Open Access	Yes
Location for Movement	Open Access	Yes
Other Setting (not home)	Open Access	Yes
Physical/Motor Equipment	Open Access	Yes
Special Seating/Grouping	Open Access	Yes
Visual Environment	Open Access	Yes
Audio – Acoustic Environment	Open Access	Yes
Extra Time **(Time-and-a-half, Double Time, or Three Hours)**	Accommodation	Yes
Break: Securely extend session over multiple days	Accommodation	Yes

Computer Testing

Presentation Support	Support Level	Graphic Literacy
Test Directions Available on Demand (on screen)	Embedded	Yes
Magnifier Tool	Embedded	Yes
Browser Zoom Magnification (full page)	Embedded	Yes
American Sign Language (ASL) Directions Only	Open Access	Yes
Line Reader	Open Access	Yes
Color Contrast (High/Low Contrast Colors)	Open Access	Yes
Translated Audio: Directions Only	Open Access	Yes
Signed Exact English (SEE): Directions Only	Open Access	Yes
Cued Speech	Accommodation	Yes
Word-to-Word Bilingual Dictionary, ACT Approved	Accommodation	Yes

Interaction & Navigation Support	Support Level	Graphic Literacy
Answer Eliminator	Embedded	Yes
Highlighter	Embedded	Yes
Keyboard Navigation	Embedded	Yes
Scratch Paper	Embedded	Yes
Calculator	Embedded	No
Answer Masking	Open Access	Yes
Custom Masking	Open Access	Yes
Abacus	Accommodation	No

Response Support	Support Level	Graphic Literacy
Mark Item for Review	Embedded	Yes
Dictate Responses	Open Access	Yes
Respond on Separate Paper	Open Access	Yes
Electronic Spell Checker	Accommodation	Yes
Accessible Keyboard or AAC device, with local print-out	Accommodation	Yes

Computer Testing (continued)

General Test Conditions	Support Level	Graphic Literacy
Proctor ability to add Extra Time (in event of test administration incident)	Embedded	Yes
Audio Indicator of Time Remaining	Open Access	Yes
Audio Indicator: 5-minute Warning	Open Access	Yes
Break: Supervised within each day (stop the clock)	Open Access	Yes
Individual Administration (not home)	Open Access	Yes
Location for Movement	Open Access	Yes
Other Setting (not home)	Open Access	Yes
Physical/Motor Equipment	Open Access	Yes
Special Seating/Grouping	Open Access	Yes
Audio—Acoustic Environment	Open Access	Yes
Visual Environment	Open Access	Yes
Extra Time **(Time-and-a-half, Double Time, or Three Hours)**	Accommodation	Yes

C h a p t e r   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 ACT 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 ACT WorkKeys.

Testing staff must be alert to activities that can compromise the fairness of the test and the validity of the scores. Such activities include, but are not limited to, cheating and questionable test taking behavior (such as copying answers or using prohibited electronic devices during testing); accessing questions prior to the test; taking photos or making copies of test questions or test materials; posting test questions on the internet; or test proctor or test administrator misconduct (such as providing answers or questions to test takers or permitting test takers to engage in prohibited conduct during testing).

In addition to these security-related administration protocols, ACT engages in additional test security practices designed to protect the WorkKeys assessment and the validity of its scores. These practices include: (1) use of a reporting hotline through which individuals with information about misconduct on an ACT WorkKeys test can anonymously report such information to ACT; (2) data forensics in support

of ACT WorkKeys related investigations; and (3) web monitoring to detect testing misconduct, possible unauthorized disclosure of secure ACT WorkKeys test content, and other activity that might compromise the security of the ACT WorkKeys test or the validity of its scores.

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. The categories covered by the framework and brief statements of their importance to ACT are:

1. Information Security Program Management: This is overseen by the Information Security Officer at ACT. The Information Security Officer has responsibility for providing guidance and direction to the organization to ensure compliance with all relevant security-related regulations and requirements. The program itself is designed to cover all security domains identified in the ISO 27001 standards and provides comprehensive oversight for Information Security at ACT.

2. Information Security Risk Management: The cornerstone of the ACT Information Security program is a risk assessment that conforms to the ISO 27005 standard. The identification, management, and mitigation of information security risks are managed using the ISMS (Information Security Management System) guidelines defined in the 27005 standard. ACT also makes use of the SP NIST 800-37 Risk Assessment which complies with FISMA security requirements for risk management (National Institute of Standards and Technology, 2017).

3. Information Security Policies and Standards: ACT established an Information Security policy to set direction and emphasize the importance of safeguarding information and data assets. Additional supporting policies, standards, and procedures have been developed to communicate requirements.

   ACT's Information Security Policy and the Assessment Data Sharing procedures govern the handling of examinee data that is classified as confidential restricted. The policy states that confidential restricted information must meet the following guidelines:

   - Electronic information assets must only be stored on ACT-approved systems/media with appropriate access controls.
   - Only limited authorized users may have access to this information.
   - Physical records must be locked in drawers or cabinets while not being used.
     - ACT also has Access Management, Business Continuity Standard, Clear Desk/Clear Screen, End User Storage, External Authentication, Information Security Incident Management, Malware Protection, Mobile Device, Network Security Management, Payment Card Security, Secure Application Development, Secure System Configuration, Security Event Logging and Monitoring Standard, System Vulnerability and Patch Management and Web Content Standard to form a system of control to protect examinee data.

4. Information and Technology Compliance: The systems that store, maintain, and process information are designed to protect data security through all lifecycle stages. The security considerations surrounding ACT's systems include measures such as encryption, system security requirements, and logging and monitoring to verify systems are operating within expected parameters.

5. Business Continuity and Disaster Recovery: ACT maintains a Business Continuity program designed to provide assurance that critical business operations will be maintained in the event of a disruption. An essential part of the program includes a cycle of planning, testing, and updating. Disaster Recovery activities are prioritized by the criticality of systems and recovery times established by the business owners.

6. Security Training and Awareness: At ACT, Information Security is everyone's responsibility. All employees take part in annual Information Security awareness training on topics covered in the Information Security policy. Additionally, ACT has individuals within the organization who are responsible for the management, coordination, and implementation of specific Information Security objectives and who receive additional Information Security Training.

7. Identity and Access Management: ACT addresses data integrity and confidentiality by implementing policies and procedures that limit access to individuals who have a business need to know the information and that verify the individual's identity. Access to ACT systems and data requires authorization from the appropriate system owner. Active Directory, file permissions, and VPN (Virtual Private Network) remote access is administered by an Identity and Access management team who are part of the Information Security organization.

8. Information Security Monitoring: The foundation of ACT's Information Security Program is reflected in the Information Security Policy which is presented and reinforced with training to all ACT employees. ACT is held accountable to following the Information Security Program through internal assessments of the security control environment. Additionally, ACT works with independent third-parties to provide assessment feedback.

9. Vulnerability and Threat Management: ACT has several mechanisms in place to identify vulnerabilities on networks, servers, and desktops. Monthly vulnerability scanning is performed by a qualified ASV (Approved Scanning Vendor). ACT has always maintained a "compliant" status in accordance with PCI-DSS (Payment Card Industry Data Security Standards) requirements. In addition to the scans performed for PCI compliance, ACT has a suite of vulnerability scanning tools which are coordinated with a log management and event monitoring tool to provide reporting and alerting.

10. Boundary Defense: ACT utilizes multiple intrusion protection and detection strategies, tools, processes, and devices to look for unusual attack mechanisms and detect any kind of compromise of these systems. Network-based IDS sensors are deployed on Internet and extranet DMZ systems and networks which provide alerting and procedures for review and response. Procedures include security review and approval of changes to configurations and semi-annual firewall rule review and restrictions to deny communications with, or limit data flow to known malicious IP addresses.

11. Endpoint Defenses: A variety of tools are utilized to ensure that a secure environment is maintained at the end-user device level. This includes segmentation within the ACT network,

anti-virus programs, and data-loss prevention programs. VPN is required for all remote access to the ACT network. Wireless access on the ACT campus requires authentication credentials and continuous scanning for rogue access points is performed.

12. Physical Security: Maintaining security on the premises where information assets reside is often considered the first line of defense in Information Security. ACT has implemented several security measures to ensure physical locations and equipment used to house data are protected, including card-key access to all facilities and camera monitoring at all entry points.

13. Security Incident Response and Forensics: Planning for how to handle information security incidents is a critical component of ACT's Information Security program. Formal policy guidance outlines response procedures, notification protocols, and escalation procedures. Forensic investigations are performed at the direction of the Information Security Officer. ACT maintains a subscription service with a third-party specializing in computer forensics in the event of a declared incident.

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.

This 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.

Chapter 7

Reporting

7.1 Graphic Literacy Reports

ACT WorkKeys® Graphic Literacy reports are designed to provide detailed information to examinees, test administration officials, employers, workforce development officials, and educators. With the updated assessments and systems, the WorkKeys Online Reports Portal (WKRP) has been designed to provide real-time electronic information to test users. This information is available through the portal whether an examinee takes an assessment online or on paper.

The objectives of the Graphic Literacy reports are:

- To clearly communicate to examinees, employers, educators, and workforce development officials the skills demonstrated by examinees
- To provide examinees with insights on their current skill levels and how they might improve
- To provide employers and educators actionable information to assist in decision making
- To provide workforce development officials and educators insights needed to improve examinee performance
- To provide information that connects skill levels to worker success
- To leverage technology to make the reporting user experience faster and more effective through the use of the WKRP

The Graphic Literacy assessment is a criterion-referenced test. A criterion-referenced test differs from a norm-referenced test in that scores are interpreted based on the skills demonstrated through testing. The Graphic Literacy Performance Level Descriptors (PLDs) provide a detailed summary of the skills demonstrated by the examinee at each score level. (See Chapter 2 for the complete Graphic Literacy PLDs.)

For the person who takes the assessment, performance is summarized through the Individual Examinee Score Report. For each WorkKeys assessment that a person takes, a separate Individual Examinee Score Report is generated. It provides the following information:

- ACT WorkKeys Realm Name
- Test Date
- Report Date
- Examinee's name
- Examinee's ID
- Assessment Title
- Scale Score (including possible scale score range)
- Level Score (including possible level score range)
- What your score means—a section that includes the PLD for the specified Level Score
- How you can use your scores—a statement that directs the examinee to a WorkKeys URL where additional score interpretation information is found

In addition to the Individual Examinee Score Report, ACT provides other reports that are available to either examinees or institutions. Table 7.1 presents the list of available Graphic Literacy reports.

Table 7.1: Graphic Literacy Reports and Their Function

Report	Function
Individual Examinee Score Report	This report provides information to the examinee about his or her score and what it means to be at a specified skill level.
Individual Summary Score Report	This report provides information to the examinee about his or her scores and skill levels for all tests taken online.
Roster Score Report	This report is a list of all examinees, the tests taken, and the scores examinees received.
Data Export Report	This report exports data from the Validus system into an Excel file format. It provides all of the information about the examinee including demographics, date tested, test titles, and scores.
Individual Score Reports (by Group)	This report provides information to the examinee about their score and what it means to be at that skill level. This report is run for all examinees in the selected group.
Individual Score vs. Profile Report	This report is used to show a comparison of a required skill level with the skill level the examinee achieved. For example, a company may want this report if they are hiring for a job that has been job profiled and they know the level required for a specific skill area. This report will print with the skill level required and the skill level of the applicant.
Group vs. Profile Report	This report displays the scores that a group of examinees achieved compared to a score that is required for a job. For example, a company may want this report if they are hiring for a job that has been job profiled and they know the level required for a specific skill area. This report will print with the skill level required and the skill level of all applicants in the group.
Registered to Test Report	This report provides a list of examinees registered for tests who have not yet tested. Proctors of a realm who are not administrators of that realm will be able to run the Registered to Test Report.
Test Usage Report	This report provides a count of the tests launched at the site for a given test date range.

Chapters 8–11 of the Technical Manual describe in detail Graphic Literacy assessment scores, metrics, and interpretations.

C h a p t e r   8

Scores and Score Scales

8.1 Overview

This chapter describes the rationales, procedures, and outcomes for scoring the WorkKeys® Graphic Literacy items, establishing scale scores, and defining level scores for the assessment.

Raw and scale scores are two types of scores used to facilitate score interpretation and use. *The Standards for Educational and Psychological Testing* (referred to as *the Standards* below) defines a raw score as "a score on a test that is calculated by counting the number of correct answers, or more generally, a sum or other combination of item scores" (AERA et al., 2014, p. 222). Raw scores are frequently transformed to scale scores to facilitate and standardize score interpretations. To produce scale scores for a new assessment, a scaling analysis is required; that is, "the process of creating a scale or a scale score to enhance test score interpretation by placing scores from different tests or test forms on a common scale or by producing scale scores designed to support score interpretations" (AERA et al., 2014, p. 223). For the Graphic Literacy assessment, an Item Response Theory (IRT) approach with arcsine transformation was applied to produce a scale with nearly equal conditional standard error of measurement for most score points.

Any WorkKeys foundational skill assessment, including the Graphic Literacy assessment, classifies an examinee into score levels that are aligned to the Performance Level Descriptors (PLDs). Combining the score level with the associated PLD provides the examinee and the test user with a description of the Graphic Literacy skills demonstrated by the examinee. To achieve this alignment, cut scores are established on the reported score scale to support level score interpretations. A cut score is defined as "a specified point on a score scale, such that scores at or above that point are reported, interpreted, or acted upon differently from scores below that point" (AERA, et al., 2014, p. 218). For the Graphic Literacy assessment, cut scores are established through a standard setting process drawing upon a panel of Subject Matter Experts (SMEs) to ensure the alignment of the level scores to the PLDs (AERA et al., 2014).

8.2 Selected-Response Item Scoring

All items on the Graphic Literacy assessment are selected-response items (e.g., multiple choice items). Selected-response items require examinees to select a correct answer from a set of alternative choices. For the Graphic Literacy assessment, each selected-response item has four choices or options. Each item that an examinee answers correctly provides the examinee with a score value of one raw point. An incorrect response, a missing response (items that an examinee did not answer), or multiple responses yield a value of zero raw point. The examinee's raw score is calculated by summing the correct responses.

ACT strives to write each Graphic Literacy item so that there is only one correct response. To ensure that there is only one correct response, ACT follows the process outlined in Chapter 2 that includes item writing, editing, review, and pretesting. Following these steps, an item may be selected for inclusion on a Graphic Literacy form. ACT psychometricians and content specialists regularly conduct preliminary item analysis and review the results for key validation for all the items on a form when initial form administration reaches acceptable sample size.

8.3 Scale Score and Level Score Differences and Rationale

Each item on the assessment is written to assess a specified skill level defined by the Graphic Literacy assessment construct. Graphic Literacy skills associated with each of the five levels (Levels 3 to 7) were defined through the design process described in Chapter 2. Each Graphic Literacy form is composed of the items to assess the skills defined by the level, and it is built to the test specifications described in Chapter 3. When examinees complete the Graphic Literacy assessment, they receive a report that includes the scale and level scores. The scale and level scores serve two distinct purposes in facilitating score interpretations and uses.

Scale scores provide finer grain score distinctions than level scores and they are designed to assist in analyzing growth or improvement over time, evaluating group comparisons on outcome measures, and providing evidence of benefit from educational or training programs. The scale scores, ranging from 65 to 90, are constructed such that the Standard Error of Measurement (SEM) is approximately equal at each score point (Kolen, 1988). When the SEM is the same for all scores across the distribution, ACT is able to report all test scores with the same level of precision. Doing so increases the fairness of score interpretation, and it removes the need for ACT to report the SEM at the different score points.

Level scores provide examinees with information as to whether they were able to master the defined skills associated with a specified level. The levels are defined through the PLDs. (See Chapter 2 for the PLDs associated with each level.) ACT implemented a standard setting process by which data was gathered from SMEs to enable the establishment of cut scores to identify the scale score performance required to achieve a specified level score.

8.4 Procedures for Establishing the Score Scale

A scaling study was conducted in spring 2017 as part of a series of field studies to establish the score scale for the updated WorkKeys assessments. ACT recruited examinees to participate in the field studies from various regions in the United States. The sampling plan was designed to achieve a representative sample corresponding to the WorkKeys test taking population in terms of geographic region, gender, and ethnic groups. Following data cleaning, the scaling study included a sample of 1,170 examinees.

Forty sites participated in the scaling study. It included 13 high schools and 27 adult testing centers across 22 states. For the scaling study, female examinees outnumbered male examinees by 51% to 46%. In terms of ethnicity, White examinees comprised approximately 61% of the examinees, while African-American examinees comprised 18%, and Hispanic examinees comprised 7%. ACT concluded that the sample was representative of the current WorkKeys test taking population.

The examinees took the Graphic Literacy assessment—Form G2C_S1 – in the scaling study. ACT analyzed examinee data from the scaling study applying a three-parameter logistic (3-PL) IRT model to calibrate item parameters. Figure 8.1 presents the raw score distribution from the sample. The distribution appears to be slightly left skewed, which is consistent with distributions observed from previous administrations of the Graphic Literacy assessment.
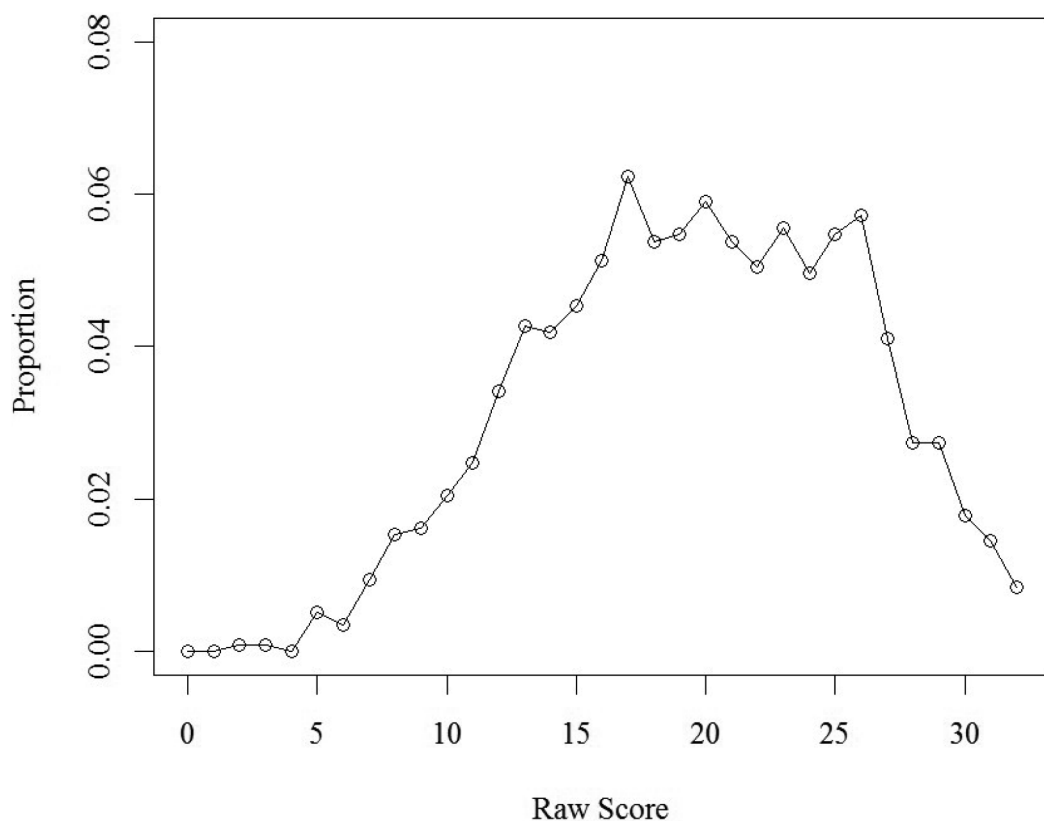
Figure 8.1: Raw Score Distribution for the GL Scaling Study Form (Form G2C_S1)

Note. Mean and standard deviation are 19.66 and 6.03 respectively.

Figure 8.2 illustrates the item p-values (ranging from 0.2 to 0.9) and *b*-parameter estimates by corresponding levels for this form, where the red dots represent the average item p-value or *b*-parameter estimate for that level. The item p-values tend to decrease as the item difficulty increases as expected. The plot on *b*-parameter estimates shows the similar trend (average *b*-parameter values increases as the level increases). Figure 8.3 shows the test characteristic curve (TCC) and test information function (TIF) for the Scaling Study form.
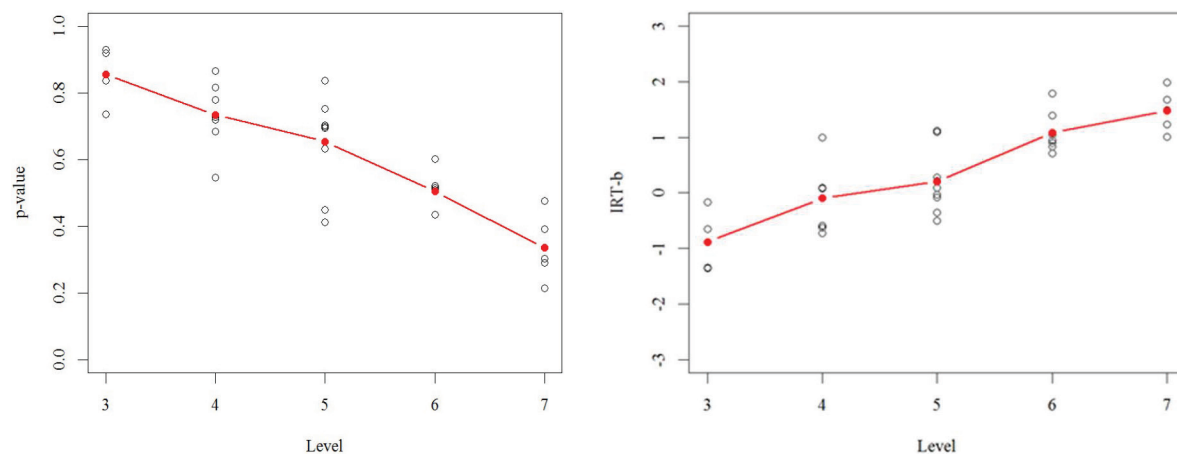
Figure 8.2: Item p-values and *b*-parameter estimates by Item Levels for Form G2C_S1
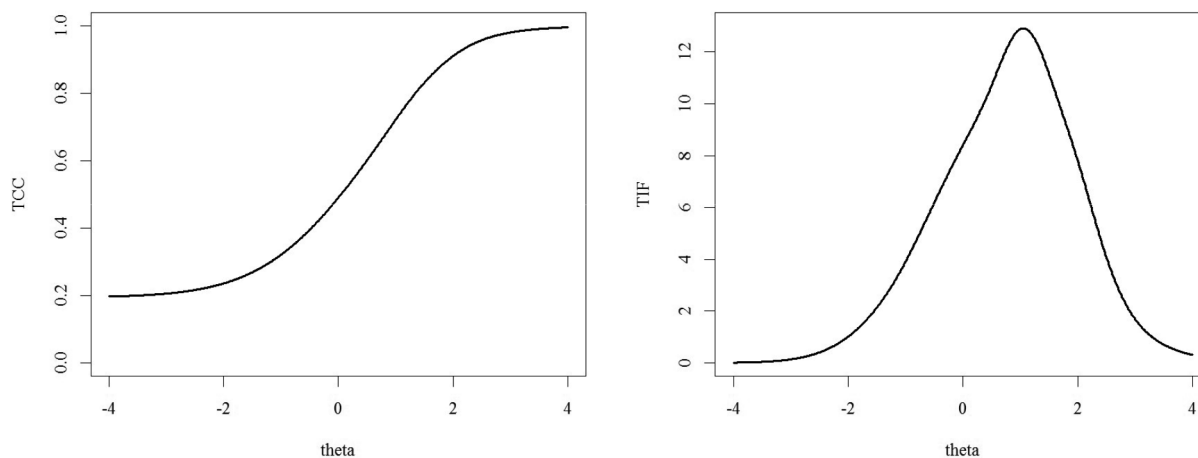
Figure 8.3: Test Characteristics Curve (left) and Test Information Function (right)

To be consistent with the Graphic Literacy assessment and the other NCRC assessments, the average scale score was set to be about 78 and the scale score score Conditional Standard Error of Measurement (CSEM) was set to less than 2. In addition, the scale score range was defined as 65 to 90, which is identical to the range of NCRC 1.0 assessment scale score. The target scale score mean and target scale score SEM are required to conduct the scaling. IRT (Ban & Lee, 2007) was used to derive the raw-to-scale score conversion, and the arcsine transformation (Kolen, 1988; Kolen & Brennan, 2014)

was used to equalize the CSEM along the score scale. The following five steps were implemented for deriving the raw-to-scale score conversion:

1. Item parameters were calibrated based on the 3-PL IRT model.

2. Theta estimates (ability estimates) for each examinee were calculated based on the item scoring vector data and the item parameter estimates calibrated in step one.

3. The expected raw score distribution was estimated based on the item parameter estimates from step one and theta estimates from step two using the Lord-Wingersky recursive formula (Lord & Wingersky, 1984).

4. Arcsine transformation was used to transform the expected raw scores to g-scores.

5. The g-scores from step four were linearly transformed to the scale scores using the target scale score mean and target scale score SEM. The slope and intercept of the linear transformation are $A = \dfrac{\sigma(E_S)}{\sigma(E_g)}$ and $B = \mu(S) - \dfrac{\sigma(E_S)}{\sigma(E_g)} \times \mu[c(\chi)]$, respectively, where $\mu(S)$ and $\sigma(E_S)$ are the target mean and SEM of the scale scores, and $\mu[c(\chi)]$ and $\sigma(E_g)$ are the mean and SEM of the g-scores.

In applying the process to create the raw to scale score transformation, the following requirements were met:

• The reported score scale covered the full range from 65 to 90.

• No more than two raw score points corresponded to one scale score, except at the two ends.

• No gaps were allowed in the score scale except at the two ends.

• Rounding error was minimized. In other words, the number of scale scores with the first decimal place of 0.5 was small.

• CSEM was as similar as possible across the score scale.

The target scale score mean and target scale score SEM were specified to be 77.9 and 1.7. These values were obtained through several explorations using the data from the scaling study and the requirements defined above.

Along with achieving the same conversions as the NCRC 1.0 assessments (e.g., same scale score range and constant CSEM), the base form conversion for the Graphic Literacy assessment included the following characteristics: (a) fewer truncated points at the lower end of the scale, (b) fewer and smaller score gaps at the higher end of the scale, and (c) defined target scale score average and CSEM.

The results indicated that the scaling procedures achieved the following goals:

• As shown in Figure 8.4, the scale score CSEMs is flat below 2.0 along the scale score except for two score ends. Note that the CSEMs of the raw scores tend to be larger in the middle and smaller at the two ends.

• The mean scale score (78) is very close to the target scale score mean (77.9) used as the input for the arcsine transformation. Table 8.1 presents the summary of the unrounded scale scores (USS) and rounded scale scores (RSS) for this form. Figure 8.5 illustrates the relative and cumulative frequency distributions of the scale scores.
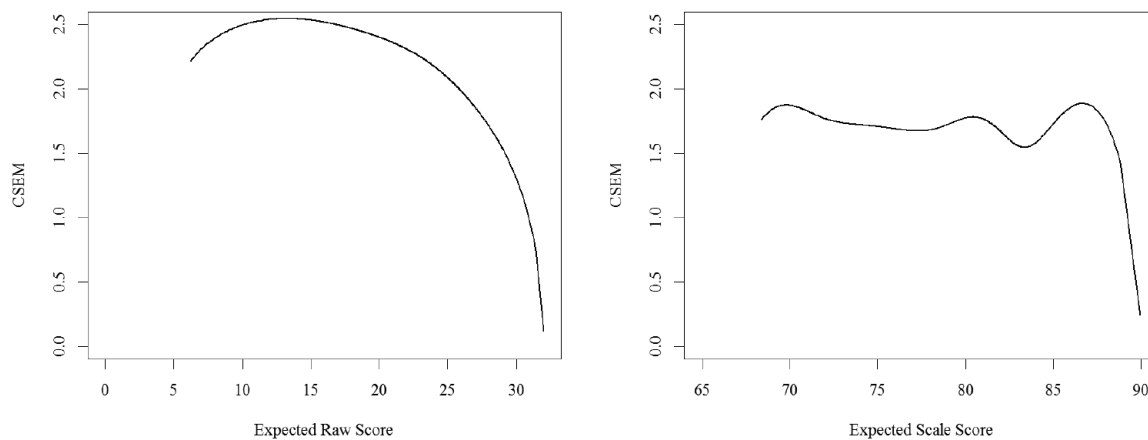
Figure 8.4: CSEM for Raw Scores (left) and Scale Scores (right)

Table 8.1: Summary of Unrounded and Rounded Scale Score

Form	Mean	SD	Min	10th	25th	50th	75th	90th	95th	Max
USS	78.10	4.45	64.37	72.77	74.77	78.10	80.95	83.39	85.33	90.63
RSS	78.02	4.45	65	73	75	78	81	83	85	90
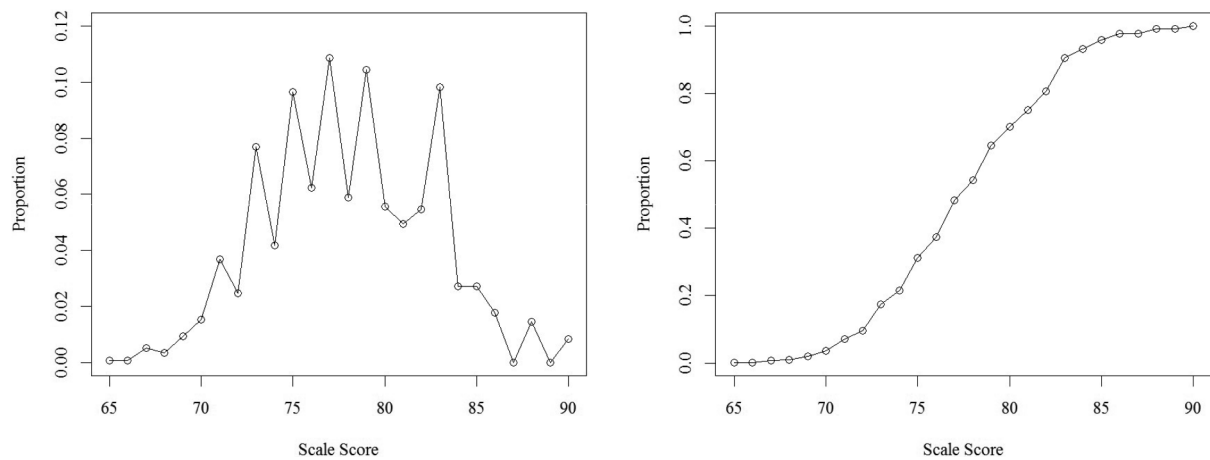
Figure 8.5: Relative Frequency Distribution (left) and Cumulative Frequency Distribution (right)

8.5 Procedures for Establishing the Level Scores

As identified above, when examinees complete the Graphic Literacy assessment, they receive a score report that includes a scale score and a level score. Following the establishment of the score scale, ACT undertook a standard setting process to establish the minimum scale scores required to achieve each of the five Graphic Literacy levels. To establish the minimum scale scores, ACT assembled a panel of SMEs consisting of educators and business people, some of whom are current WorkKeys customers. The Mapmark standard setting method (Schulz & Mitzel, 2005) with Whole Booklet Feedback was used to establish the cut scores for each of the Graphic Literacy score levels.

Mapmark builds on the popular Bookmark procedure (Lewis, Mitzel, Mercado, & Schulz, 2012). The key difference between Mapmark and Bookmark methods is the Item Map contained within the Order Item Booklet (OIB). The OIB contains a sample of items from the Graphic Literacy item pool ordered from easiest to hardest. The Mapmark process includes within the OIB the item map, which provides the difficulty of each item mapped to the actual scale value. The item map, therefore, shows "how much" more difficult one item is than another. In other words, the item map provides additional information on item difficulty.

A total of 92 items were selected to create the OIB. The IRT parameter estimates for all the items in the OIB were calibrated and scaled to the base form. All the items were ranked in order by the corresponding scale score (convert item difficulty to scale score) to form the OIB.

ACT conducted a standard setting study with a panel of SMEs (see Chapter 2 for the credentials of the panel), including appropriate training sessions. The purpose of the standard setting process was to gather data to assist ACT in establishing the standards for achieving a defined performance level on the Graphic Literacy assessment. Because the Graphic Literacy assessment is a criterion-referenced measure, reported scores on the assessment are aligned to the PLDs (see Chapter 2) that a test taker has demonstrated through responding to items on the assessment. Specifically, the purpose is to identify a cut point on the score scale per skill level where examinees who score at or above the point have demonstrated the ability to perform the skills corresponding to that skill level, and examinees who score below the point have not demonstrated the ability to perform the skills. In implementing the Mapmark procedure, ACT instructed the SMEs to define the level scores such that:

- an examinee is expected to correctly respond to at least 67% of the items that belong to his or her reported level.
- an examinee is expected to have demonstrated mastery for all levels below his or her reported level.
- an examinee is NOT expected to correctly respond to more than 67% of the items that belong to levels higher than his or her reported level.

The Mapmark standard setting included a three-round process, with Whole Booklet Feedback. For each of three rounds, the SMEs set cut scores for each level. In Round 1, the SMEs (a) took the Graphic Literacy assessment, (b) reviewed the Graphic Literacy PLDs, (c) reviewed test items and their associated scale scores, (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 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).

In Round 2, the panelists received feedback regarding their bookmark placement 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 16 test takers (two test takers in each level and one test taker between each level) answered each of the items on Form G2C_S1. Data was provided for two examinees that scored at or near the Round 1 cut score for each skill level and data for a borderline examinee at each level. The purpose was to help the panelists understand what examinees at the Round 1 cut scores "can" do and consider whether this is what examinees "should" be able to do according to the PLD for each skill level. Using all of this information, panelists were asked to repeat the process of placing bookmarks in the OIB for each level.

In Round 3, the panelists received feedback regarding their bookmark placement in Round 2. The feedback included consequences or impact data showing the percentage of examinees performing at or above the cut scores set for each skill level. ACT emphasized to the panelists that the PLDs should take precedence since the assessment is criterion-referenced. With that, they set their bookmarks for the third round.

During the final meeting, the panelists reviewed the Item Map with lines representing the Round 3 median cut scores drawn on the map. Next, they received instructions for recording the Round 3 cut scores in their OIB, and reviewed a Cut Score Distribution Chart showing the distribution of panelists' Round 3 cut scores across all the skill levels. Finally, the panelists discussed consequences data based on the final cut scores. Following these discussions, the panelists approved the final median cut score to define the five performance levels.

ACT reviewed the work of the Standard Setting panelists and 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 cut scores for the five levels were approved for the Graphic Literacy assessment. The final median cut scores will be used to define each performance level on the Graphic Literacy assessment, and the cut scores are presented in Table 8.2.

Table 8.2: Median Cut Scores for Graphic Literacy Assessment

| Levels | Final Scale Score Cut Points | | |
| | Median Cut | Range of Median Cut | |
		Min	Max
Level 3	72	71	73
Level 4	76	74	76
Level 5	78	77	78
Level 6	82	81	83
Level 7	86	85	88

With the establishment of the scale scores and cut scores, new forms will be built to be parallel based on the test specifications (see Chapter 3) and will be equated to the base form to achieve score comparability. As a result, scale scores and level scores for different forms of the Graphic Literacy assessment will be comparable (see Chapter 9).

Chapter 9

Equating and Linking

This chapter contains three sections. The first section describes the equating methods used for ACT® WorkKeys® Graphic Literacy assessment. Because multiple alternate forms of the Graphic Literacy assessment are required, ACT applies equating methods to ensure that scores from different forms are interchangeable and comparable across forms. The second section reports the findings of the mode comparability study. ACT administers the ACT® WorkKeys® NCRC® assessments in both paper and online formats. The mode comparability study was conducted to learn if scores earned by an examinee using the paper mode are interchangeable and comparable to scores earned by an examinee using the online mode. The third section presents the findings of a linking study to provide concordance scale scores between the previous version of the Locating Information and current Graphic Literacy assessments. WorkKeys test users want to understand the relationship between scores earned on the Locating Information assessment and scores earned on the Graphic Literacy assessment. Although scores earned on the Graphic Literacy assessment are not interchangeable with scores earned on the Locating Information assessment, the linking study will assist users in understanding the relationship of the current assessment to the previous assessment.

9.1 Equating Method and Procedures

New test forms for the WorkKeys Graphic Literacy assessment are developed on a regular basis to ensure the fairness and security of the test scores. Though each form is constructed to meet the same content (See chapter 3 for the detailed content blueprint) and statistical specifications, the forms may vary slightly in form difficulty. Equating is the process of making statistical adjustments to achieve score interchangeability across the forms so that the reported scale scores have the same meaning regardless of the forms administered (Kolen & Brennan, 2014). Using Item Response Theory (IRT) true-score equating, the Graphic Literacy forms are either pre-equated or post-equated to produce scale scores and level scores. Pre-equating refers to the process by which conversions from raw to scale scores are established prior to test delivery. Pre-equating enables test takers to receive their score reports in a relatively short period of time following testing. To construct a Graphic Literacy new test form, items are

selected from an item pool which meets the content classification specifications and the item statistical specifications. Test development content specialists and research psychometric specialists review the proposed form to ensure that it meets the complete test specifications. After item selection is approved and finalized, ACT applies pre-equating to derive the raw-to-scale score conversion table (see greater detail about skill level and scale scores in Chapter 8). However, if pre-equating cannot be applied due to a lack of calibrated item statistics, post-equating can be conducted following the test administrations, assuming a sufficient number of examinees have taken the assessment.

To be able to apply pre-equating to a newly developed form, all items in the form need IRT-calibrated parameter estimates that have been placed on the same scale. For the Graphic Literacy assessment, ACT is continually developing new items. When newly developed items have been reviewed and approved, they are embedded as pretest items in operational form administrations (see Chapter 8). ACT routinely conducts item calibrations using a three-parameter logistic (3-PL) IRT model. The Stocking-Lord method (Stocking & Lord, 1983) is used to place the item parameter estimates, including those for pretest items, onto the same scale. After each form calibration, the item statistics are reviewed in terms of classical test theory (CTT) and IRT. For example, items with very low discrimination indices (e.g., point biserial correlation or IRT a-parameter estimate) or extreme difficulty indices (e.g., p-value or IRT b-parameter estimate) are either archived or revised for additional pretesting. Through the process of item development, pretesting, and calibrations, new items whose content and statistical properties are reviewed and found to be acceptable, are added to the WorkKeys item pool which is continually expanded and maintained.

In addition, ACT periodically reviews the item pool for the purpose of archiving outdated or overused items. ACT also monitors the stability of item parameters to ensure that all items contained in the pool are suitable for the assembly of new test forms.

9.2 Mode Comparability

ACT developed the Graphic Literacy assessment to be administered using both paper and online formats. The *Standards for Educational and Psychological Testing* (AERA et al., 2014) state that evidence supporting score interpretations and use should be provided when a testing program maintains test forms "administered under different test administration conditions are comparable for the same purpose" (see standard 5.17 of *the Standards*) (AERA et al., 2014).

Mroch, Li, and Thompson (2015) proposed a framework of score comparability focusing on construct and score equivalence, while considering a variety of test conditions. For the Graphic Literacy assessment, forms are built independently of test mode, using the same item pool and test specifications. ACT applies the same test equating methods for both paper and online forms to derive raw-to-scale score conversions. The mode comparability study for the Graphic Literacy assessment includes an evaluation of items, scores, and score conversions.

9.2.1 Mode Comparability: Study Design

ACT conducted a field study to evaluate the comparability of scores between paper and online administrations. In the field study, test centers were to randomly assign examinees to one of three proposed testing conditions. ACT directed the proctors to randomly assign test takers to take one of the three test forms: a Graphic Literacy online Form (G2C_LM1), a Graphic Literacy paper Form (G2P_LM2), or a Locating Information online Form (G1C_LM3). Examinees responded to the items on Forms G2C_LM1 and G2P_LM2 were used to evaluate mode comparability, and examinees responded to items on Forms G2C_LM1 and G1C_LM3 were used for the Linking Study. ACT directed the centers to have each test taker take all three WorkKeys NCRC assessments on the same or different days, with the test order counterbalanced across the sites. The test takers also completed a survey regarding their testing experience either at the end of each online assessment or after finishing all three paper assessments.

9.2.2 Mode Comparability: Sample

Similar to the scaling study presented in Chapter 8, ACT recruited a sample of examinees representative of the WorkKeys test-taker population.

Although ACT had instructed test centers to randomly assign examinees to the three conditions, ACT discovered that in some cases these instructions were not followed. Consequently, ACT did extensive review and cleaning of the test data. ACT removed data from a few centers where examinee distribution in the three conditions was extremely unbalanced (ACT defined an unbalanced test center as a center with a difference of 10 or more examinees between the different test conditions). Following data cleaning, ACT conducted further reviews to ensure that the remaining data represented random equivalent groups. A total of 37 testing sites participated in this study including 10 high schools and 27 adult testing centers across 20 states from different regions. Because the data may contain additional sampling error, measurement precision may be affected. As a result, the interpretations of the results below should be made with caution.

Final examinee counts were 701 and 668 for online (Form G2C_LM1) and paper (Form G2P_LM2) testing conditions, respectively. Table 9.1 presents the demographic distribution information. In general, the recruited samples for the two mode conditions are acceptable to represent the current WorkKeys test population, and are quite similar except for Caucasian groups (63% vs. 57% for online and paper testing).

Table 9.1: Sample Demographic Information for the Two Test Delivery Modes

| Mode | N | M (SD) | Gender | | Sample | | Ethnicity | | |
			F	M	HS	AD	W	B	H
Online	701	20.09 (6.76)	54%	44%	46%	54%	63%	15%	9%
Paper	668	20.17 (6.57)	52%	45%	46%	54%	57%	15%	9%

Note. Non-respondent or multi-races not included; F = Female. M = Male; HS = High School; AD = Adult; W = Caucasian; B = African American; H = Hispanic.

Across two mode conditions, the omit rates (no-answer) at each item are compared. As shown in Figure 9.1, the omit rates are generally below 10% for both conditions. The omit rates tend to be similar between modes.
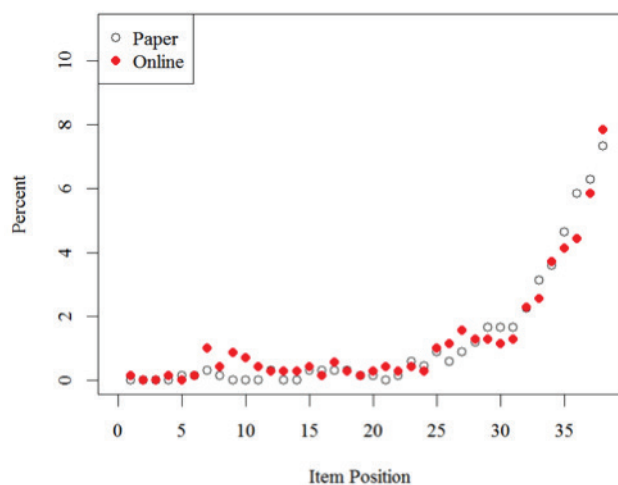
Figure 9.1: Comparison of Item Omit Rates for the Two Delivery Modes

9.2.3 Mode Comparability: Comparisons on Items, Tests, and Score Conversions

Item Level Comparison. Separate calibrations were conducted for the online and paper forms, and the item parameter estimates were transformed to the same pool scale. Table 9.2 shows the summary

statistics between the online and paper forms, and Figure 9.2 presents the scatterplots of item p-values and IRT b-parameter estimates. These results indicate that the item statistics were similar across the two mode conditions.

Table 9.2: Test Summary Statistics for Graphic Literacy

Mode	P	PBIS	IRT-a	IRT-b	IRT-c
Online	0.616 (0.173)	0.474 (0.100)	1.040 (0.228)	0.436 (1.096)	0.185 (0.052)
Paper	0.618 (0.171)	0.465 (0.084)	1.061 (0.307)	0.509 (1.002)	0.176 (0.043)

Note. P = p-value; PBIS = point biserial correlation; standard deviations are in parentheses.
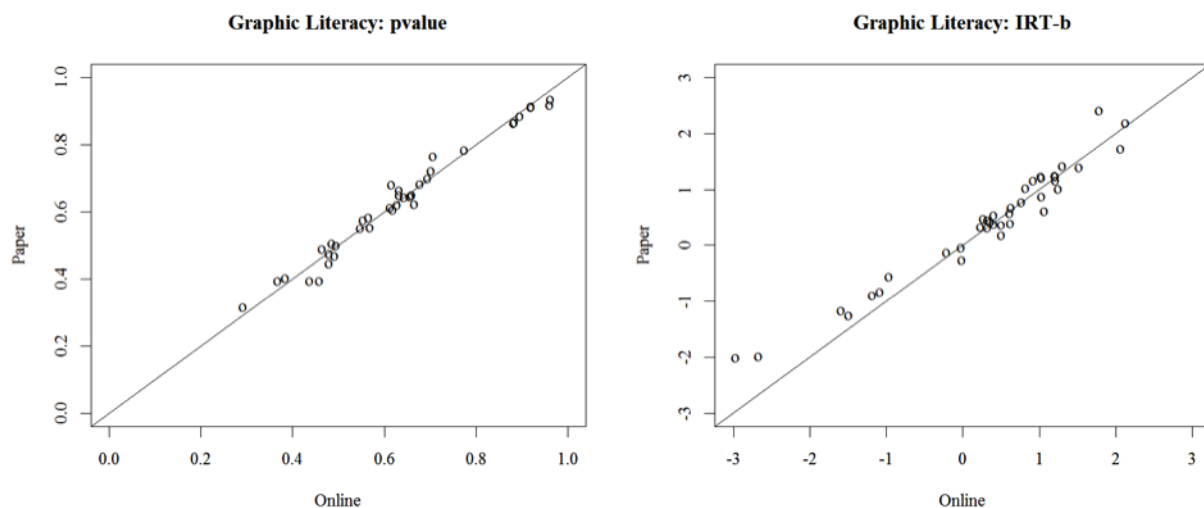
Figure 9.2: Scatterplots of Item p-values (left) and IRT *b*-parameter estimates (right) for the Two Delivery Modes.

Differential item functioning (DIF) analysis was also conducted on the items between paper and online forms. Only one item is flagged as Category C (favoring online testing) using the Mantel-Haenszel method.

Test Comparison. Figure 9.3 shows the comparisons of the Test Characteristics Curve (TCC) and Test Information Function (TIF). The TCCs are almost identical and the TIFs are very similar between modes, which indicate that the average mode effect is negligible.
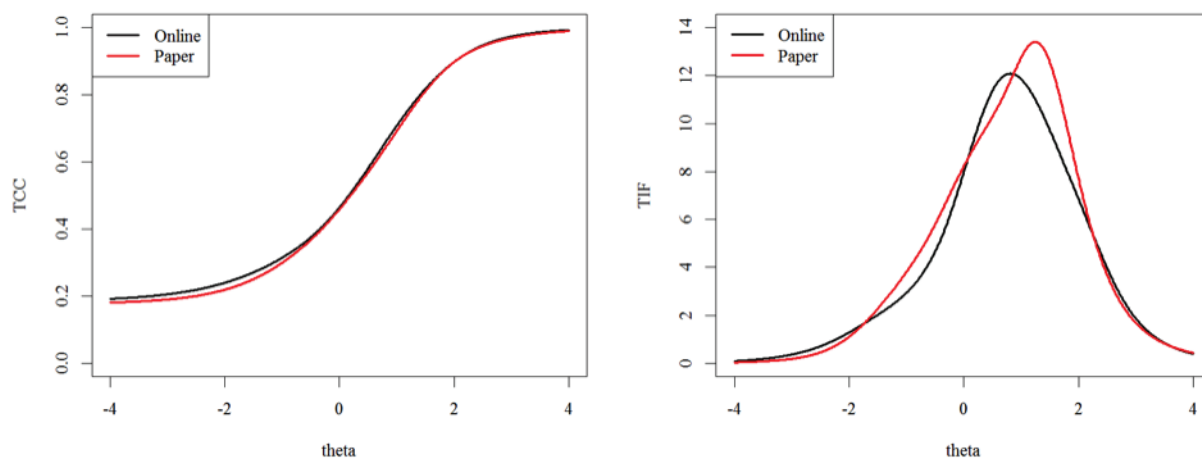
Figure 9.3: Comparisons of Test Characteristic Curves (left) and Test Information Functions (right) for the Two Delivery Modes

Score Conversion Comparison. Item 7 on Graphic Literacy forms was found to be flawed with no key during the analysis. As a result, the following procedure was applied: (a) conducted IRT true score equating from a 31-item to a 32-item base form to obtain the raw-to-scale score conversions; (b) added one score point to the raw score in the conversion to reach the 32 raw score points; and (c) added "0, 60.00000, 65" to the beginning of the conversion obtained in (b) to get the final conversion. Item 7 was scored as all correct for everyone who took the Graphic Literacy forms, and the scale score for each test taker was then obtained by applying the conversion from (c).

Figure 9.4 compares the raw-to-scale score conversions. Comparing the raw-to-scale score conversion between modes, five raw score points differed on their conversion to scale scores; however, the raw score conversions to level scores were identical for the two modes.
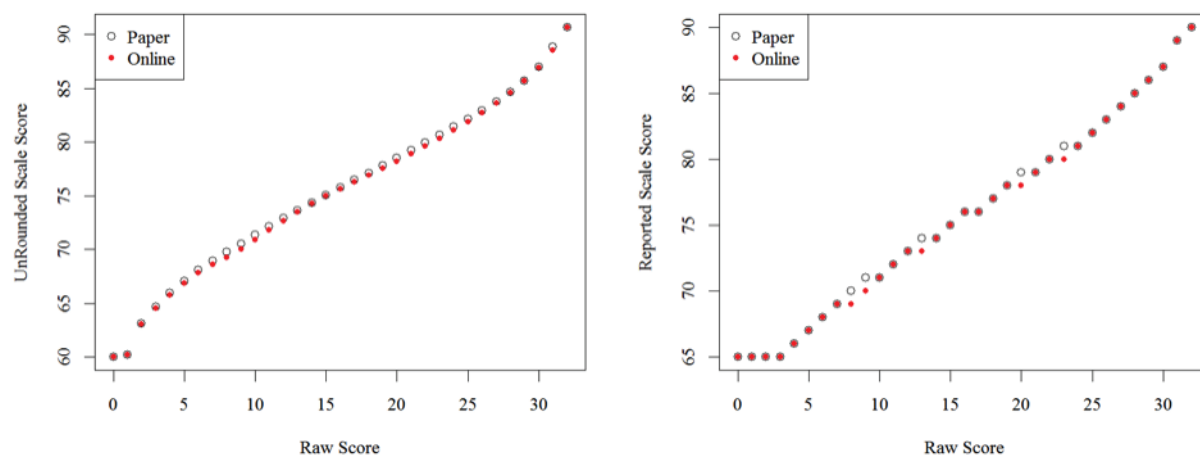
Figure 9.4: Comparison of Unrounded (left) and Reported (right) Raw-to-Scale Score Conversions for the Two Delivery Modes

Figure 9.5 shows the Conditional Standard Error of Measurements (CSEMs). The raw score CSEMs tend to be larger in the middle and smaller at the two ends and the scale score CSEMs tend to be flat for most of the score points. The CSEMs for both scores appear to be similar between modes.
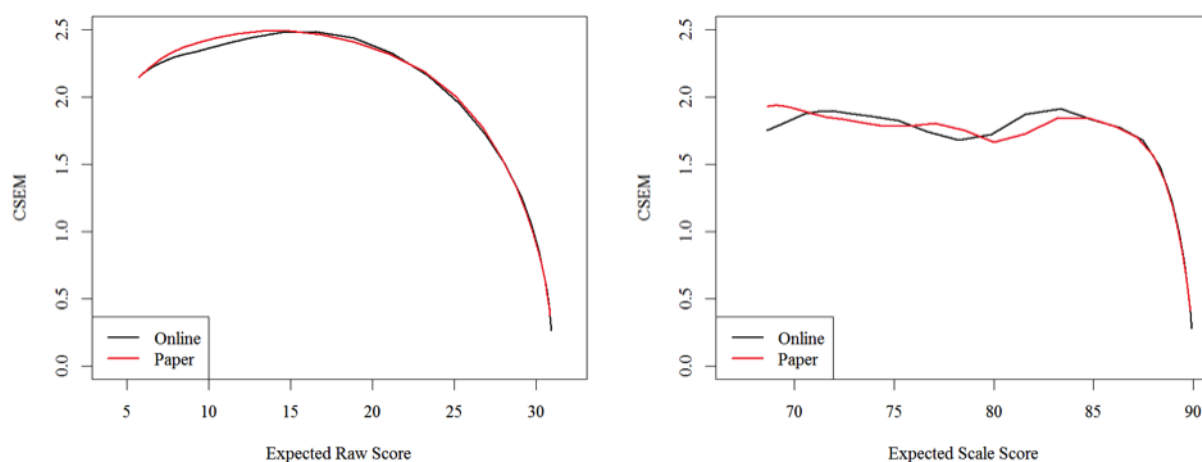
Figure 9.5: Comparison of CSEMs for Raw Scores (left) and Scale Scores (right) for the Two Delivery Modes

9.2.4 Mode Comparability: Score Comparisons

Table 9.3 presents the summary statistics for the raw and scale scores by mode. Figure 9.6 presents the raw score distributions, and Figure 9.7 presents the scale score distributions. The results are very similar for the two modes. For both type of scores, mean differences are below 0.21 and the effect sizes are below 0.041, indicating nearly identical score distributions.

Table 9.3: Summary for Raw and Scale Scores for the Two Delivery Modes

Score	Mode	M	SD	P10	P25	P50	P75	P90	P95	M Diff.	ES	t-test prob
Raw Scores	Online	20.09	6.76	11	15	20	26	29	30	0.08	0.012	0.829
	Paper	20.17	6.57	11	15	20	26	29	30			
Scale Scores	Online	79.70	5.33	72	75	78	83	86	87	0.21	0.041	0.453
	Paper	79.91	5.09	72	75	79	83	86	87			

Note. M Diff. = mean difference; ES = effect size.
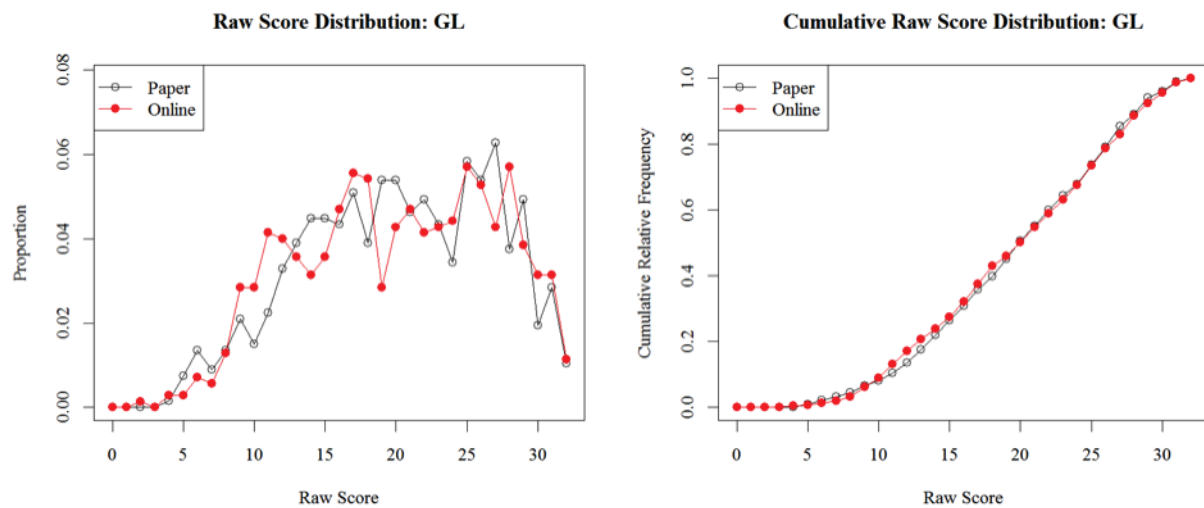
Figure 9.6: Comparison of Raw Score Distributions for the Two Delivery Modes
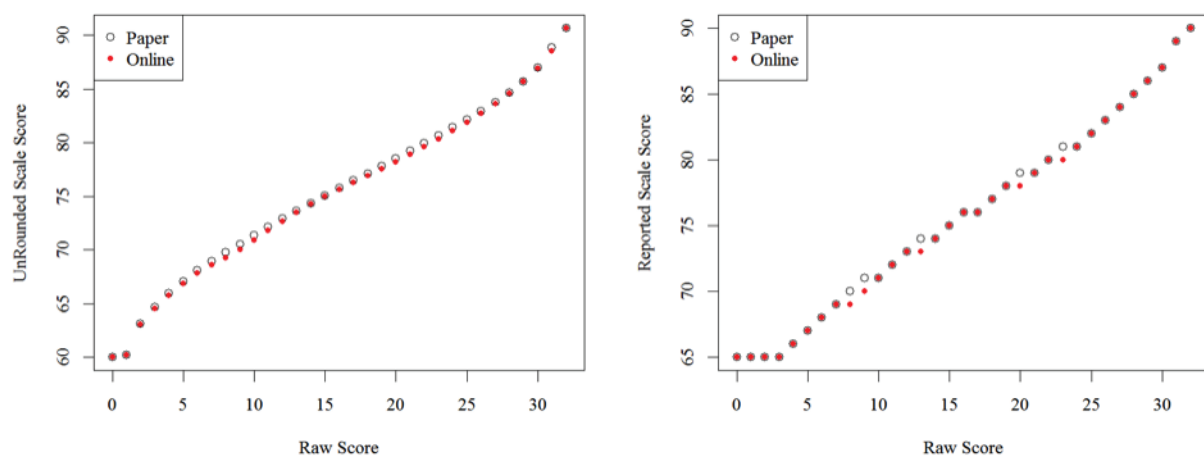
Figure 9.7: Comparison of Scale Score Distributions for the Two Delivery Modes

Based on the findings of the analysis, ACT concluded that no significant mode effect existed. Due to the limitations of the field test data, ACT will continue to monitor potential mode effects on the Graphic Literacy assessment to ensure the comparability of test scores for paper and online administrations.

9.3 Linking Locating Information to Graphic Literacy Score Scale

When a test publisher needs to modify the test construct, update test specifications, or refresh content to improve an existing assessment, test score users often need to understand the relationships between the old and new assessments. To facilitate a better understanding of the relationship between the different tests or different versions of a test, a statistical procedure is often used to make adjustments to link the scores from one test to another. There are generally four types of linking which are ordered in terms of the "strength" of the resulting relationship: equating, calibration, projection, and moderation (Linn, 1993; Mislevy, 1992). Concordance is a type of statistical moderation of "matching distributions" using percentile ranks to derive a table that links the scores between two tests. Holland (2007) points out that "Concordance represents scaling of tests that are very similar but that were not created with the idea that their scores would be used interchangeably" (p. 19). Different from the equating of two forms of a same test which produces comparable scores, scores from concordance of two tests are *not* interchangeable.

The Graphic Literacy assessment was developed based on redesigned test specifications of the Locating Information assessment (see Chapter 3 for the test specifications). To facilitate a smooth transition from Locating Information to Graphic Literacy assessments, ACT conducted a Linking Study in the spring of 2017. The focus of the Linking Study was to develop a concordance between Locating Information and Graphic Literacy assessments. Concordance between the two assessments is defined by identifying the scale scores on the Locating Information assessment that have the same percentage of test takers at or below the given scale score points on the Graphic Literacy assessment within the linking study sample. This document summarizes the findings from the Linking Study, as a means to better understand the relationships between the two assessments and ultimately to assist users in appropriately interpreting the scores or score trends derived from the two assessments.

9.3.1 Study Design and Sample Representativeness

A total of 43 testing sites were administered both Forms G2C_LM1 (Graphic Literacy online) and G1C_LM3 (Locating Information online) including 10 high schools and 33 adult testing centers across 20 states. More than 800 test takers took one of the two Linking forms and they were given 55 minutes to complete each test. The sample sizes were similar between the two forms. In general, the recruited sample is representative of the WorkKeys test population based on the demographic characteristics (see Table 9.1).

Although the Graphic Literacy assessment is a new assessment with new constructs and test specifications that differ from the Locating Information assessment, resulting scores are not interchangeable, it is desirable to have similar difficulty and measurement precision to strengthen the concordances. A series of analyses were conducted to evaluate and compare psychometric properties of the two assessments in terms of omit rates, testing time, scale score summary statistics, reliability, and Standard Error of Measurement (SEM).

9.3.2 Comparison of Omit Rates and Testing Time Between Locating Information and Graphic Literacy

Figure 9.8 presents the omit rates for each item in both Graphic Literacy and Locating Information forms administered in the Linking Study. In general, the figure indicates that the omit rates are less than 10% for most items except for the last item in Locating Information Form G1C_LM3. In addition, as summarized in Table 9.4, test takers spent an average of four minutes less time testing on Form G2C_LM1 than on Form G1C_LM3. The less testing time spent on this form can be explained in part by the Graphic Literacy assessment having fewer graphics than the Locating Information assessment (17 or 18 in Graphic Literacy versus 32 in Locating Information).
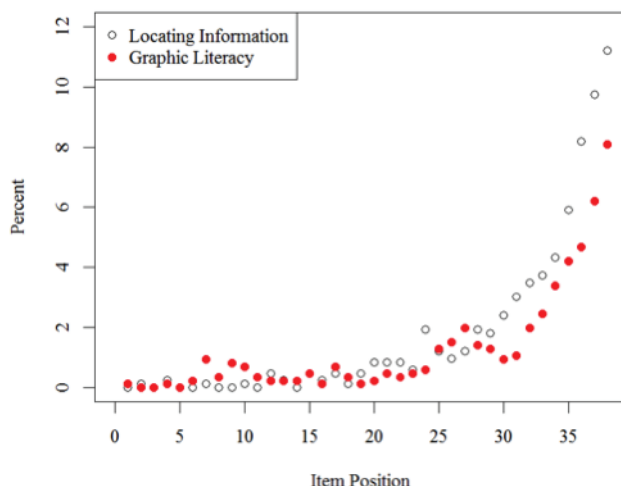
Figure 9.8: Comparison of Item Omit Rate Between Locating Information and Graphic Literacy

Table 9.4: Summary for Total Testing Time (in minutes)—Locating Information and Graphic Literacy

Form	N	Mean (SD)	Min	P5	P10	P25	P50	P75	P90	P95
G1C_LM3	830	38.77 (13.58)	5	13	18	29	41	52	54	55
G2C_LM1	854	34.91 (11.95)	6	13	18	27	35	45	51	53

9.3.3 Scale Score Distributions for Locating Information and Graphic Literacy

Because no significant mode effect was observed in the Mode Study, the item parameter estimates were then re-calibrated using the combined data from both paper and online administrations to derive the conversion for the Graphic Literacy (G2_LM) Form. It should be noted that the Graphic Literacy Form was equated to the base form for Graphic Literacy due to the scoring issue with item 7. Tables 9.5 and 9.6 provide the summary statistics for the raw scores and the scale scores for the Linking Study. Based on average IRT-b statistics, the Graphic Literacy Form, G2_LM, appears to be slightly more difficult than the Locating Information Form, G1C_LM3.

Table 9.5: Test Summary Statistics for Locating Information and Graphic Literacy

Form	p	PBIS	IRT-a	IRT-b	IRT-c
G1C_LM3	0.604 (0.256)	0.381 (0.121)	0.973 (0.282)	0.297 (1.769)	0.172 (0.052)
G2_LM	0.614 (0.177)	0.462 (0.100)	1.022 (0.261)	0.435 (1.124)	0.178 (0.054)

Note. p = p-value; PBIS = point biserial correlation; standard deviations are in parentheses.

Table 9.6: Scale Scores Summary Statistics for Locating Information and Graphic Literacy

Form	N	Mean (SD)	P5	P10	P25	P50	P75	P90	P95
G1C_LM3	830	76.71 (4.11)	70	71	74	76	79	82	84
G2_LM	854	78.64 (5.14)	70	72	75	78	82	86	87

Figure 9.9 presents the relative frequency distributions (left) and cumulative relative frequency distributions (right) for the Locating Information and Graphic Literacy Forms. These plots suggest that the scale score distributions are different for two assessments where significant modifications were made to the Graphic Literacy assessment.
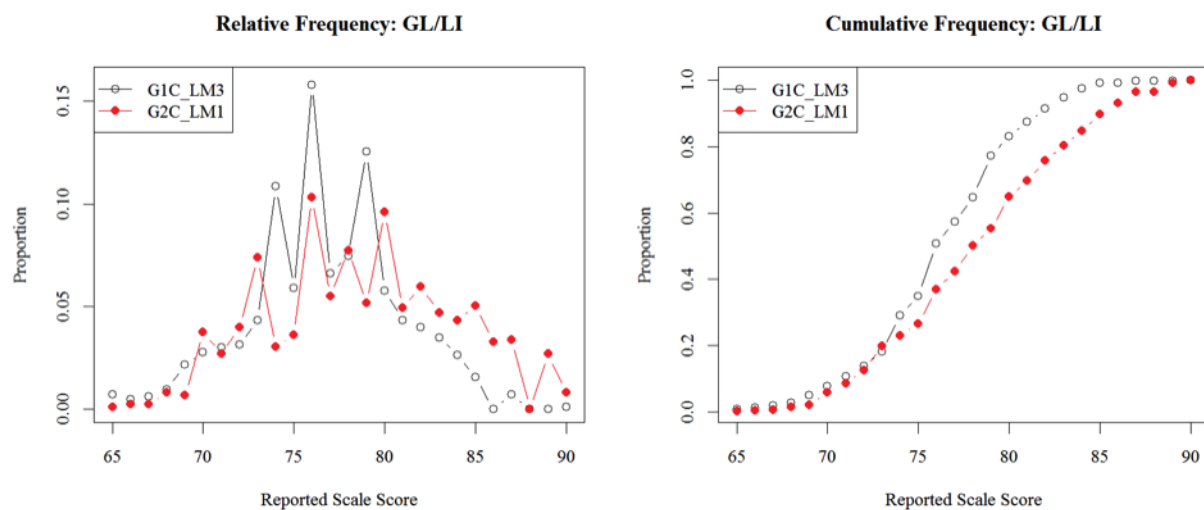
Figure 9.9: Comparison of Relative (left) and Cumulative (right) Frequency Distribution for Graphic Literacy and Locating Information

9.3.4 Concordance from Locating Information to Graphic Literacy

Given the changes in test specifications and the need to link the Locating Information and Graphic Literacy assessments, statistical moderations using an equating method were performed to link scores from Locating Information (LI 1.0) to Graphic Literacy (GL 2.0) assessments. The concordance was based on the equipercentile method with smoothing (S) of 0.05 for Locating Information to Graphic Literacy.

9.3.5 Evaluation of Locating Information Forms After Linking

Table 9.7 provides the summary statistics of the scale scores for the original Locating Information Form (G1C_LM3) before and after it was transformed to the Graphic Literacy scale (G1C_LM3*), and the Graphic Literacy Form (G2_LM). It can be observed that the means, standard deviations, and quantiles of the transformed scale score on the Locating Information Form (G1C_LM3*) are very similar to the Graphic Literacy Form (G2_LM).

Table 9.7: Summary Statistics of Scale Scores Before and After Concordance

Scale	Form	N	Mean (SD)	P10	P25	P50	P75	P90	P95
LI 1.0	G1C_LM3	830	76.71 (4.11)	71	74	76	79	82	84
GL 2.0	G1C_LM3*	830	78.73 (5.06)	72	75	78	82	85	87
GL 2.0	G2_LM	854	78.64 (5.14)	72	75	78	82	86	87

Note. G1C_LM3* = G1C_LM3 implemented GL 2.0 scale score concordance table.

Table 9.8 provides summary statistics of the Level Scores for the previous version of the Locating Information Form (G1C_LM3) before and after it was transformed to the Graphic Literacy scale (G1C_LM3*), and the Graphic Literacy Form (G2_LM). The means and standard deviations are very similar between G1C_LM3* and G2_LM, except for the P90 quantile. The Level cuts for the Graphic Literacy assessment were developed based on a standard setting study using a Mapmark method (see Chapter 8 for greater detail on the Standard Setting process).

Table 9.8: Summary for Level Scores Before and After Concordance

Scale	Form	N	Mean (SD)	P10	P25	P50	P75	P90	P95
LI 1.0	G1C_LM3	830	3.62 (1.42)	<3	3	4	4	5	5
GL 2.0	G1C_LM3*	830	4.57 (1.81)	3	3	5	6	6	7
GL 2.0	G2_LM	854	4.46 (1.83)	3	3	5	6	7	7

The results suggest that in order to compare the scores from the Locating Information and Graphic Literacy assessments and to understand the score relationships between the two assessments, the scale scores on the Locating Information assessment need to first be transformed to the Graphic Literacy scale based on the concordance table. Test users need to be aware that the concordance scale scores do not always represent the test scores that a test taker would achieve if he or she were to take the Graphic Literacy assessment. Similarly, comparing group performance averages or analyzing year-to-year performance trends using concordance scores from a test that has not been taken need to be made with a good deal of caution.

Chapter 10

Reliability and Measurement Error

10.1 Overview

This chapter reports the reliability evidence of the WorkKeys® Graphic Literacy assessment. Reliability and measurement error are fundamental for evaluating the psychometric qualities of an assessment in order for the assessment claims defined in Chapter 1 to be substantiated. As the *Standards for Educational and Psychological Testing* (referred to as *the Standards* below) states, "for each total score, subscore, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/ precision should be reported" (AERA et al., 2014, p. 43 as Standard 2.3).

According to *the Standards*, reliability is the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker; the degree to which scores are free of random errors of measurement for a given group (AERA et al., 2014). As a quantitative measure of the consistency of an assessment, reliability is closely related to Standard Error of Measurement (SEM). SEM is the standard deviation of an individual's observed scores from repeated administrations of a test (or parallel forms of a test) under identical conditions (AERA et al., 2014). The SEM summarizes the amount of error or inconsistency in test scores.

Because any WorkKeys foundational skill assessment, including the Graphic Literacy assessment, classifies examinees into skill-level groups, classification consistency is important to support level score uses. Classification consistency is defined as the extent to which the classification of examinees into groups is identical when obtained from two independent administrations of a single form or two parallel forms of a test. Because assessments are usually administered only on one occasion to the same examinee, classification consistency is estimated from a single test administration with strong assumptions made about distributions of measurement errors and true scores.

The following sections provide results related to (a) reliability coefficients and SEM estimates of raw scores and scale scores based on Classical Test Theory, (b) reliability coefficients of level scores based on Generalizability Theory, and (c) classification consistency of level scores.

10.2 Reliability Coefficients and Standard Error of Measurement (SEM)

Reliability coefficients quantify the consistency level of test scores. They typically range from zero to one, with values near one indicating high consistency and those near zero indicating little or no consistency. Based on a single test administration, internal consistency reliability, usually measured by Coefficient Alpha (Cronbach, 1951), is one of the most widely used indices of test score reliability. Coefficient Alpha is computed as a reliability estimate for raw scores using the following formula:

$$\hat{\alpha} = \left(\frac{k}{k-1} \right)\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_x^2} \right),$$

where k is the number of test items used for scoring, s_i^2 is the sample variance of the ith item, and s_x^2 is the sample variance of the observed raw score.

For scale scores of test t, the reliability estimate (r_t) can be obtained using the following formula:

$$r_t = 1 - \frac{SEM_t^2}{s_t^2},$$

where SEM_t is the average of estimated scale score CSEMs and s_t^2 is the sample variance of the observed scale score. 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 scale scores for about two-thirds of the test-taker group are within plus or minus one SEM of their scale score.

Table 10.1 presents the Coefficient Alphas and the SEMs for the Graphic Literacy assessment for both raw scores and scale scores. The reliability and SEM estimates are based on the sample utilized for the Scaling Study described in Chapter 8. The sample included 1,170 examinees following data cleaning. For score use, a minimum value of 0.80 is required for reliable test score interpretations. The reliability estimates for both the raw and scale scores exceed the threshold of 0.80. (Corresponding plots of Conditional Standard Error of Measurement (CSEM) on raw scores and scale scores are presented in Chapter 8.)

Table 10.1: Coefficient Alphas and SEMs for Graphic Literacy Form G2C_S1

Form	N	Raw Score		Scale Score	
		Coefficient Alpha	SEM	Coefficient Alpha	SEM
Form G2C_S1	1,170	0.85	2.34	0.85	1.71

10.3 Generalizability Theory

Reliability based on Generalizability Theory was also investigated. Generalizability Theory provides a broad conceptual and statistical framework for evaluating measurement precision (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Generalizability Theory not only produces reliability-like coefficients known as generalizability and dependability coefficients but also disentangles and estimates multiple sources of error. Multivariate Generalizability Theory (Brennan, 2001) can address issues involved in analyzing data for a stratified test under a table of specifications. In Graphic Literacy forms, items are nested (stratified) within specific levels of difficulty, that is, Levels 3 to 7. A mixed model of *persons x (items:strata)* or *p x (i:h)* from a multivariate perspective was used, and the results are presented in Table 10.2 with the following highlights:

- The estimated universe score variance which is analogous to the true score variance, $\hat{\sigma}^2(p)$, is relatively larger at the middle levels of items, suggesting that the average performances can be differentiated more on the moderately difficult items than the easy or difficult items;

- Variability of item difficulty, $\hat{\sigma}^2(i)$, is small, suggesting that difficulty is similar among items within each level;

- Interactions of person-by-item, $\hat{\sigma}^2(pi)$, are greater for the items at Levels 5 to 7 than those at Levels 3 and 4, indicating that performance is less consistent across the items at Levels 5 to 7 than at Levels 3 or 4;

- The estimates of error variances, $\hat{\sigma}^2(\delta)$ for norm-reference decisions and $\hat{\sigma}^2(\Delta)$ for criterion-reference decisions, are similar due to the small $\hat{\sigma}^2(i)$;

- The reliability-like coefficients, $E\hat{\rho}^2$ for norm-reference decisions and $\hat{\Phi}$ for criterion-reference decisions, are 0.43 or higher at each level with Level 3 having the lowest value;

- The estimated effective weights which indicate relative contributions of each level of items to the total variance are higher for the middle levels (Levels 4 through 6) than for Levels 3 and 7. The results, related to the numbers of items by level, suggest that moderately difficult items are more heavily weighted in forming the total scores than the other items in the test;

- For total scores, the reliability-like coefficients for both rank-ordering test takers and judging performance levels of test takers are both equal to 0.85.

Table 10.2: Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Graphic Literacy Form G2C_S1

Level	I	$\sigma^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(\delta)$	$\hat{\sigma}^2(\Delta)$	$E\hat{\rho}^2$	$\hat{\Phi}$	Effective Weight
3	4	0.020	0.008	0.098	0.024	0.026	0.45	0.43	0.08
4	7	0.028	0.011	0.158	0.023	0.024	0.55	0.54	0.20
5	9	0.041	0.019	0.169	0.019	0.021	0.69	0.66	0.32
6	7	0.040	0.003	0.207	0.030	0.030	0.57	0.57	0.24
7	5	0.040	0.010	0.175	0.035	0.037	0.53	0.52	0.16

10.4 Classification Consistency of Level Scores

The Standards (AERA et al, 2014, p. 46 as Standard 2.16) recommends that test publishers provide information about the percentage of test takers who would be classified in the same way for classification tests if they were to take a test twice using alternate forms. Classification consistency ranges from 0 to 100 percent, with values near 100 indicating higher consistency and those near zero indicating little or no consistency.

According to Subkoviak (1984), two important classification consistency indices are:

 • agreement index p, which is the proportion of consistent classification based on two parallel forms, and

 • coefficient κ, which is the proportion of consistent classification adjusted for chance agreement.

The classification consistency indices computed using the IRT methodology (Schulz, Kolen, & Nicewander, 1997, 1999) for Graphic Literacy Form G2C_S1 data are presented in Table 10.3. The second row of the table, labeled "Exact," shows the percentages of test takers who would receive the same level score from two parallel forms. For example, if a test taker were to take two parallel forms of the test and score at Level 3 on both forms, this would be a case of exact agreement. For Graphic Literacy Form G2C_S1, the estimated exact agreement is 52 percent. The remaining rows show the consistency of aggregated classifications (i.e., at-or-above) at each level. Aggregated classification consistency for a level score is the summary of test-taker percentages of two groups: Both scores are either below the level score, or at-or-above it. For example, a test taker who scores at Level 4 and Level 5 on two testing occasions would not be consistently classified as Level 5, but would be consistently classified as Level 4 or above. In this study, aggregated classification consistency of level scores is estimated to be 84 percent or higher. As expected, the values of coefficient κ are lower than those of agreement index p.

Estimates of classification consistency are sensitive to the distribution of skill levels in the test taker sample. For example, the mean of the test taker sample is at the Level 5 theta cutoff, suggesting that the true skill of a relatively large proportion of these test takers was close to the Level 5 theta cutoff. Generally, test takers are more likely to be misclassified because of measurement error when their true skill is closer to the cutoff.

Table 10.3: Estimated Classification Consistency Indices for Level Scores for Form G2C_S1

Level	p	κ
Exact	52%	39%
3	93%	45%
4	85%	65%
5	84%	69%
6	88%	67%
7	96%	52%

In summary, the reliability and classification consistency findings above are deemed acceptable based on the available field study data presented in Chapter 8. As the Graphic Literacy assessment is administered to large numbers of examinees, ACT will continue to update the findings related to test score reliability and classification consistency.

Chapter 11

Validity

11.1 Validation of Test Score Uses and Interpretations

The *Standards for Educational and Psychological Testing* (AERA et al., 2014) define validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses" (p. 11). In adhering to this understanding of validity, the ACT WorkKeys® Assessments incorporated an approach of gathering evidence as a means to enable users to evaluate the appropriateness and reasonableness of test score interpretations and uses.

To validate test score interpretations and uses 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 henceforth becomes a scientific process designed to evaluate the degree to which the analytic and empirical evidence supports the assessment claims.

Validation, as a scientific process, entails the careful articulation of test claims along with the inferences and assumptions required to build the connections from examinee task performance to score-based interpretations and uses. The assessment claims are explicit statements regarding the purpose of the assessment and how test scores are to be interpreted and used. As such, the claims provide the framework for validation. When clearly specified, an evidentiary chain is built between the claims and associated evidence. If the claims are rational, and their associated inferences and assumptions are plausible based on evidence, then the defined test score should also be considered plausible or valid (Kane, 2013; Messick, 1989).

Validation of test score interpretations and uses through the evaluation of evidence does not lead to a Yes/No validity determination. Validation is a matter of degree, requiring interpretation and insight into the underlying theory supporting the meaning of the test scores and the potential uses and consequences of score-based decisions. As several theorists have argued, a test may be interpreted as appropriate and

valid for one usage, but altogether inappropriate and problematic for a second usage. As a result, it is the usage and decisions stemming from test scores that are validated and not the test itself (Cronbach, 1988; Kane, 2006; Messick, 1989).

In collecting and evaluating evidence regarding Graphic Literacy test score interpretations and usage, WorkKeys Assessments subscribed to the concept of validity as a claims-based argument (Cronbach, 1988; Kane 2006, 2013; Mislevy, 2006). In adhering to a claims-based validation approach, WorkKeys Assessments also utilized the principles of Design Science (Johannesson & Perjons, 2014; Van Aken & Romme, 2012) as a means of clearly defining the assessment problem, developing proposed solutions, gathering feedback and test data, and documenting evidence and decision making (Langenfeld & Thomas, 2016).

The Graphic Literacy Design Team implemented a process that began by articulating the purpose of the assessment and its associated claims; it culminated with the collection of data from various sources to evaluate the validity use argument. The purpose of the validity chapter is to present the Graphic Literacy assessment claims and assumptions, then provide evidence to evaluate the appropriateness of the proposed interpretations and uses.

11.2 Purpose of the Graphic Literacy Assessment

The Graphic Literacy assessment provides information to examinees, employers, workforce development officials, and educators. For examinees, the assessment provides them with insights in regards to their foundational graphic literacy skills and their career readiness. In some cases, scores on the assessment may assist examinees in finding employment. For employers, the assessment provides information that may be used, with other information, for employment decisions. For workforce development officials, the assessment provides information regarding the work-ready status of individuals requesting services and also assists them in guiding individuals toward jobs. For secondary educators, the assessment provides information related to foundational skills and career readiness that may be used as an accountability measure. For postsecondary educators, the assessment provides information related to program readiness or program evaluation. For the assessment to be used appropriately for each of these purposes, ACT needs to collect evidence and evaluate it. Additionally, ACT needs to provide guidance in regards to the proper use of the assessment for each purpose.

An additional purpose of the Graphic Literacy assessment relates to the issuance of the ACT® WorkKeys® National Career Readiness Certificate (NCRC®). The assessment constitutes one of three assessments that are used to determine an examinee's achievement of a WorkKeys NCRC. The WorkKeys NCRC is an evidence-based career readiness credential, which assists both examinees and employers in various ways. For the examinee, the WorkKeys NCRC provides them with a better understanding of their level of foundational skills. The WorkKeys NCRC level and assessment scores provide both examinees and counselors with insights in regard to their skill levels and how these relate to various occupations.

WorkKeys developed the Graphic Literacy assessment as a replacement for the Locating Information assessment. Locating Information measured examinees' ability to locate, compare, summarize, and

analyze information presented in graphical format (ACT, 2008). The new Graphic Literacy assessment was designed as an update to the Locating Information content and measured construct.

Using data and knowledge gained through over 20 years of administering the Locating Information assessment, ACT was able to more fully develop the updated Graphic Literacy construct and content. Graphic literacy utilizes words, pictorial shapes and symbols, and numbers as visual representations to effectively communicate information and inform decision making (Few, 2012; Friel & Bright, 1996; Shah & Freedman, 2011). In ACT's definition, graphical representations are used to communicate both quantitative and qualitative information. The basic skill in comprehending graphical representations is locating information; however, the assessment measures more than just the basic skill. It also measures examinees' ability to interpret and apply 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 decisions. As such, scores on Graphic Literacy cannot be used interchangeably with scores from the previous assessment, Locating Information.

The WorkKeys assessment program was conceived to mitigate the "skills gap" problem. The skills gap is a term used to describe the challenge that employers and hiring managers face. The skills gap occurs because many well-paying jobs exist; but, due to the shortage of qualified workers, employers are unable to find workers to fill them. Goldin and Katz (2008) provide evidence demonstrating that, since 1970, United States educational achievement has increased only marginally while technological advances and requirements in business and industry have increased greatly. Many of the problems associated with businesses being unable to find quality workers is the result of average worker skill levels being little better than the skill levels of 40 years ago. As a result, a discrepancy exists between employer needs and the skill sets many workers bring to the job (Autor, 2015; Goldin & Katz, 2008). (For more detailed information on the skills gaps, see Chapter 1.)

Because of the discrepancy between educational achievement and job requirements, WorkKeys provided a means of addressing the skills gap for both employers and workers (ACT, 2011). Through the use of WorkKeys assessments and the WorkKeys NCRC, workers can demonstrate the foundational skills needed in today's economy. For the employer, WorkKeys assessment scores allow them to use skills-based hiring practices as a means of identifying the right person for the job.

11.3 Graphic Literacy Assessment Claims

Drawing on its understanding of the skills gap and skills-based hiring practices, the Design Team developed three primary claims for the Graphic Literacy assessment. The primary claims are each supported by critical assumptions.

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 and interpretable 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 scores) 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 employee 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 who 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) from new employees than if the company had hired 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 who hire U.S. examinees of high school or work force 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 if the companies had hired 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 who score at higher levels on Graphic Literacy are less likely to quit in the first 6 months of employment than employees with lower score 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 addressed questions concerning examinee job 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 prepare students to take jobs in the state or region, and examinees who are, or will be, seeking employment or career advancement.

The *Standards* (AERA et al., 2014) identify five sources of validity evidence: (a) evidence based on test content, (b) evidence based on internal structure, (c) evidence based on relationships to other variables, (d) evidence based on response processes, and (e) evidence based on consequences of testing. The remainder of the chapter applies a validity use argument (Kane, 2013) to provide evidence first related to the assumptions associated with the claims and then for the claims themselves.

11.4 Graphic Literacy–A Measure of Foundational Workforce Skills

All three primary claims are dependent on the validity of initial assumptions:

1. graphic literacy is a foundational workplace skill and is required for success in a large number of jobs;

2. ACT has developed a valid and appropriate construct definition of graphic literacy;

3. ACT's Graphic Literacy assessment provides reliable and interpretable scores that reflect the construct. ACT's Graphic Literacy assessment elicits observable evidence of the construct;

4. ACT has defined appropriate Graphic Literacy PLDs and has established standards aligned to the PLDs; and

5. cut scores used to delineate each performance level have sufficient classification accuracy.

For the primary claims to be plausible, evidence supporting each of the five assumptions needs to be evaluated.

The next subsections present data and analysis related to the five assumptions. The analysis draws on the professional literature from the fields of educational measurement and industrial-organization psychology, as well as data that ACT collected from over 20 years of job profiling, from three separate field test studies, and from a series of standard setting meetings.

11.4.1 Foundational Workplace Skills

Foundational workplace skills are 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, licensed electricians require skills in working with electrical circuits and wiring to perform their jobs. Foundational skills are more general than job skills; they are the skills that enable a person to learn specialized job skills.

Foundational skills are often referred to as basic or academic skills taught through formal schooling, but they may be also learned from other sources. The foundational skills are frequently 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. Chinn (2017) described one facet of foundational skills as "being able to understand and ask questions in a variety of forms, including graphs, charts, and diagrams, as well as written or spoken words" (p. 1).

Foundational skills are fundamental in that they serve as the basis for supporting additional learning. They are "portable" in that, rather than being job specific, they can be applied at some level across a wide variety of jobs and occupations (Symonds, 2011). In the 21st century, multiple studies and surveys have identified the need for employees to be engaged in lifelong or fluid learning (Infosys, 2016; NNBIA, 2014; Organization of Economic Cooperation and Development [OECD], 2016; Society for Human Resource Management [SHRM], 2010). As the economy has become more technical and global, the pace of change has increased greatly. The concept of a job for life has become outdated. Successful workers will have a flexible mind set and the basic skills needed to continually learn and re-train themselves to remain relevant and successful in a dynamic and shifting economy (Infosys, 2016).

11.4.2 Graphic Literacy—A Foundational Workplace Skill

In the assumptions supporting the assessment claims, ACT identified Graphic Literacy as one facet of foundational workplace skills. As stated above, foundational skills commonly are identified as academic subjects, yet few students have taken a class in graphics or graphic literacy. Why then does ACT claim that graphic literacy is a foundational workplace skill? ACT based its argument on three sources of evidence: (1) job analysis data that has consistently indicated that graphic skills are needed to achieve job success, (2) professional literature and job competency models that identify graphic literacy as a critical 21st century skill, and (3) descriptions of the Programme for the International Assessment of Adult Competencies (PIAAC) assessments in which the ability to understand and interpret information presented in graphic format is a component of adult literacy and numeracy.

Since initiating its job profiling services in 1993, ACT has collected data on over 21,000 job profiles representing a wide cross-section of U.S. jobs. Job profiles have been conducted on jobs in manufacturing, health care, construction, financial services, public administration, leisure and hospitality, agriculture, and other sectors. ACT has profiled 193 (just under 50%) of the 387 Bright Outlook Occupations as identified by O*NET using Bureau of Labor Statistics projection data (U.S. Bureau of Labor Statistics, 2013). Analysis of the job profile database indicates that the skills associated with graphic literacy were included in 13,725 profiles or slightly more than 64% of all ACT profiles. When ACT assigned each completed profile to an O*NET job code, graphic literacy appeared as a required skill for 664 distinct O*NET job codes or 61% of all O*NET job codes.

In recent years, several business and industry associations have built 21st century workplace competency models that provide support for the inclusion of graphic literacy as a foundational workplace skill (Association for Career and Technical Education [ACTE], 2010; Infosys, 2016; NNBIA, 2014).

The competency model developed by the Business Roundtable (NNBIA, 2014) defined common employability skills. Under the category of workplace skills, they emphasize Planning and Organizing, Problem Solving, Decision Making, and Working with Tools and Technology. Within each of these areas, they define specific skills that involve graphic literacy, which they state is required to interpret and understand data in order to make decisions.

ACTE (2010) argues that students must be able to apply academic knowledge to authentic situations that they might encounter in their careers. The report asserts that students will receive workplace communications in varied formats designed to provide information needed to successfully perform tasks. It maintains that students require instruction and practice to interpret information and data presented in various modes. The report concludes that students are not receiving sufficient instruction in various modes of communication (including graphics) where information is conveyed to produce a workplace action.

Graphic literacy as a foundational workplace skill is further supported by PIAAC's assessments of adult competencies. PIAAC evaluates the status of adult workplace competency through three different assessments: Literacy, Numeracy, and Problem Solving in Technology-Rich Environments (OECD, 2016). In the Literacy assessment, they present reading passages to adult test takers using continuous and non-continuous texts. OECD (2016) defines non-continuous texts as "organized in a matrix format or around graphic features. Several different organizing structures are identified, including simple and complex lists and graphic documents (e.g., graphs, diagrams), locative documents (e.g., maps), and entry documents (e.g., forms)" (p.19). In the Numeracy assessments, examinees are required to identify trends and relationships using data presented in graphic formats. Thus, OECD has included the ability to understand, use, and interpret information presented in graphics as part of their definition of adult literacy and numeracy.

Based on the understandings gained from studying ACT's job profiling data, the workforce competency models, and the construct definitions developed for the PIAAC assessments, graphic literacy is a critical foundational workplace skill that contributes to employee success and lifelong learning.

Of course, graphic literacy, like math and other skills, is not universally required across all jobs. As stated above, ACT has found that graphic literacy skills are used in 664 distinct O*NET job codes or approximately 61% of all O*NET job codes. When the Graphic Literacy assessment is used as a part of the hiring process, ACT recommends that the employer gathers evidence to support the relevancy of the assessment and level score requirements. ACT provides its job profiling service as a valid method for gathering the required evidence to demonstrate both assessment relevancy and score level requirements.

11.4.3 Graphic Literacy—Construct Defined

A detailed description of the Graphic Literacy construct is provided in Chapter 2. Summarizing Chapter 2, Graphic Literacy is designed to assess an essential 21st-century workplace domain that employees use to find, summarize, compare, and analyze graphic resources to make decisions. Graphic resources are visual representations of information designed to convey understanding to a potential user. Graphical resources may use words, but they also use symbols, shapes, lines, numbers, and pictures to enable understanding in ways that are more effective than words alone. The Graphic Literacy assessment measures skills that individuals use when they read and comprehend graphical resources to solve work-related problems.

At a more concrete level, the construct is defined in terms of both the complexity of the graphical resources and cognitive processes required to solve a problem. WorkKeys has defined the construct as the interaction of the complexity level of the graphic and the cognitive skill required by the specific task. Scores on the Graphic Literacy assessment should be interpreted in terms of whether an examinee can solve problems based on the interaction of the graphic complexity and the cognitive demand of the tasks.

11.4.4 Graphic Literacy—Field Test Sampling

Graphic Literacy was theoretically defined and supported through analyses of professional literature on the use of graphics to convey information, data collected by ACT through its job profiling services, and through input provided by a panel of Subject Matter Experts (SMEs).

ACT engaged in a series of three field test studies to evaluate the psychometric properties of initial Graphic Literacy forms. For each of the field test studies, ACT attempted to recruit samples that were representative of the WorkKeys test population. In recruiting for the field test studies, ACT was cognizant of recruiting a sufficient number of adult test takers due to the workforce orientation of the assessment. Table 11.1 provides a comparison of the percentages of test takers from the WorkKeys test population (2013–2014) to the three field test samples.

Table 11.1: Comparison of WorkKeys Test Population and Field Test Samples by Student/Adult, Gender, and Ethnicity

Group	WorkKeys Test Population	Field Test #1 Sample	Field Test #2 Sample	Field Test #3 Sample
Age Groups				
High School Age	40.6%*	60.7%	60.9%	43.6%
Adults	59.4%	39.3%	39.1%	56.4%
Gender Groups				
Women	46.0%	52.6%	53.0%	54.7%
Men	54.0%	47.4%	47.0%	45.3%
Ethnic Groups				
White Examinees	58.0%	60.7%	61.0%	63.4%
African-American Examinees	21.2%	17.4%	17.7%	16.9%
Hispanic Examinees	8.2%	6.7%	6.6%	8.1%

Note. The WorkKeys test population percentages are based on examinees self-identifying with a specific group during the testing period from July 1, 2013 and June 30, 2014.

*Based on test-takers who reported their age as 20 and below.

The field testing was designed to (a) determine an acceptable time allotment for testing, (b) develop a standardized score scale that was interpretable and could be applied for developing subsequent Graphic Literacy forms, (c) evaluate model-data fit for the three-parameter logistic (3-PL) IRT (Hambleton & Swaminathan, 1985), and (d) evaluate the mode effect on test scores (paper vs. online administration).

11.4.5 Measuring Graphic Literacy

Testing Time. ACT conducted two separate studies to assess the appropriate amount of time examinees should be allowed to complete the Graphic Literacy assessment. In the first study, examinees were assigned to take either the online or the paper version of the assessment. They were also assigned to have either 55 or 60 minutes to test. (ACT selected these two testing times for the study based on the amount of time examinees required in small sample studies completed in ACT's Cognitive Laboratory.) Based on the study, ACT wanted to determine (a) whether the test mode (online vs. paper) required the same or different time allotments, and (b) the appropriate amount of time to provide examinees in testing.

ACT defined the assessment as a power test, which is a test that provides examinees sufficient time to answer all items or tasks, and the speed by which an examinee solves the items or tasks should not affect test scores. In a speeded test, examinees' ability to work quickly through the items or tasks is considered a relevant facet of the construct. For Graphic Literacy, whether examinees work through the items quickly or slowly, their speed should not affect their scores. Any effect that speed might have on test scores is interpreted as construct irrelevant variance. Regardless, ACT establishes an assessment time limit because administrators at test centers need to be able to schedule examinees for testing and a time limit provides structure for examinees. (In cases where a test taker requires extra time due to a documented need, ACT and the test center are able to provide the additional time period. See Chapter 5 for more information on accessible test features.)

ACT evaluated test speededness by analyzing the percentage of examinees who were able to answer the last item on the assessment and the omit rate of items across the complete assessment. Over 500 examinees participated in the first field study.

From the first field study, ACT found that examinees took approximately the same amount of time to complete the assessment regardless of mode (online vs. paper). They also found that the completion rates for the assessment were only slightly different for the 55-minute time limit compared to the 60-minute time limit. For online testing, where ACT was able to track the amount of time examinees spent on each item, examinees in the 60-minute condition used an average of one additional minute for testing than examinees in the 55-minute condition. Ninety-five percent of the examinees in both conditions completed the assessment in 47 minutes or less. The omit rate for the final test item in both conditions was less than 1 percent. For examinees in the 55-minute condition, 94% either strongly agreed or agreed with the statement that they had sufficient time to test. For examinees in the 60-minute condition, 98% either strongly agreed or agreed with the statement that they had sufficient time to test.

Based on these results, ACT concluded that for both online and paper administration, the allotted testing time should be 55 minutes. In the second field study, ACT continued to evaluate testing time. The findings from the second study confirmed the conclusion of the first study; 55 minutes was a sufficient amount of time to allow examinees. With a 55-minute time allowance, the findings from the field study analyzing completion rates, omit rates, and survey responses indicate that speededness should not affect examinees' Graphic Literacy scores.

Scale Scores. Results from the field test studies related to the establishment of the scoring scale are presented in Chapter 8.

Score Reliability and Generalizability. Score reliability or generalizability is essential for interpreting and using scores derived from any measure (Kane, 2013). For test scores to be interpretable, they must be consistent across various testing occasions and across different forms of an assessment. Chapter 10 summarizes analyses of field test data to provide estimates of score reliability and measurement error. Based on the analysis, Graphic Literacy scores are reliable and generalizable (i.e., measurement error is minimal) for use in estimating examinee skill levels.

Mode Effects. ACT develops graphic resources and items to be used for both paper and online delivery. ACT conducted a field study to determine if scores achieved when taking the Graphic Literacy assessment online were comparable to scores achieved when taking the assessment on paper. ACT evaluated the mode effects at the item level, by comparing the similarity of item p-values, point biserial correlations, IRT item parameter estimates (a, b, and c parameters), and omit rates. The evaluation of the different item statistics indicated that examinees responded to the items similarly across modes. Differential Item Functioning (DIF) analyses were conducted to determine if examinees of similar ability had similar probabilities of answering an item correctly in different modes. ACT also evaluated the mode effect by analyzing raw scores across the two modes. Examinee raw scores across the two modes were nearly identical, as was the raw score variance. ACT further analyzed the mode effect by analyzing the factor structure of the assessment delivered in two different formats. Overall, ACT concluded that the mode effect was negligible. (For greater detail regarding the mode analyses, see Chapter 9.)

11.4.6 Graphic Literacy—Evidence Based on Internal Structure

ACT analyzes WorkKeys assessment item data using a unidimensional Item Response Theory model (Hambleton & Swaminathan, 1985; Lord, 1980). WorkKeys has traditionally applied unidimensional IRT models to make inferences about examinee proficiency based on observed item scores. This requires the assumption that observed score variance be attributable to a single underlying factor.

Graphic Literacy Dimensionality. ACT applied exploratory factor analysis (EFA) to assess dimensionality for the Graphic Literacy assessment. EFA uses an inter-item correlation matrix to identify the factors underlying observed item variance. In the analysis, four criteria are evaluated to assess dimensionality. A scree plot of eigenvalues is one of the most commonly used tools for determining test dimensionality. When there is only one eigenvalue above the "elbow" in the scree plot, this indicates a unidimensional test. Hatcher (1994) suggested that a factor should be retained if it accounted for at least 10% of 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) maintained that the first factor is relatively strong if the factor difference ratio index (FDRI) (Johnson, Yamashiro, & Yu, 2003) is greater than 3. FDRI is the ratio of the difference between the eigenvalue of the first factor and the second factor to the difference between the eigenvalue of the second and the third factor.

The EFA was conducted using data from the second field study. Over 2,100 examinees participated in the second field study taking two forms of the assessment. The participants were representative of the WorkKeys testing population in that approximately 60% of the examinees were high schoolers and 40% were adults; approximately 53% of test takers were women and 47% were men.

Figure 11.1 presents the scree plot derived from the correlation matrix of item scores on the Graphic Literacy base form. Table 11.1 summarizes the eigenvalues and FDRI for both test forms. Figure 11.1 reveals that the "elbow" appears immediately after the first eigenvalue. Table 11.1 indicates that the percentage of variances accounted for by the first factor is approximately 30% and, for the second factor, it is less than 10%. Additionally, Table 11.2 indicates that the FDRI is 18.32 or significantly greater than 3. These findings consistently indicate that a single factor underlies item scores on the Graphic Literacy assessment.
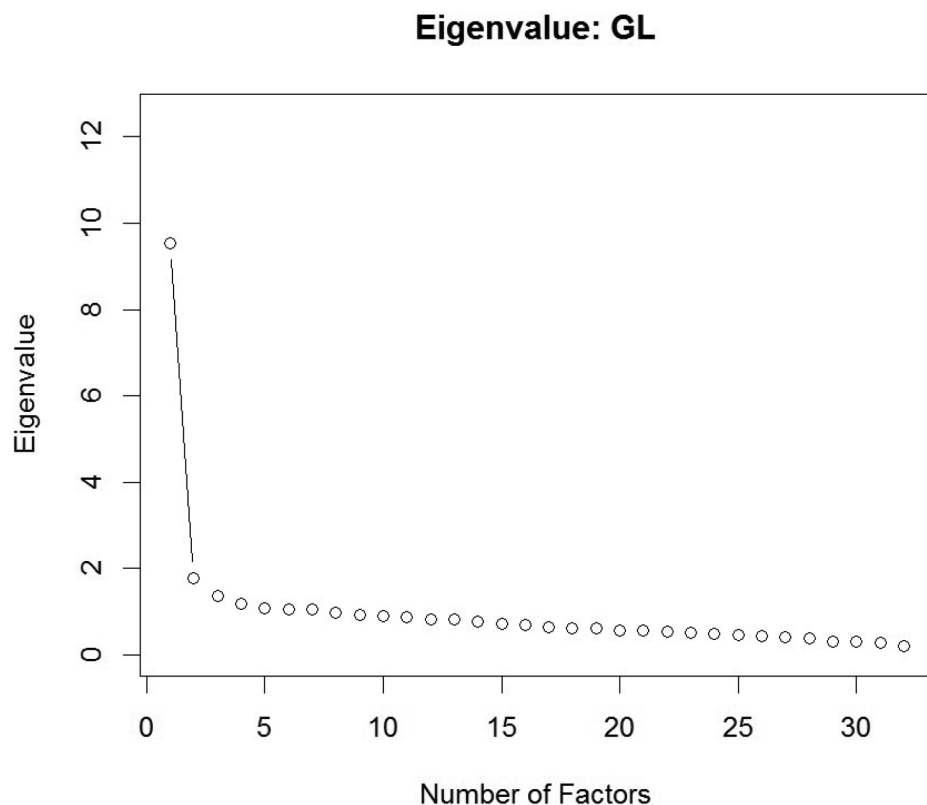
Eigenvalue: GL

Figure 11.1: Graphic Literacy—Eigenvalue Scree Plot

Table 11.2: Summary of Eigenvalues and Factor Difference Ratio Index (FDRI)

Factor	Eigenvalue	Difference between Eigenvalues	FDRI
1	9.53 (29.8%)		
2	1.78 (5.6%)	7.75	
3	1.36 (4.2%)	0.42	18.32

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

IRT Modeling – Local Item Independence. The 3PL IRT model assumes that items are locally independent, which means that examinees' scores on different items in an assessment are statistically independent of each other after controlling for the examinee's ability. For the assumption to be met, examinees' responses to one item cannot be affected or prompted by other items. When local independence is achieved, the probability of any pattern of item responses for an individual is the product of the probability of the correct response for each individual item based solely on examinee ability (Hambleton & Swaminathan, 1985). The Graphic Literacy assessment design includes a series of graphics where each graphic has two or three items associated with it. This raises the possibility of statistical dependencies between items that share graphics, so determining whether items and item sets are independent or dependent is critical for applying an IRT model to Graphic Literacy.

ACT used Q_3 (Yen, 1984) to evaluate the local item dependence for the items within a Graphic Literacy form. 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 a 3-PL IRT model. In this study, items not in the same set were assumed to be locally independent. The Q_3 indices for all items that were not in a set were computed and served as the baseline. Then, the Q_3 for the items within a set were compared to the baseline to evaluate whether the items in a set were more dependent than the items not in 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 was considered to be dependent.

ACT used test data from the second field study to generate the Q_3 matrix to evaluate whether local item dependence was present. Analyzing the Q_3 matrix for the Graphic Literacy scaling form indicated that the items within a set do not show higher correlations than those items in the form that were independent. Consequently, after reviewing all of the items in the form, it was concluded that no compelling evidence of item dependence existed. Thus, the items on the Graphic Literacy assessment met the assumption of Local Independence.

11.4.7 Graphic Literacy—Evidence Based on Response Processes

ACT conducted two cognitive laboratory studies to analyze the response processes that test takers use to solve Graphic Literacy tasks. ACT implemented a think-aloud protocol (Van Someren, Barnard, & Sandberg, 1994) in the first study to gather data for the purpose of gaining a more thorough understanding of the construct. In the second study, ACT utilized eye-tracking software to identify the item features participants focused on and the level of effort they exerted as they tried to solve the graphic literacy tasks (Beatty, 1982; Marshall, 2002; Porter, Troscianko, & Gilchrist, 2007). In the second study, participants answered items constituting a complete form of the Graphic Literacy assessment, built to the content specifications defined in Chapter 3. The purpose of both studies was to elicit evidence based on test takers' response processes to support the interpretation and use of Graphic Literacy scores as indicators of ACT's graphic literacy construct.

Think-aloud Study and Findings. ACT cognitive labs used think-aloud protocols in the initial stages of development to gain greater understanding of graphic literacy. Twenty-one individuals participated in the think-aloud protocols, including 10 high school students, three college students, and eight adults currently holding jobs in the workforce. Participants were recruited from two different high schools, one university, and several places of employment. Participants were videotaped as they worked through the assessment items.

ACT staff members reviewed the videotapes and coded their response processes to specific graphic literacy skills. Participants in the think-aloud study provided evidence supporting the proposed model (defined in Chapter 2) of participants using cognitive steps to solve problems. Overall, the participants tended to start by finding or locating the needed information within the graphic. Based on the task, some participants were able to take that information and then performing additional cognitive tasks. For example, they might find several data points and then analyze them to determine if a trend existed. Additionally, a few of the participants were able to take this information and work beyond the data to predict or justify a decision. Participants who scored well tended to first make sense of the entire graphic before looking at the questions associated with it. This strategy enabled them to answer questions in a graphic set more quickly because of their understanding.

Eye-tracking Study and Findings. After Graphic Literacy items had been developed and pre-tested, eye-tracking research was conducted on a newly constructed form of the Graphic Literacy assessment. The purpose of the eye-tracking study was to analyze gaze and pupillometry data to determine how participants interacted and engaged with the graphics and questions to solve the test items. ACT proposed that differences would exist in how participants with high graphic literacy would solve tasks compared to participants with low graphic literacy. The hypothesis was based on observed eye movement differences comparing experts and non-experts in problem solving tasks (Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, 2010; Obersteiner, Moll, Beitlich, Cui, Schmidt, Khmelivska, & Reiss, 2014).

Thirty-eight individuals participated in the eye-tracking study. They included 16 high school students, four college students, and 18 workforce age adults. Participants were given instructions on the protocol to answer each item as they would as if they were taking the test to obtain a usable score. All participants completed the assessments in less than the allotted time.

ACT collected eye-tracking gaze data from the participants as they worked through the test form. Gaze data included fixations, saccades, sequence, and heat maps, as well as pupillometry (Langenfeld, Thomas, Zhu, & Morris, 2019). Heat maps show the location, order, and time spent looking at locations on the screen. By using data gathered through heat maps, it is possible to infer the thought patterns used by a participant to derive a response. For example, the heat map may indicate that the participant initially spent time studying the graphic and then proceeded to the question. It also provides information on the section of the graphic on which the examinee focused. For example, did the examinee focus on the critical information, or was he or she distracted by other information (Djamasbi, 2014)?

Pupillometry is the measure of pupil size and reactivity. Through observing individuals solve mental tasks, researchers learned that, as humans become more highly engaged in cognitive activities and exert greater effort, their pupils dilate or enlarge. As a result, cognitive research now uses pupillometry as a measure of cognitive engagement and effort (Marshall, 2002; Porter, Troscianko, & Gilchrist, 2007). The Index of Cognitive Activity (ICA) (Marshall, 2002) measures small reliable changes in pupil size stemming from cognitive activity (Marshall, 2002). It provides a measure of the instantaneous cognitive load based on rapid changes in pupil dilation caused by mental effort and minimizes the slower pupillary response caused by changes in the intensity of light (Bartels & Marshall, 2012; Marshall, 2002; Marshall, 2007). ACT defined Total Cognitive Effort (TCE) as the ICA multiplied by elapsed time.

The findings from the analysis of eye-tracking data for each item provided support for the graphic literacy construct. Analysis of sequence maps and areas of interest showed that individuals who answered an item correctly not only looked where predicted based on the graphic literacy skills, but also that they generally followed the skill list path describing the optimal way to solve the task (Langenfeld et al., 2019; Thomas & Langenfeld, 2017).

Gaze pattern data were analyzed to compare participants who achieved high scores (Levels 6 and 7) to participants who achieved low scores (Level 3 or below). The eye-tracking data analysis supported the position that individuals who were high performers utilized different strategies and interacted with the graphics differently than low performers. When a new graphic set was presented, high performers tended to spend significantly more time on the contextual information for the new graphics compared to low performers. High performers were more likely to fully utilize the information presented in the graphic; whereas, low performers struggled when multiple graphics were required to answer an item, when a graphic contained two vertical axes representing two different variables, and when it was critical for the test taker to apply the information presented in the legend. High performers appeared to recognize the importance of this graphical information and utilized it to answer items correctly.

TCE for each participant was calculated for each item. ACT proposed that, as items became more complex regarding graphic complexity, cognitive level, and performance level, the test takers' TCE would increase. The TCE distribution was highly skewed; therefore, ACT adopted the median item TCE as the best measure of central tendency. TCE summary statistics were calculated using the mean of the item median TCE for graphic complexity, cognitive level, and performance level. Table 11.3 provides the TCE summary statistics. For all three model factors, from the lowest levels to the higher levels, the TCE increased with each incremental level. For all three factors, the mean TCE for the highest-level items was three times greater than the mean TCE for the lowest-level items. Proceeding from the lower-level items to the higher-level items, the variance tended to also increase. Reviewing participant TCE measures, the increase in variance is attributed to two of the low performers disengaging as items became more difficult.

Table 11.3: Total Cognitive Effort (TCE) Description Statistics for Items by Graphic Complexity, Cognitive Process Steps, and Performance Levels

	n^a	Mean	SD
Graphic Complexity			
Simple (1)	3	2437	606
Low Moderate (2)	8	5092	1986
High Moderate (3)	9	6493	3385
Complex (4)	12	9335	4647
Cognitive Process Steps			
One-step	6	3157	1511
Two-step	20	6386	3842
Three-step	6	10,306	3002
Performance Levels			
Level 3	4	2512	517
Level 4	8	5054	1401
Level 5	8	6839	3619
Level 6	7	7272	4977
Level 7	5	11,555	2448

[a]A single form of the assessment was used for the eye tracking study; as a result, N represents the number of items classified in specified model category.

The correlations of item median TCE to levels of graphic complexity, cognitive process steps, and performance levels were 0.55, 0.56, and 0.62, respectively. The correlations of item TCE medians to the different assessment factors provided support for the skills-based cognitive model, by indicating that participants' TCE increased as they responded to items that had more complex graphics, required more cognitive process steps, and were at higher performance levels.

The eye-tracking analysis findings provide support for ACT's construct definition and potentially can assist individuals who want to improve their graphic literacy skills. ACT defined the Graphic Literacy assessment construct and item levels as an interaction of the graphic complexity with the cognitive complexity of the task. The findings from the eye-tracking study supported for this definition. Additionally, ACT found that high performers tend to start by studying the graphic to gain an understanding of its overall purpose. After gaining an understanding of its overall purpose, high performers are able to focus on the critical information needed for solving the problem. On the other hand, low performers tend to start with the question and then search the graphic for something relevant. This strategy frequently has them missing critical information. High performers have a more complete understanding of graphics and utilize information contained in labels, legends, and scales.

11.4.8 Graphic Literacy—Evidence Based on Relations to Other Variables

The *Standards* identifies evidence based on the relation of assessment scores to other variables as a source of validity evidence. This type of evidence includes the relationship of scores on the assessment to other assessment scores, and the strength of the relationship of the assessment scores with future relevant behaviors. In terms of the *Graphic Literacy* assessment and its associated claims (see Chapter 1 and Section 11.6 of this chapter), the analysis of the relationship of *Graphic Literacy* scores to workplace performance and training programs is critical.

Chapter 11 presents data analyses from studies evaluating the relationship of scores achieved on the ACT WorkKeys *Locating Information* assessment to workforce performance ratings and educational training programs. (See Sections 11.5.2-11.5.5.) The *Graphic Literacy* assessment was developed in part from the *Locating Information* construct.

ACT is currently seeking to collaborate with businesses, industries, and community colleges to gather performance and educational data to evaluate the relationship of scores on *Graphic Literacy* to important outcome variables. In the next year, ACT plans to complete several studies and report directly on validation evidence based on relations to other variables.

11.4.9 Graphic Literacy—Standard Setting

The goal of the standard setting process is to translate the Graphic Literacy PLDs into a set of cut scores. Essentially, the process is designed to identify a point on the score scale where examinees who score at or above the point have demonstrated that they can perform certain skills, and examinees who score below the point have not demonstrated that they can perform those skills. To provide data and input for setting the cut scores, ACT recruited an external panel of SMEs consisting of educators and business people, some of whom have used WorkKeys products.

ACT implemented the Mapmark standard setting procedure (Schulz & Mitzel, 2005) with Whole Booklet Feedback to establish the standards or cut points for each of the five Graphic Literacy score levels. The Mapmark procedure, which was first implemented by ACT for the Grade 12 mathematics National Assessment of Educational Progress (NAEP) achievement level setting project, builds on the widely used Bookmark method (Lewis, Mitzel, & Green, 1996). The Bookmark method was introduced in 1996 and has gained wide acceptance in state educational assessment programs and in professional certification and licensing programs. Mapmark supplements the Bookmark method by including spatially-representative item maps (Masters, Adams, & Lokan, 1994).

To establish the cut points for each of the five score levels, ACT led the SMEs through three rounds of ratings and reviews. In Round 1, the SMEs applied the Mapmark procedure to establish the initial cut points. The initial cut points were refined in Round 2 by providing the SMEs with whole booklet feedback in the form of examinee test answer sheets. In Round 3, ACT provided the SMEs the estimated level score distribution based on data from the second field study to assist them in finalizing the cut points.

Chapter 8—Scores and Score Scales—provides a complete description of the Standard Setting process.

11.4.10 Graphic Literacy Measurement—Summary

Based on the data analysis presented in Section 11.4, ACT has provided psychometric support for the interpretation and use of Graphic Literacy scores. This support was accomplished by starting with the information and data that ACT had gathered through 20 years of conducting job analyses and profiling various jobs requiring graphic literacy skills. The information and data were then supplemented by a thorough review of the professional literature around the use of graphics both in education and the workforce. The external SMEs further assisted ACT in refining the construct definition of graphic literacy and the development of exemplary items.

With the development of the initial forms of assessment, ACT then conducted field tests to learn more about graphic literacy and the assessment. In addition to traditional psychometric analyses, ACT conducted a series of cognitive labs to gather qualitative data on how examinees work through these types of tasks.

The preponderance of the research and data analyses indicated that the Graphic Literacy assessment provided a reliable measure of a unidimensional construct built around the concept of examinees working through problems requiring different levels of graphic literacy. The analysis gathered by having external SMEs evaluate the construct and exemplary items indicated that ACT had appropriately defined graphic literacy. The analysis of examinee responses to the testing materials in the cognitive labs provided evidence that the assessment was measuring their ability to solve graphical problems. The analysis of field test responses assisted ACT in identifying the appropriate amount of time for testing. Field testing also allowed ACT to conclude that the assessment was a measure of a unidimensional construct and that construct irrelevant variance was minimal. Analysis of field test data further demonstrated that scores achieved taking the assessment by paper administration were comparable to scores achieved by online administration. Lastly, the analysis found that scale scores and level scores earned on the assessment were reliable.

11.5 Graphic Literacy—Primary Claims and Relevant Evidence

The purpose of the WorkKeys system is to help build a high-performance workforce by connecting job skills, training, and testing in a manner that benefits both employers and employees. WorkKeys also assists educators in identifying skill gaps between student skills and employment needs, so that they may better address the gaps and thereby improve students' employment prospects.

The three primary claims articulate how scores from the Graphic Literacy assessment may provide actionable information to examinees, employers, educators, and workforce development officials to make these connections. The claims differ in who is the focus of the claim and how score information may be used to accomplish the intended result.

The focus of Claim #1 is the examinee or person seeking employment. Scores on the Graphic Literacy assessment are related to workplace success. In other words, an examinee who scores at a prescribed level (as defined by a Job Profile) will have a greater probability of achieving success in corresponding jobs (based on levels established through a Job Profile) than an examinee who did not score at the

prescribed level. Additionally, examinees who score at higher levels on the Graphic Literacy assessment will have a higher probability of obtaining jobs with greater responsibilities and wages. Claim #1 provides the structure for evaluating how high scores on Graphic Literacy may help an individual in the labor market.

The focus of the second and third claims is the employer or business. Scores on the Graphic Literacy assessment are related to workplace success in ways that will result in improved business productivity and efficiency. Claim #2 states that, if a business determined the Graphic Literacy assessment scores required for specific jobs through a rigorous job analysis or Job Profile, and if the business then hired people who achieved those scores, the productivity gains provided by the new employees would be greater than if the business had not used the assessment scores to help select employees. Claim #3 states that, if a business follows the hiring process outlined for Claim #2, the business would experience less employee turnover (i.e., more new hires retained) than if the business had not used the assessment to help select employees.

Claims #1 and #2 can be supported by the development of the content-related and construct evidence provided in Section 11.4. Additionally, they can be supported through the analysis of outcome data. Claim #3 requires the analysis of employee turnover rates to be plausible. ACT has embarked on a series of outcome studies collecting data from employers and educators to assess the extent that the claims are plausible. With Graphic Literacy being a new assessment, ACT is in the process of working with businesses and states to collect the necessary data. Following the collection and analysis of the data, ACT will publish the findings and update the technical manual. In the meantime, this section of the technical manual presents information and data derived from the Locating Information assessment (the predecessor to Graphic Literacy), the WorkKeys NCRC, and meta-analyses relating cognitive assessments to workplace performance.

ACT has amassed substantial evidence to support the use of the WorkKeys NCRC and the former assessment Locating Information. Clearly, the Graphic Literacy assessment is significantly different from the former assessment in content and construct and requires additional evidence. However, previous research on Locating Information should not be summarily dismissed. The evidence based on Location Information suggests both the appropriateness of the test content and its relationship to critical outcome variables. The updating of the content and the revision of the specifications have led to better measurement of the construct. The evidence suggests relationships with outcomes should not change substantially as a result of the updated content and improved measurement. The next four sections summarize much of this evidence.

11.5.1 Graphic Literacy—Evidence Based on Test Content

Evidence based on content comprises one source of evidence to establish the validity of test score interpretations and uses (AERA et al., 2014). Content evidence often comprises the first line of evidence to support employment selection practices. The *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission [EEOC], Civil Service Commission, Department of Labor, & Department of Justice, 2000), the *Standards* (AERA et al., 2014), and the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial Organizational Psychology [SIOP], 2003) all describe the need to demonstrate that knowledge and skills in employment measures should

be demonstrably linked to work behaviors and job tasks. Both the *Standards* (2014) and the *Principles* (2003) suggest that expert judgment can be used to determine the importance and criticality of job tasks and to relate such tasks to the content domain of a measure. This process is commonly conducted through a job analysis that identifies the tasks required for performance on a job and subsequently for the development of the content blueprint and item development to ensure content validity (Cascio, 1982; Dunnette & Hough, 1990). The Graphic Literacy assessment was designed to assess foundational skills and skill levels associated with many jobs. As such, the content-related validity evidence for the assessment was originally established by the SMEs across numerous jobs that aligned the Graphic Literacy skills and PLDs to specific tasks and job behaviors for a particular job.

ACT applies a job profiling procedure that focuses on the skills and behaviors present across the ACT WorkKeys assessments. It is a multi-step process that includes the creation of one or more groups of SMEs who are typically job incumbents or supervisors. An ACT-trained and authorized job profiler conducts the profiling procedure. Each profile that is conducted represents a content validation study at the organizational level.

The job profiling process involves several steps to establish a link between the PLDs and the requirements of a particular job. Ideally, the SMEs participating in the job analysis comprise a representative sample across a variety of demographic variables (e.g., race, ethnicity, gender, geographic region).

The process begins with a task analysis where the group of SMEs generates a task list that accurately represents the job at an organization and to rate each task in terms of its importance. Figure 11.2 details the steps in the job profiling procedure where tasks and skills are identified leading to the completion of the job profile.

Equally important is the skill analysis where the SMEs review each skill measured by the Graphic Literacy assessment. Once the SMEs understand the definition of the skill and have determined its relevancy to the job, they independently identify the important tasks on the Final Task List that require the skill. They also identify the ways in which a task uses an identified skill. After discussing the relationship of the skills to the tasks, only those tasks identified as important by a majority of the SMEs are included in subsequent discussions, and only those tasks are used to determine the level of skill required for the job through a consensus process.

As part of the skill analysis segment, the SMEs use successive approximation to determine the skill level required for the final set of tasks. Each skill level denotes a level of difficulty, with the lowest level representing the simplest of tasks related to the skill construct and the highest level representing the most complex. The SMEs typically begin with the lowest skill level. They then determine whether the job requires skills at, above, or below the level described. If the SMEs determine that the skills required for the job are higher than skills described in a level, they proceed to the next higher level; if they determine the required skills are lower, they review the next lower level. If they determine that the skills are about the same as the level they are reviewing, they are still shown the next higher level before confirming agreement between skills and a designated level to confirm their judgment.

No decision is reached until the SMEs have considered a range of skill levels: those skills they have identified at the required level, at least one level above it, and at least one level below it (unless they have chosen the highest or lowest level available).

The process described in this section is documented by the job profiler in a content validity report that is provided to the client. Currently, ACT WorkKeys clients have completed over 21,000 job profiles.

```
┌─────────────────┐
│ Initial Task List│
│  Preparation     │
│  (With SMEs)     │
└─────────────────┘
         │
         ▼
┌─────────────────┐
│  Task Analysis   │
└─────────────────┘
         │
         ▼
┌──────────────────────┐      ┌──────────────────────┐      ┌──────────────────┐
│ • Edit initial task  │      │ Skill Analysis       │      │      SME         │
│   list               │ ───▶ │                      │      │  Demographic     │
│ • Rate tasks for     │      │ • Define ACT WorkKeys│      │  Information     │
│   importance         │      │   Skill              │      └──────────────────┘
│ • Finalize task list │      │ • Identify tasks     │
└──────────────────────┘      │   requiring skill    │
         ▲                    │ • Use consensus      │
         │                    │   process to determine│
         │                    │   the level of skill │
         │                    │   required           │
         │                    └──────────────────────┘
         │                              │
         │                              ▼
┌──────────┐           ◇ Replicate ◇        ┌──────────┐      ┌──────────────────┐
│   Yes    │ ◀──────── ◇    or     ◇ ─────▶ │    No    │ ───▶ │  Completion of   │
└──────────┘           ◇ Reconcile ◇        └──────────┘      │ the Job Profile  │
                                                              └──────────────────┘
```
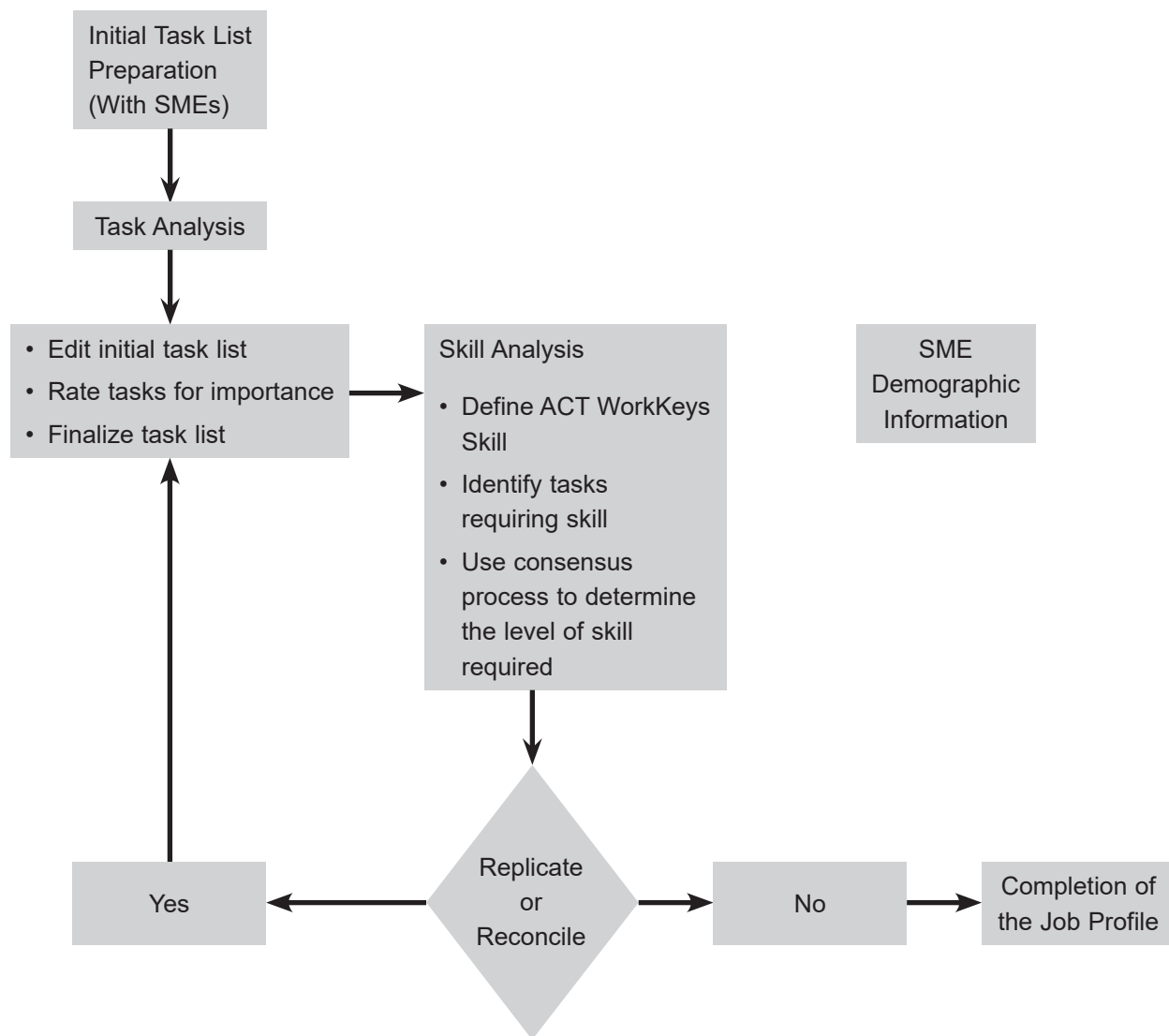
Figure 11.2: Job Profile Process Designed to Align Job Tasks to Skill Levels

11.5.2 Locating Information—Evidence Based on Relationships to Work-related Variables

LeFebvre (2016) summarized 19 workplace outcomes studies for the WorkKeys suite of assessments, including the Locating Information assessment. Those studies examined the relationship between scores on the Locating Information assessment and job performance with sample sizes ranging from 13 to 1,216 participants. The studies included health care service providers, manufacturing workers, motor coach drivers, and students in career technical programs. She concluded that individuals who achieved higher Locating Information scores tended to receive higher job performance ratings and were retained

for longer periods in their jobs. Further, individuals who achieved higher Locating Information scores had fewer work-related safety incidents, had lower rates of absenteeism, experienced fewer customer complaints, and had lower turnover rates. Table 11.4 presents a summary of the validity coefficients between scores on the Locating Information assessment and different outcome measures. Table 11.4 also presents the relationship of composite scores from Locating Information, Applied Mathematics, and Reading for Information with different outcome measures (LeFebvre, 2016).

Table 11.4: Correlations between Scores on the WorkKeys Locating Information Assessment and Different Outcomes

WorkKeys Assessment	No. of Studies	Sample Size or Range	Validity Coefficient*	Outcome Variable
Locating Information	1	2,162	.21	Career Tech Course Grades
Locating Information	1	96	-.33	HRIS Data—Turnover
Locating Information	1	96	-.22	HRIS Data—Absenteeism
Locating Information	1	96	-.11	HRIS Data—Safety Incidents
Locating Information	1	96	-.11	HRIS Data—Customer Complaints
Locating Information	13	13–314	.16	Overall Job Performance—Supervisor Ratings
Composite of LI, AM, and RFI	3	68–951	.29	Overall Job Performance—Supervisor Ratings
Composite of LI, AM, and RFI	1	951	.25	Career Tech Course Grades

* When multiple studies are included, the table presents the median validity coefficient for the set of studies.

Hendrick and Raspiller (2011) analyzed data from 12 different companies that used the WorkKeys NCRC to determine its effect on worker retention. They found that businesses using the WorkKeys NCRC as part of the hiring process saw their retention rates increase from 84% to 93%. Further, they found that the higher the WorkKeys scores, the more positive the effect on retention. In follow-up interviews with hiring managers, Hendrick and Raspiller (2011) learned that using the WorkKeys NCRC as part of the hiring process also resulted in new employers requiring less training time and less of a need to be closely supervised.

Greene (2008) analyzed the use of the WorkKeys cognitive assessments in business and industry in North Carolina. She surveyed employers of small and large companies focusing primarily on the use of the WorkKeys NCRC. She found that employers viewed the WorkKeys NCRC as a useful tool to assist in hiring. In using the WorkKeys NCRC to assist in hiring decisions, 60% of hiring managers agreed that training time was reduced, 52% agreed that worker turnover rates were reduced, 40% agreed that company teamwork increased, and 36% agreed that re-work was reduced. In follow-up interviews, the hiring managers stated that the WorkKeys NCRC provided a pre-employment screening device that allowed them to select workers who learned job tasks more quickly, reached production targets more quickly, and produced better overall quality work.

These studies specifically analyzed scores on the Locating Information assessment or levels achieved on the WorkKeys NCRC to outcome measures, including job performance ratings and grades in career and technical education programs. Other researchers have analyzed measures of cognitive ability and their usefulness in the employment sector. The most reputable of these studies have combined data from many studies and incorporated meta-analysis techniques to draw conclusions.

Prior to the use of meta-analysis and today's understanding of measurement problems associated with outcome variables, researchers believed that validity coefficients varied a great deal from one job to the next. For much of the first 70 years of the 20th century, researchers evaluated employment selection methods by correlating scores on selection tests to measures of job performance. They found that using the same tests for nearly identical jobs often resulted in quite different validity coefficients. They concluded that the differences in validity coefficients stemmed from subtle differences in job requirements resulting in situational-specific validity (Ghiselli, 1966).

Many of the differences reported across different validity studies have been shown to be the result of statistical and measurement artifacts (Schmidt & Hunter, 1977; Schmidt, Hunter, Pearlman, & Shane, 1979). Subsequently, meta-analytic methods were developed to account for sampling error, selection bias, low reliability of criterion measures, and other artifacts. When statistical and measurement artifacts were accounted for, the findings indicated that the variability of validity coefficients was reduced to near zero (Hunter, 1980). The finding that validity coefficients could be generalized across selection methods and jobs made it possible to compare and analyze different personnel selection methods.

In a comprehensive review, Schmidt and Hunter (1998) examined 85 years of research on personnel selection and concluded that the best predictor of job performance and the ability to benefit from job-related training was general cognitive ability. As an update to the 1998 paper, Schmidt, Oh, and Schaffer (2016) evaluated 31 different methods of personnel selection from cognitive ability testing to job interview rating systems to the analysis of handwriting. They concluded that general cognitive ability was the "gold standard" of selection methods, and they then assessed how much additional predictive power was gained by combining other methods with cognitive ability testing.

Schmidt and Sharf (2010) evaluated the three assessments constituting the WorkKeys NCRC. They concluded that "measures of general cognitive ability such as WorkKeys are the most job related (i.e., most valid) predictors of job performance in both the military and civilian workforces" (p. 12). They defined the Locating Information assessment as a measure of technical/problem-solving skills highly relevant to job performance and learning.

Combining Schmidt and Sharf's (2010) results with LeFebvre's summary reveals a median correlation of .29, which appears similar to correlations of the SAT and ACT to first-year college grades. Taking into account selection effects, range restriction, and low reliability of outcome measures, similar to the validity coefficients of the SAT and ACT in predicting student grades, the correlation of 0.29 is a conservative estimate. The disattenuated correlation is likely much greater (Sackett, Borneman, & Connelly, 2008).[1]

11.5.3 Locating Information—Return on Investment

Hunter, Schmidt, and Judiesch (1990) published a ground breaking analysis indicating that the return on investment (ROI) of hiring the best people was potentially large, and for jobs that required complex information processing, it was very large. They utilized meta-analytic methods to evaluate data from several hundred studies involving thousands of employees doing different jobs. They concluded that, for jobs that required low levels of information processing, a person who was in the top 1% of the applicant

pool would be 1.52 times more productive than a person who was at the median of the applicant pool. For jobs that required moderate levels of information processing, a person who was in the top 1% of the applicant pool would be 1.85 times more productive than a person who was at the median of the applicant pool. Lastly, for jobs that require high levels of information processing, a person who was in the top 1% of the applicant pool would be 2.27 times more productive than a person who was at the median of the applicant pool. They concluded that differences in individual productivity were large and businesses that hire the best people tend to experience a competitive advantage. This difference would be particularly pronounced for a business where large numbers of employees were engaged in high levels of information processing.

Mayo (2012) analyzed hiring data for New Options New Mexico evaluating the ROI of using the WorkKeys NCRC as part of the hiring process. Preexisting data for each employer was collected and outcomes compared pre- and post-WorkKeys NCRC implementation. She found that by implementing the WorkKeys NCRC, businesses experienced a 25–75% reduction in turnover, a 50–70% reduction in time to hire, a 70% reduction in cost-to-hire, and a 50% reduction in training time. Overall, she concluded that using the WorkKeys NCRC as part of the hiring process resulted in employers making a minimal investment in order to receive a very large return. Of course, additional studies with a variety of employers are required for ACT to support its claim related to employee retention.

11.5.4 Locating Information—Evidence Based on Relationships to Educational Variables

LeFebvre (2016) reviewed studies that related Locating Information scores to post-secondary educational outcomes (see Table 11.3). In career and technical education programs, individuals who achieved higher Locating Information scores tended to have higher completion rates and earn higher grades. Also, individuals who achieve higher Locating Information scores tended to have higher grade point averages in their postsecondary studies.

Schultz and Stern (2015) studied changes in examinee perceptions of career readiness following the administration of the NCRC assessments to high school students in Alaska. They surveyed students in their junior year of high school and asked them if taking the assessments and reviewing their scores were helpful. Students reported that the assessments assisted them in evaluating their career readiness, were useful in career planning, and caused them to think more seriously about different career options. Most interestingly, scores from the assessments provided students with information that appeared to contradict the feedback they had received from their high school course grades. Whereas nearly 75% of the students reported receiving class grades of A's and B's, and they regarded their skills as strong, based on their WorkKeys scores, slightly more than 50% of the students did not meet the college or career readiness standards.

11.5.5 Locating Information—Regional Data

LeFebvre (2016) analyzed statewide workforce studies where the WorkKeys NCRC was used to assist individuals in finding employment. Using data from workforce development agencies in Indiana, Iowa, Ohio, and southwest Missouri, she found that individuals who achieved higher levels experienced faster time to hire, earned higher wages, and stayed in their jobs longer.

11.6 Graphic Literacy—Evaluation of Claims

The cited studies analyzed data from the Locating Information assessment, the WorkKeys NCRC, and general measures of cognitive ability. As mentioned earlier, the Graphic Literacy assessment constituted one of three assessments of the WorkKeys NCRC. Graphic Literacy was designed building on the information that ACT had collected over the past 25 years from the original Locating Information assessment. The Graphic Literacy construct includes many of the facets of the Locating Information construct, but it built on and extended the Locating Information construct. For Graphic Literacy, test content was updated to better reflect current uses of graphical information in the workforce. Psychometrically, the updated Graphic Literacy assessment met or exceeded the psychometric standards that were used to develop forms of the Locating Information assessment. For these reasons, data collected from the Locating Information assessment can tentatively be used to evaluate the claims, even though ACT is currently collecting outcome data related to Graphic Literacy performance.

From the individual examinee perspective, based on the findings, when score information from the Locating Information assessment and the WorkKeys NCRC were used as part of employment selection or for educational evaluation, it appeared that individuals who achieved sufficient scores on Locating Information tended to experience the following:

- Reduction in time to hire (LeFebvre, 2016; Mayo, 2012)
- Higher wages (LeFebvre, 2016; Mayo, 2012)
- Longer job tenures (LeFebvre, 2016; Mayo, 2012)
- Better job performance evaluations (LeFebvre, 2016)
- Better post-secondary grades and higher career-technical program completion rates (LeFebvre, 2016)
- Information that provided insight useful in evaluating career readiness and career planning (Schultz & Stern, 2015)

The findings from the studies provide evidence supporting Claim #1 that examinees who score at given levels of the Graphic Literacy assessment are more likely to successfully perform in more and higher levels of U.S. jobs than examinees whose scores do not reach that level.

From the employer's perspective, based on the findings, when score information from the Locating Information assessment and the WorkKeys NCRC were used as part of the employment selection process, it appeared that businesses tended to have the following outcomes:

- Higher levels of productivity (LeFebvre, 2016; Greene, 2008; Hunter, Schmidt, & Judiesch, 1990)
- Lower rates of re-work (Greene, 2008)
- Lower turnover rates/higher retention rates (LeFebvre, 2016; Hendrick & Raspiller, 2011; Mayo, 2012; Greene, 2008)
- Less training time (Hendrick & Raspiller, 2011; Mayo, 2012; Greene, 2008)
- Fewer safety incidents (LeFebvre, 2016)
- Less absenteeism (LeFebvre, 2016)

The findings provided initial evidence supporting Claims #2 and #3 that businesses that use the Graphic Literacy assessment as part of the hiring process will experience increases in business productivity reduced worker turnover rates.

From the state and regional perspective, based on the findings of using test scores from the Locating Information assessment and the WorkKeys NCRC to promote local workforce development, it appeared that states and regions that have a large number of workers who have earned high scores and credentials have the following characteristics:

- workers with higher levels of the WorkKeys NCRC tend to be hired more quickly (LeFebvre, 2016)

- workers with higher levels of the WorkKeys NCRC tend to earn higher wages (LeFebvre, 2016)

- workers with higher levels of the WorkKeys NCRC tend to stay in jobs for longer periods of time (LeFebvre, 2016; Hendrick & Raspiller, 2011; Mayo, 2012; Greene, 2008)

As ACT builds up the Work Ready Communities, it is collecting data on economic and business productivity. It is also collecting data on job growth and wages.

11.7 Graphic Literacy—Evidence Based on Consequences of Testing

Kane (2013) defined consequential evidence that should be evaluated and weighed in making decisions about test use. Two critical components of consequential evidence that need to be evaluated are intended outcomes and adverse impact. The intended outcomes of the Graphic Literacy assessment are articulated by the three primary assessment claims. Empirical evidence should indicate that an assessment program achieves its intended outcomes and not unintended negative outcomes. Adverse impact refers to possible performance differences between demographic groups and how decisions derived from scores might adversely affect a specific group. The *Uniform Guidelines on Employee Selection Procedures* (EEOC et al., 2000) defined adverse impact in the area of employment selection.

11.7.1 Intended Outcomes

An evaluation of the three primary claims is presented in Section 11.5. Based on analyses of the Locating Information assessment and the WorkKeys NCRC, it appears that scores from the assessment and levels of the credential assist individuals in finding suitable employment and assisting businesses in finding qualified workers.

With the newly released Graphic Literacy assessment, ACT is collecting outcomes data relating test scores to outcomes such as job performance, successful completion of educational programs, and other evaluative measures.

11.7.2 Adverse Impact

Chapter 12—Fairness—specifically addresses the Graphic Literacy assessment and adverse impact. The chapter defines adverse impact and provides analysis and recommendations to employers regarding fair employment procedures.

When the Graphic Literacy assessment or any WorkKeys assessment is used for pre-employment screening or other employment decisions, employers should conduct a well-documented job analysis that provides appropriate evidence linking the skills required on the job with the skills measured in the assessment. When cutoff scores are used to assist in decision making, they should be established at appropriate levels, and the process for identifying the levels should be clearly documented (AERA, et al., 2014; SIOP, 2003).

11.8 Graphic Literacy—Ongoing Validation

ACT continually collects and analyzes data related to the validation of its products. With the development of the new Graphic Literacy assessment, ACT has begun the process of collecting data and evidence to determine the plausibility of its claims.

As outcome data is collected and analyzed, ACT will publish the findings through research reports and it will be supplementing the Technical Manual. In collecting and analyzing the data, ACT is cognizant of the two main populations served by the Graphic Literacy assessment: adults in the workforce and students in high school, college, or career and technical programs. It is critical that validity evidence is collected and analyzed from both populations to confirm that it meets the needs of both. While specific details of the analyses are dependent on the available outcome data, ACT will analyze the relationships of scores on the Graphic Literacy assessment to critical outcome variables including job performance, job attendance, job retention, and completion of training programs. With sufficient sample sizes, ACT will additionally analyze assessment scores and relationships by demographic groups such as gender, ethnicity, and job types.

Note

1. By applying meta-analytic methods to address range restriction and low reliability of outcome measures, Sackett, Borneman, and Connelly (2008) estimated the disattenuated correlation of general cognitive ability to job performance as 0.47.

Chapter 12

Assessment Fairness

This chapter contains evidence to address assessment fairness related to the WorkKeys® Graphic Literacy assessment. The chapter adheres to the conceptual framework of fairness defined in the *Standards for Educational and Psychological Testing* (AERA et al., 2014). The *Standards* maintain that fairness is a fundamental validity component that requires evaluation throughout the assessment process, from design to test administration to score interpretation and use.

12.1 Test Fairness—Overview

Striving for the fairness of all tests is a professional responsibility and a fundamental component for the validation of test score use. The most recent edition of the *Standards* (AERA et al., 2014) devotes an entire chapter to fairness. The *Standards* divide fairness into four elements, each requiring evaluation: (1) fairness in treatment during the testing process, (2) fairness in access to the construct(s) measured, (3) fairness as lack of measurement bias, and (4) fairness as validity of individual test score interpretations for the intended uses.

Whenever tests are used as part of the decision making process, whether for educational or workforce purposes, it is critical for the testing program to be developed and carried out in a fair and unbiased manner. ACT subscribes to the *Standards* definition of fairness regarding validation and test score usage.

> A test that is fair within the meaning of the *Standards* reflects the same construct(s) for all test takers, and scores from it have the same meaning for all individuals in the intended test population; a fair test does not advantage or disadvantage some individuals because of characteristics irrelevant to the intended construct (AERA et al., 2014, p. 50).

As a component of validation, evaluations of fairness are ongoing, with evidence being collected and reported throughout the life of a testing program. Evidence regarding the fairness of the Graphic

Literacy assessment is not limited to this chapter and is drawn from other chapters in the technical manual. Further, ACT continually collects and analyzes assessment data. As additional data is collected and analyzed, ACT will continually issue reports related to the fairness of Graphic Literacy score interpretations and use.

12.2 Fairness and Test Administration

Fairness during the testing process refers to examinees being assessed in a way that maximizes their opportunity for showing their standing on the construct (Wollack & Case, 2016). In other words, the entire testing process, from test design to scoring, facilitates test takers being able to perform their best and does not adversely affect the performance of an individual examinee or a group of examinees.

The design, development, and scoring of the Graphic Literacy assessment incorporated principals of Universal Design (CAST, 2011) and Evidence-Centered Design (Mislevy et al., 2004) to assist in ensuring fairness to all test takers. ACT developed and documented standardized procedures for the training of test center staff for test administration. They have articulated room and equipment standards in an effort to support standardized and fair conditions for all test takers. They further have defined protocols for the handling of secure information to safeguard sensitive information and protect the privacy of examinees. When unexpected events occur at a test center, the Test Coordinator is required to file an Irregularity Report detailing the event and allowing ACT to make a determination as to whether the event compromised validity. WorkKeys has implemented these procedures as a means to attain fairness for all examinees in the administration of the Graphic Literacy assessment. (See Chapter 4 of the Technical Manual for a comprehensive review of the test administration procedures.)

The Graphic Literacy assessment is administered to examinees in both paper and online formats. To provide evidence of the fairness of scores across both administrative formats, ACT conducted a mode comparability study. ACT evaluated the mode effects at the item and score level. Through the analysis, ACT concluded that modes effects on examinee responses and scores were negligible. (For greater detail regarding the mode analysis, see Chapter 9.)

Although ACT recognizes that the standardization of procedures for test administration is critically important for ensuring that all examinees have an equal opportunity to demonstrate their standing on the construct, ACT also recognizes that flexibility is required to achieve true fairness. When the standardized administrative procedures hinder a test taker from demonstrating his or her standing on the construct, and the test taker provides proper documentation, accommodations to the standardized procedures are considered fair and appropriate.

12.3 Fairness in Access to the Construct Measured

Accessibility in the context of fairness refers to the extent to which examinees can access the knowledge, skills, and/or abilities intended to be measured by the test without being unduly burdened by aspects of the test or test administration that may affect or limit access (Stone & Cook, 2016). For example, an examinee with a visual impairment may not be able to appropriately answer questions on the Graphic Literacy assessment because he or she cannot clearly see the test materials. In such cases,

the lack of accessibility to the test materials creates construct irrelevant variance. A second example might involve an examinee who has been diagnosed with mild Autism Spectrum Disorder (ASD). This examinee may require a special testing location, free from distractions with additional time to complete the test. ACT provides a variety of accessibility options for examinees designed to provide access to the intended test construct, while not violating the construct or giving the test taker an unfair advantage.

The supports provided on the Graphic Literacy assessment are structured along a continuum of increasingly intensive supports designed to meet the needs of all potential examinees. Three levels of accessibility supports are provided: 1) Embedded Tools, 2) Open Access Tools, and 3) 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 in advance. Accommodation supports and tools are the most intensive level of support. Accommodations are available to those who are qualified to use them. Examinees who receive accommodations have a formally documented need and have therefore been identified as qualifying for resources that require expertise, special training, and/or extensive monitoring to select and administer effectively and securely.

All accessibility supports permitted for the Graphic Literacy assessment are designed to remove unnecessary barriers to performance, while not violating or interfering with the measurement of the intended construct. (See Chapter 5 for a comprehensive review of test accessibility features for paper and online administrations.)

12.4 Fairness as Lack of Measurement Bias

Measurement bias has been characterized as "a source of invalidity that keeps some examinees with the trait or knowledge being measured from demonstrating that ability" (Shepard, Camilli, & Williams, 1985, p. 79). Measurement fairness requires that examinees of equal standing on the construct average equal scores on the assessment, regardless of group membership (Sackett et al., 2008). Consequently, measurement bias occurs when score interpretations are differentially valid for any group of examinees. To investigate the potential for measurement bias, ACT evaluates the internal structure of the Graphic Literacy assessment by evaluating the invariance of the items and the overall assessment.

ACT evaluates measurement bias at the item level by applying a Differential Item Function (DIF) procedure (Holland & Wainer, 1993). DIF refers to a set of statistical methods used to identify items that individuals from one demographic group respond to differentially than individuals from another demographic group. DIF occurs when equally able examinees have different probabilities of answering an item correctly based on their group membership (AERA et al., 2014). Items flagged as demonstrating DIF contain statistical evidence of bias; but, statistical evidence alone is not sufficient to conclude measurement bias. ACT WorkKeys has established a process for conducting DIF analyses followed by external reviews of flagged items to determine measurement bias.

In conducting the DIF analyses, ACT compares item responses for two groups of test takers. The two groups are termed the Focal Group and Reference Group. The Focal Group is the group of primary interest, and it includes protected classes under federal employment anti-discrimination laws. The Reference Group serves as the basis for comparison.

For WorkKeys DIF studies, for each item, three separate DIF analyses are conducted using three different comparison group pairs. The group pairs are identified in Table 12.1.

Table 12.1: Differential Item Functioning Evaluations—Group Comparisons

	Focal Group	Reference Group
1	Women	Men
2	African American	White non-Hispanic
3	Hispanic	White non-Hispanic

An item is flagged as containing DIF when one group of matched test takers has a higher probability of answering an item correctly than the other group. Because groups may differ on ability, the DIF analysis matches test takers on ability. (For the WorkKeys DIF studies, ACT matches test takers using their total test score.)

For Graphic Literacy items, the Mantel-Haenszel Delta DIF statistics (Dorans & Holland, 1993) are computed to classify items into three DIF categories: Group A—negligible DIF, Group B—moderate DIF, and Group C—large DIF. (The rules for classifying items into the three groups are presented in Table 12.2.) Items classified as either Category B or C are interpreted as flagged items requiring further review.

Table 12.2: WorkKeys DIF Classification Rules

Group A	MH delta (MHD) not significantly different from 0 (based on Chi Square test, alpha = .05) **or** |MHD| < 1.0
Group B	MHD significantly different from 0 (based on Chi Square test, alpha = .05) and {|MHD| ≥ 1.0 and < 1.5}; **or** MHD not significantly different from 0 and |MHD| ≥ 1.0
Group C	MHD significantly different from 0 (based on Chi Square test, alpha = .05 and |MHD| ≥ 1.5

Note. Classification rules adopted from National Assessment of Educational Progress (NAEP) guidelines (Allen, Carlson, & Zelenak, 1999).

After ACT has analyzed the DIF statistics and classified items into groups A, B, or C, content specialists evaluate all flagged items (Category B and C) for possible bias. Item bias occurs when an aspect of item content places a group at a disadvantage. As a result, to determine if an item contains bias, item content must be thoroughly reviewed by external evaluators. ACT contracts with external evaluators who have training and expertise in cultural anthropology or multicultural education to review the flagged items. The review includes evaluating the item's vocabulary or use of numbers and symbols, the knowledge needed to correctly answer, how accessible the knowledge is to test takers, the cognitive processes required,

and possible test taker misinterpretations that might occur because of differences in life experiences or opportunity to learn. To assist in this review, ACT has identified five questions for use in the item review:

Status: Are the members of a particular group shown in situations that do not involve authority or leadership?

Stereotype: Are the members of a particular group portrayed as uniformly having certain aptitudes, interests, occupations, or personality characteristics?

Familiarity: Is there greater opportunity on the part of one group to be acquainted with the vocabulary? Is there a greater chance that one group will have experienced the situation or have become acquainted with the processes presented by an item?

Offensive Choice of Words: Has a demeaning label been applied or has a male term been used where a neutral term could be substituted?

Other: Are there any other indications of bias?

After the review of each item, the evaluators recommend one of the following actions:

1. Maintain the item as it is currently constructed and continue to use.

2. Send the item back to the content team for revision; evaluator identifies what aspect of the item should be revised.

3. Remove the item from the item pool.

In the case of the decision to maintain the item as it is currently constructed, the evaluator is essentially stating that the item appears to be fair and the DIF flag was a statistical anomaly. In this case, when the item is used on the next occasion, DIF statistics are again generated. If on the second testing occasion, it is not flagged for DIF, it is assumed to be a fair item and is maintained for use on future forms. If on the second occasion, it is flagged for DIF, it is now assumed to be a biased item, and it is marked in the pool and should not be used.

DIF procedures are an effective method for assessing measurement invariance (Liu & Dorans, 2016). Measurement invariance presumes that an assessment is measuring the same construct for all examinees, regardless of group membership.

12.4.1 DIF Analysis Results from Graphic Literacy Field Testing

During the second step in the field testing process, ACT administered the two forms of the Graphic Literacy assessment to 2,266 field test participants. Forty testing sites in 22 states participated. Of the sites, 13 were high schools and 27 were adult testing centers. Approximately, 61% of the examinees were high school students and 39% were adults. Prior to administration, ACT required the field test participants to answer a series of questions related to their age, educational background, gender, and ethnicity. From the information the participants provided, ACT was able to conduct a series of analyses to better understand the fairness of the forms and items. Table 12.3 presents the demographic characteristics by test form for the Graphic Literacy assessment.

Table 12.3: Graphic Literacy—Number and Percent of Field Test Participants by Demographic Group

| Demographic Characteristic | Graphic Literacy | | | | | |
| | Form GS1 | | Form GS2 | | Total | |
	Number	Percent	Number	Percent	Number	Percent
Total Participants	1170	51.6%	1096	48.4%	2266	100%
Men	534	45.6%	471	43.0%	1005	44.4%
Women	601	51.4%	586	53.5%	1187	52.4%
African American	207	17.7%	225	20.5%	432	19.1%
American Indian	24	2.1%	18	1.6%	42	1.9%
Asian American	7	0.6%	4	0.4%	11	0.5%
Hispanic	77	6.6%	51	4.7%	128	5.6%
Native HI/Pacific Islander	1	0.1%	1	0.1%	2	0.1%
Two or more ethnicities	90	7.7%	87	7.9%	177	7.8%
White non-Hispanic	714	61.0%	667	60.9%	1381	60.9%
Prefer not to respond	50	4.3%	43	3.9%	93	4.1%

DIF analyses were generated for comparisons of Women and Men, and for comparisons of African-American and White, non-Hispanic examinees. (The number of Hispanic-American examinees in the field test sample was too small to conduct a DIF analysis.) For the two forms, consisting of 76 items (12 items were common for the two forms) six items were flagged for C-Level DIF. The summary of the DIF analyses for the two forms are presented in Table 12.4.

Table 12.4: Identifications of C-Level DIF items on the two Graphic Literacy Forms

Test	Form	# of Flagged Items	Favored Group
Graphic Literacy	GS1	4	Women, African Americans
Graphic Literacy	GS2	2	Men

The DIF analysis from the field study needs to be interpreted with caution. First, the sample sizes for African Americans for each form was small (n = 207 and n = 225). Due to the limited size of the samples, generalizing from the analysis could result in unwarranted interpretations. As a result, ACT will continue generating DIF analyses for test forms and will continue to update the technical manual as new data becomes available through the national and statewide testing programs. Because DIF methods require

large sample sizes, for other demographic group comparisons, insufficient test sample sizes preclude ACT from conducting additional DIF analyses.

12.5 Fairness as the Validity of Individual Score Interpretations

Fairness of individual score interpretations becomes an important consideration when an assessment score is used as part of a process for making high-stakes decisions. ACT concludes that when a WorkKeys score is used as part of the process to make a decision related to employment, it constitutes high-stakes test use. In these cases, federal rules and procedures should be followed by those using the WorkKeys scores in order for them to have valid, fair, and legal score interpretations.

Federal agencies responsible for enforcing civil rights legislation collectively published the *Uniform Guidelines on Employee Selection Procedures* (EEOC et al., 2000), which regulate how an assessment process may be used to assist in employment selection. If a selection procedure produces adverse impact for a protected group, the procedure should not be used unless the employer is able to demonstrate that the assessment measures skills that are job-related.

Adverse impact occurs when a seemingly neutral employment selection practice has a disproportionately negative effect on members of a protected group (Society for Human Resource Management [SHRM], 2015). Under applicable federal law, adverse impact does not require any intention on the part of the employer to discriminate. The EEOC has defined disproportionally negative effect using two different methods. The first method is frequently referred to as the 80% rule. Adverse impact occurs when the protected group is selected at a rate that is less than 80% of the reference group. The second method is referred to as the statistical significance test. This method attempts to answer the question is the difference in selection rates greater than that which would be expected by chance. It uses Fisher's Exact Test and interprets a difference of two standard deviations as indicating adverse impact.[1]

When a selection process that uses assessment scores shows adverse impact, the burden of proof shifts to the employer. The employer must then demonstrate that the assessment measures job-related skills and is justified by business necessity. Business necessity requires that the employer demonstrate a clear relationship between the selection procedure and job requirements.

Differences in scores is not evidence of test bias. There are many reasons why such differences may exist with a cognitive ability test. Ultimately, a differential prediction study may be conducted to examine test bias and whether there are differences in the slope and intercept of regression equations used to predict an outcome (e.g., job performance, turnover) for demographic groups. This type of analysis can be conducted with applicants if they are later employed or by administering a test to incumbents and using extant data on outcomes to examine test bias. ACT is actively recruiting organizations to participate in both validity and fairness studies to examine these issues. Further, organizations using WorkKeys should conduct a job analysis if they intend to use the Graphic Literacy test scores as a part of their employment decision.

When the Graphic Literacy assessment, or any WorkKeys assessment, is used for pre-employment screening or other employment decisions, employers should conduct a well-documented job analysis

that provides appropriate evidence linking the skills required on the job with the skills measured in the assessment. When cutoff scores are used to assist in decision making, they should be established at appropriate levels, and the process for identifying the levels should be clearly documented (AERA et al., 2014; SIOP, 2003).

The *Uniform Guidelines* along with the *Standards* recognize the use of job analysis coupled with a content evaluation as a means of validating the selection process. ACT developed its Job Profiling process to meet the validation requirements of the *Uniform Guidelines*. Table 12.5 describes the validation requirements of the *Uniform Guidelines* and how ACT's Job Profiling process meets the requirements. (Chapter 11 Section 11.5.1 provides a detailed description of job profiling and content validation.)

Anytime an employer wants to use a WorkKeys assessment as part of the selection process, ACT recommends that the employer utilize the Job Profiling process to assist in determining both the requisite skills and levels for the job. In utilizing Job Profiling, the employer is making the most efficient use of the WorkKeys assessment suite. Further, the employer is also providing job applicants a fair method of selection consistent with the *Uniform Guidelines*.

Table 12.5: Comparing the Requirements of the *Uniform Guidelines* to the ACT WorkKeys Job Profiling Procedure

Uniform Guidelines Requirement	WorkKeys Job Profiling Procedure
A job analysis that generates descriptions of job behaviors, descriptions of tasks, and measures of their criticality	SMEs (Subject Matter Experts participating in the job profiling procedure) establish a list that describes behaviors and tasks with tasks from O*NET API in SkillPro software and customize using information gained from company materials, interviews, and job shadowing. Then, SMEs rate each task for importance and the SkillPro software averages their ratings in order to yield a list of tasks in order of importance.
Demonstrate that the test is related to the described job behaviors and tasks	ACT job profilers report the percentage of important tasks that require the skill (average SME importance ratings of 2.5 or above on a 0 to 5 scale).
Definition of skills in terms of observable work outcomes	Each WorkKeys skill and skill level is defined with specific criteria and is illustrated with multiple workplace examples. SMEs link these definitions to job behaviors and tasks.
Explanation of how the skills are used to perform the tasks or behaviors	SMEs identify important tasks that require the skill under review. SMEs link specific tasks to a skill level and say how the level is used for the tasks.
No decisions can be made based on knowledge, skills, and abilities that can be learned quickly on the job or in training	SMEs identify the skill level required for job entry. New hires should enter the job with this level, not learn it on the job.
Applicants can be assessed on skills for higher-level jobs only if new hires may advance quickly to the higher-level jobs	SMEs identify the skill level required for performing the job on the first day. In addition, they may set a higher skill level for performing the job effective after training.
The rationale for setting the cutoff score must be provided	SMEs identify cutoff skill levels by describing job tasks and linking skill level descriptions and sample items to cutoff levels.
Cutoff scores are to be consistent with normal expectations of workers	SMEs identify the cutoff skill levels based on the normal requirements of the job; not on unusual situations, desired capabilities, or beliefs regarding their own skill levels.
Scores are interpreted as pass/fail only; they must not be interpreted as rank ordering of test takers	WorkKeys scores show that test takers either have the required skill levels or they do not have them. It is not appropriate to rank order test takers based on their level scores.
Documentation regarding the validation process is maintained	ACT Job Profilers present a full report documenting content-related validity evidence, and retain all related worksheets and computer records.

Note

1. In its commitment to fairness in assessment practices, ACT continually monitors examinee scores by group membership. With the recent launch of the updated assessments, ACT currently does not have sufficient volumes of examinee scores to conduct an analysis by group membership. As the updated assessments are administered to more examinees, ACT plans to analyze and publish score distributions for gender and ethnic groups. ACT plans to publish a revision to the technical manual (specifically adding score distributions by groups to Chapter 12) in the next six to twelve months.

Chapter 13

Operational Validation

13.1 Overview

It is important to continuously monitor and review the psychometric properties of operational testing forms after the launch of updated WorkKeys assessments. This chapter reports the analyses and findings from assessments administered from May 2018 to April 2019. Not only does this chapter include demographic statistics from the large sample, but it also includes psychometric analysis results from four operational form administrations as further evidence of test quality. The findings should be interpreted as an extension of the psychometric analyses presented in the earlier chapters based on the field studies since similar analyses were conducted using operational data here. Specifically, the following results are reported to provide additional support to the analyses summarized in the earlier chapters.

- Gender and ethnic group summary;
- Summary statistics for four operational forms, including three computer-based testing (CBT) forms and one paper form;
- Reliability results, including classification accuracy results for the forms;
- Dimensionality evaluation from one of the three CBT forms as an example.

13.2 Examinees

This section summarizes assessment results of different gender or ethnic groups from examinees who took at least one WorkKeys Graphic Literacy assessment from May 1, 2018 to April 30, 2019. (Note that the updated WorkKeys assessments formally launched in September 2017, and the previous assessments were formally retired on May 1, 2018.)

A total of 571,094 examinees were administered one of Graphic Literacy forms during the time period and had valid scores to be included in the analyses. Based on gender and ethnic group distributions, the assessment samples, shown in Table 13.1, are consistent with the test administrations from previous assessments, as shown in Table 11.1. Consistent demographic trends include more male test-takers

than female test-takers, and approximately 50%, 20%, and 10% of the examinees are whites, African-Americans, and Hispanics, respectively. The average scale score earned by male test-takers (78.8) is nearly identical to that earned by female test-takers (78.9). Among the four largest ethnic groups, the order of average score, from high to low, is Asian Americans, whites, Hispanics/Latinos, and African-Americans. For this large sample, the average scale score is 78.8 with a standard deviation of 4.5. The next section will present the findings from four forms administered to part of the large sample.

Table 13.1: Score Summary for Different Gender/Ethnicity Groups for WorkKeys Graphic Literacy Assessment (Tested from 2018/5/1 to 2019/4/30)

	N	%	Scale Score Mean	Scale Score SD	Percentage Distributions for Level Scores					
					Below 3	3	4	5	6	7
Full Group	571,094		78.8	4.5	5	18	15	33	20	8
Gender										
Female	257,154	45.0%	78.9	4.3	4	18	16	35	21	7
Male	298,999	52.3%	78.8	4.7	6	19	14	31	20	9
Missing	15,541	2.7%	78.2	4.3	6	20	17	34	18	5
Ethnicity										
White	298,394	52.2%	80.0	4.4	3	12	12	34	27	11
African American	134,085	23.5%	76.7	3.8	9	29	20	31	9	1
Hispanic/Latino	53,397	9.3%	78.0	4.3	7	22	17	34	16	4
Asian	11,597	2.0%	80.1	5.0	5	14	11	29	25	16
American Indian/ Alaska Native	7,267	1.3%	77.4	4.2	8	24	18	33	13	3
Native Hawaiian/ Other Pacific Islander	1,745	0.3%	77.0	4.3	9	30	18	28	13	3
Two or more races	21,653	3.8%	79.2	4.3	4	16	15	36	22	8
Missing	42,956	7.5%	78.1	4.8	9	22	15	30	17	7

Note. Based on test records with valid scale scores.
Missing groups include the response category of 'prefer not to respond' for gender and ethnic variables.
Percentages of CBT and paper test administrations are 55% and 45%, respectively.

13.3 Summary Statistics of Four Operational Forms

This section presents summary statistics for four operational forms that were selected from the large sample described in Section 13.2. As presented in Table 13.2, the results include sample sizes, gender/ethnic group distributions, test completion rates, and scale score means and standard deviations. There are three CBT and one paper forms, denoted as "CBT #1", "CBT #2", "CBT #3", and "Paper". The three CBT forms were administered from July to November 2018, and the paper form was administered from May to November 2018.

Examinees taking the four forms have comparable characteristics to the total sample with the exception of a higher percentage of male examinees (68.4%) taking the paper form. The CBT forms have larger sample sizes than the paper form (over 26,000 to nearly 9,000). The percentages of the three largest ethnic groups are similar to those reported in Table 13.1.

Table 13.2: Summary Statistics for Four Forms

|	N	Female	Male	White	African American	Hispanic	Test Completion	Scale Score Mean	Scale Score SD
CBT #1	26,222	46.5	50.5	49.4	31.7	8.3	94.8	78.65	4.09
CBT #2	26,257	46.9	50.3	49.0	32.0	8.5	93.1	78.91	4.33
CBT #3	26,498	47.0	50.1	48.9	32.2	8.3	95.7	78.58	4.16
Paper	8,880	22.9	68.4	43.6	25.6	6.9	90.3	78.43	4.60

Note. The percentage values are reported for gender/ethnic groups and test completion rates.

Test completion rates are over 90% for the four forms. The average scale scores range from 78.43 to 78.91, which are slightly higher than the target; specifically, the target scale score mean and Standard Error of Measurement (SEM) were specified to be 77.9 and 1.7 (See Section 8.4 Procedures for Establishing the Score Scale). The SEM for the four forms is consistent with the target (See Table 13.4 below for the SEM for the four forms). Figure 14.1 presents the level score distributions for the four forms. The percentages for Below Level-3 and Level-7 groups are below 10% for all the four forms.

The Test Characteristic Curves (TCCs) and Test Information Functions (TIFs) for the four forms are presented in Figure 13.2 and Figure 13.3. For comparison, the scaling form is included as the base form (identical to those in Figure 8.3). (Note that these forms were built to meet the same assessment blueprint as presented in Section 3.3). The TCCs are placed tightly across the forms as shown in the figure.
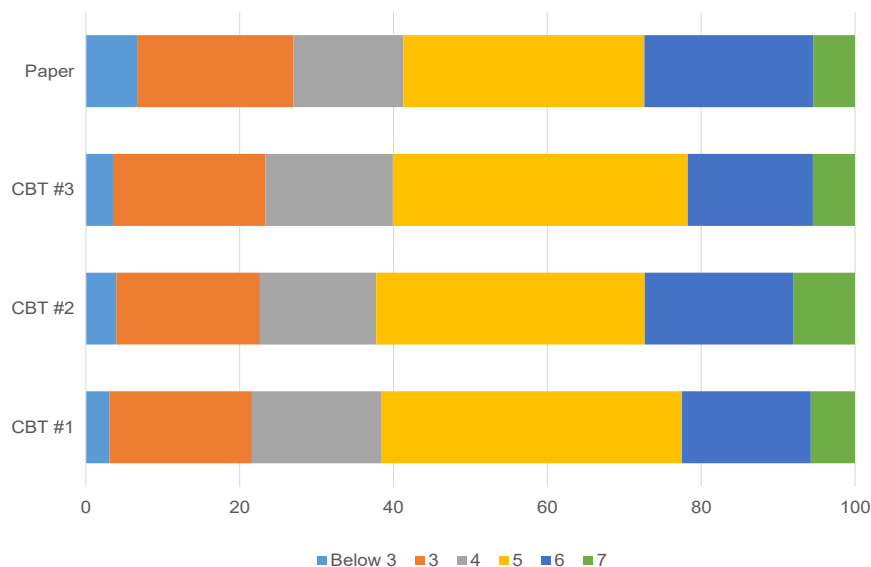
Figure 13.1. Level Score Distributions for Form Administrations
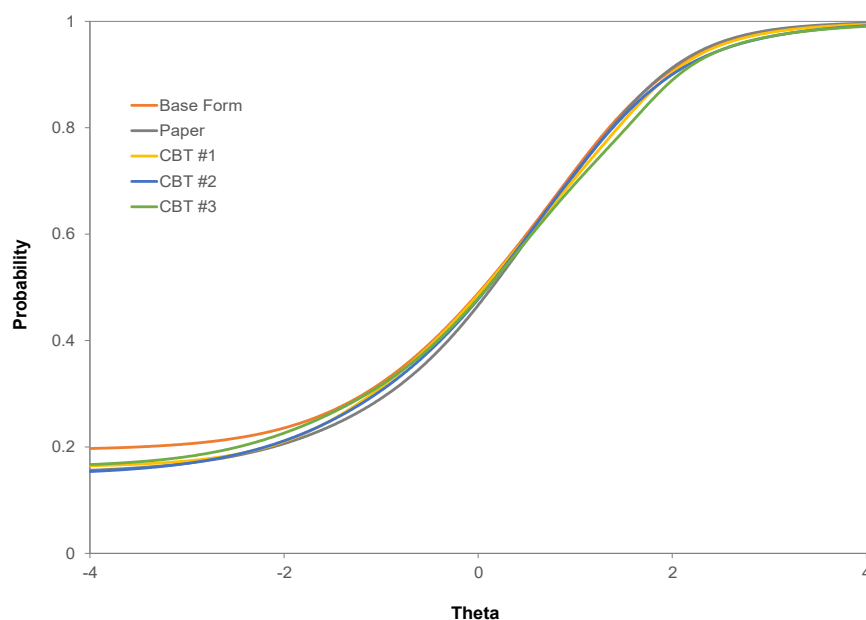
Figure 13.2. Test Characteristic Curves for Base Form and Four Operational Forms

ACT Research also continued to monitor Differential Item Functioning (DIF) for both pretest and operational items, using the same method to the analysis presented in Section 12.4.1. For the four forms, no operational items were flagged as Group C DIF based on the conditions between women and men, African Americans and whites, and Hispanics/Latinos and whites.
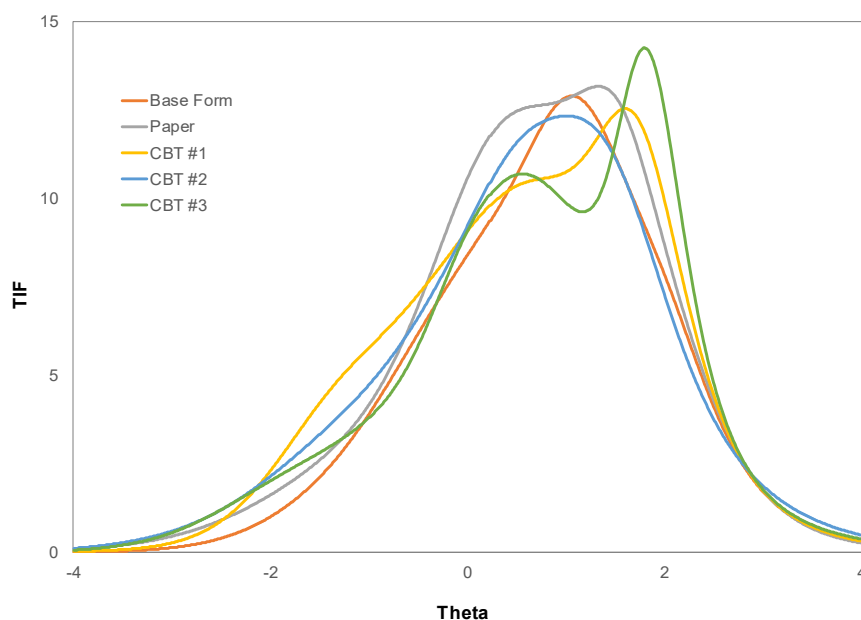
Figure 13.3. Test Information Function Curves for Base Form and Four Operational Forms

13.4 Reliability Analyses

The reliability analyses were divided into two parts. The first part is based on familiar estimates of score reliability, including Cronbach Alphas, scale score reliability, and Standard Error of Measurement (SEM) for scale scores from the four forms. Cronbach Alpha estimates range from 0.82 to 0.87 (see table 13.3), which are consistent with the scaling form (0.85, reported in Section 10.2). For the four forms, the reliability indices for scale scores are 0.85 or 0.86, and the scale score SEMs range from 1.67 to 1.73, which are slightly higher than that for the scaling form (1.61). The reliability and SEM for scale score are based on averaging CSEMs at theta values using form-specific conversion table and pool scale.

The second part analyzed the classification consistency of score levels for the forms. Based on the item parameter estimates used in preequating, the classification consistency analysis described in Section 10.4 was conducted. The classification consistency results are quite stable comparing Table 13.4 to Table 10.3.

Table 13.3: Reliability and SEM Results for the four forms

Form	Cronbach Alpha	Scale Score Reliability	Scale Score SEM
CBT #1	0.83	0.85	1.71
CBT #2	0.85	0.86	1.67
CBT #3	0.82	0.85	1.73
Paper	0.87	0.86	1.67

Table 13.4: Estimated Classification Consistency Indices for Level Scores for Three CBT Forms

Level	CBT #1		CBT #2		CBT #3		Paper	
	p	*κ*	*p*	*κ*	*p*	*κ*	*p*	*κ*
Exact	52%	39%	53%	40%	52%	39%	54%	41%
3	94%	42%	93%	46%	94%	40%	93%	41%
4	86%	64%	86%	67%	85%	64%	86%	66%
5	84%	68%	85%	70%	84%	68%	85%	70%
6	88%	64%	88%	66%	88%	66%	88%	68%
7	95%	55%	95%	56%	95%	54%	96%	55%

13.5 Dimensionality Evaluation

This section provides evidence that the test is unidimensional based on the same method used in Section 11.4.6, that is, eigenvalue comparisons of the first three factors from the exploratory factor analysis (EFA). Table 13.5 presents the EFA results for CBT #1 form. Similarly the Factor Difference Ratio Index (FDRI) value is significantly greater than 3, and the first factor explains 16% of total variance for the full set of operational items. These findings consistently indicate an underlying single factor structure on the Graphic Literacy assessment.

Table 13.5: Eigenvalues and Factor Difference Ratio Index (FDRI) - CBT #1 Form

Factor	Eigenvalue		Difference	FDRI
1	5.19	(16.21%)		
2	1.56	(4.88%)	3.63	
3	1.13	(3.53%)	0.43	8.44

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

In summary, Chapter 13 presents additional psychometric findings based on operational assessment data. The results of operational assessment data consistently support the findings from the field studies and provide strong evidence of the psychometric quality of the Graphic Literacy assessment forms. As additional WorkKeys Graphic Literacy forms are developed based on the assessment blueprint, ACT researchers will continue to implement similar analyses to review and monitor test form and item quality.

Chapter 14

# Defining Readiness for Work and Careers

There are many dimensions along which an individual needs to develop to be prepared for success throughout a lifetime. The path to success becomes more complex as individuals leave formal education systems and enter the workforce, where they must apply their knowledge and skills to demonstrate performance. College readiness, which is defined as having the skills and achievement levels needed to succeed in first-year, credit-bearing courses without remediation, is necessary for college success. On the other hand, core academic skills are necessary but not sufficient for college, career, and workplace success (Mattern, Burrus, Camara, O'Conner, Hanson, Grambrell, Casillas, & Bobek, 2014). A more holistic approach is needed to assess readiness across various transition points along the education and career continuum.

Readiness is applicable along a continuum, starting with a general or global standard for the typical level of skills needed for most jobs in the economy, to skill levels needed to be successful in a career pathway or for specific occupations. Career readiness is defined as having the Knowledge, Skills, Abilities, and Other characteristics (KSAOs) needed and the levels of those KSAOs needed to be successful in a typical job in a typical organization. Within the context of career readiness, foundational skills are the fundamental, portable skills that are critical to training and workplace success (Symonds, 2011). These skills are fundamental in that they serve as a basis—the foundation—for supporting more advanced skill development. And they are portable because, rather than being job specific, they can be applied at some level across a wide variety of occupations or within a career pathway. Readiness for a career pathway requires individuals to have the KSAOs and levels of KSAOs to be successful in a typical job within a career pathway.

In contrast to career readiness, a "work ready" individual possesses the KSAOs needed to be minimally qualified for a specific occupation as determined through a job analysis or occupational profile (ACT, 2013a). The skills needed for work readiness (a) are both foundational and occupation specific, (b) vary in both importance and level for different occupations, and (c) depend on the critical tasks identified via a job analysis or an occupational profile. Work readiness skills include foundational cognitive skills such as reading required for the workplace, applied mathematics, graphic literacy, problem solving, and critical thinking.

14.1 Work and Career Readiness Standards and Benchmarks

ACT® Work Readiness Standards and Benchmarks are precise descriptions of the knowledge and combination of skills that individuals need to be minimally qualified for a target occupation. These standards and benchmarks are determined by the level of skills profiled for a national representative sample of jobs in a given occupation (ACT, 2013a). While work readiness standards establish the mix of skills and range of levels reported by employers (i.e., minimum and maximum) for specific occupations, work readiness benchmarks are considered to be a target skill level (i.e., median) that an individual should aim for in order to be considered work ready for that occupation. The standards and benchmarks ensure that current and prospective employees' skills are aligned with employer skill requirements. They also ensure that individuals develop the foundational and job-specific skills necessary to be successful throughout a lifetime. ACT Career Readiness Standards and Benchmarks apply a similar methodology used to determine work readiness by providing individuals with a snapshot of skill requirements for different career pathways (LeFebvre, 2015). Figure 1 provides a summary of the work and career readiness definitions and corresponding examples of use cases.
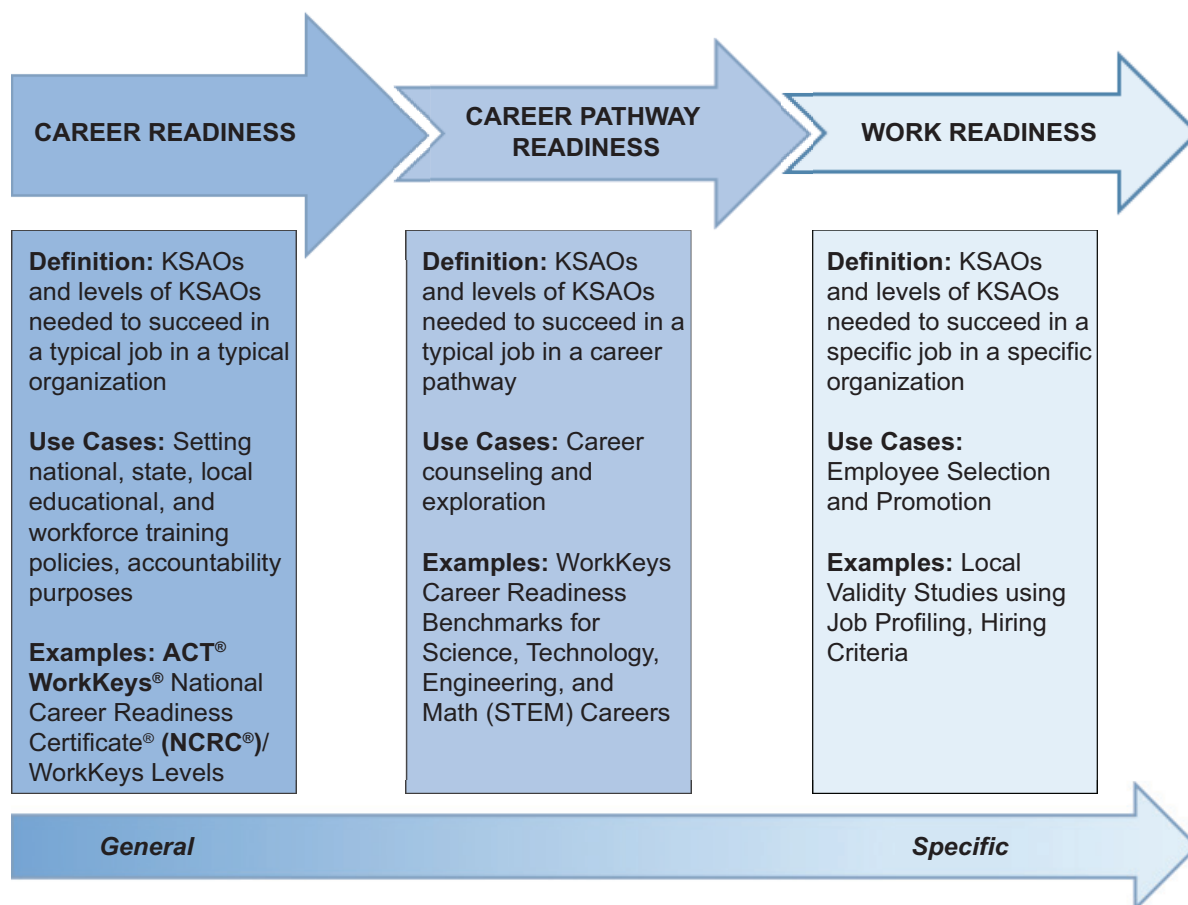
Figure 14.1: Summary of Work and Career Readiness

Hierarchical Education and Workforce Readiness Framework. From *The development of an empirical framework linking college readiness and career readiness* by M. LeFebvre and K. Mattern, in press, Iowa City, IA: ACT, Inc.

14.2 Using WorkKeys Assessments for Career and Work Readiness

The ACT® WorkKeys® Assessments can be used with ACT® WorkKeys® Job Profiling and the WorkKeys NCRC as a comprehensive system to support skill training and development, personnel selection, career planning, workforce and economic development, and accountability. While career and work readiness are closely related, the type of use determines whether specific WorkKeys Assessment scores or the WorkKeys NCRC is an appropriate measure for readiness. The following section provides a summary of the different uses of the WorkKeys Assessments and the WorkKeys NCRC.

14.2.1 Personnel Selection and Development

WorkKeys Assessments can be used for (a) pre-employment screening to identify individuals who have achieved levels of proficiency needed for a target job, (b) pre-employment screening to identify less desirable candidates based on behaviors associated with job performance, (c) employee development, and (d) developing the appropriate level of fit with occupations in terms of interests (LeFebvre, 2016).

When WorkKeys Assessments are used for pre-employment screening or other high-stakes employment decisions, employers should demonstrate that the knowledge and skills in the pre-employment measure are linked to work behaviors and job tasks either through job profiling or through research that links the assessment to job performance. When WorkKeys Assessments are used for employee development or the assessment of readiness for individuals or groups, criteria other than job performance may be more relevant (e.g., individual earnings, employment, or training completion). The WorkKeys Assessments should be used in combination with additional measures (e.g., tests, interviews, or other selection procedures) that the employer deems appropriate and relevant for pre-employment selection and other employment decisions.

14.2.2 Workforce and Economic Development

The WorkKeys Assessments and the WorkKeys NCRC are widely used in workforce and economic development programs. For example, the WorkKeys Assessments and the WorkKeys NCRC can be used by (a) an employer who uses the WorkKeys Assessments or the WorkKeys NCRC and other criteria to identify a qualified pool of applicants and requires a specific level of WorkKeys NCRC or WorkKeys scores, (b) an employer who uses the WorkKeys NCRC to make employment decisions and does not require a specific level, (c) states, communities, and schools that use the WorkKeys NCRC to document an individual's level of essential work readiness skills, and (d) states, communities, or schools that use the WorkKeys NCRC to document the aggregate career readiness of a community, region, or state.

ACT® Work Ready Communities (WRC) are an approach for workforce and economic developers to certify that their community has a qualified workforce to support industry demand. This approach uses WorkKeys Assessments and the WorkKeys NCRC to measure foundational workplace skills with goals established for the current, emerging, and transitioning workforce. In order to be certified as a Work

Ready Community, states and their counties establish goals based on the Work Ready Communities common criteria. The criteria are evaluated using the WorkKeys NCRC levels obtained across various subpopulations of the workforce (ACT, 2015). Skill gaps across various sectors of the workforce can be identified and addressed by state or local community policies and programs.

14.2.3 Accountability

State accountability systems, such as Career and Technical Education programs, have incorporated WorkKeys Assessments and the WorkKeys NCRC as a measure of employability skills or career readiness (Center on Education Policy, 2013). The WorkKeys NCRC is typically used in conjunction with other technical skills assessments such as industry-based certificates or licensing exams as part of a stackable credentialing system (ACT, 2013b). Some states also report using WorkKeys Assessment results as a requirement for graduation, for receipt of a career/technical diploma, endorsement on a standard diploma, or for scholarship eligibility.

References

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

ACT. (2011). *A better measure of skills gaps: Utilizing ACT skill profile and assessment data for strategic skill research.* Retrieved from https://www.act.org/content/dam/act/unsecured/documents/abettermeasure.pdf

ACT. (2013a). *Work Readiness Standards and Benchmarks: The key to differentiating America's workforce.* Iowa City, IA: Author.

ACT. (2013b). *Skills credentials aid displaced manufacturing workers in Ohio.* ACT Case Study. Iowa City, IA: Author.

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

ACT. (2015). *ACT Work Ready Communities: Common criteria*. Iowa City, IA: Author.

ACT. (2016). *ACT national curriculum survey 2016.* Retrieved from http://www.act.org/content/act/en/research/national-curriculum-survey.html

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

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

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 Technical Report.* Washington, DC: National Center for Education Statistics.

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

Association for Career and Technical Education (ACTE). (2010). *What is "career ready"?* Retrieved from https://www.acteonline.org/WorkArea/DownloadAsset.aspx?id=2114

Autor, D. H. (2015). Skills, education, and the rise of earnings inequality among the "other 99 Percent." *Science, 344,* 843–851.

Autor, D. H., Levy, F., & Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics, 118*(4), 1279–1333.

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.

Bartels, M., & Marshall, S. P. (2012). Measuring cognitive workload across different eye tracking hardware platforms. In *Proceedings of the 2012 symposium on eye tracking research and applications* (pp. 161–164). Santa Barbara, CA: ACM.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91,* 276–292.

BeGaze Manual. (2011). Retrieved from http://twiki.cis.rit.edu/twiki/pub/MVRL/SmiTracker/begaze2.pdf

Bessen, J. (2014). Employers are not just whining— the "skills gap" is real. *Harvard Business Review*, August 25, 2014. Retrieved from https://hbr.org/2014/08/employers-arent-just-whining-the-skills-gap-is-real/

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). New York, NY: Springer.

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

Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19–47.

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

Cappelli, P. (2012). *Why good people can't get jobs: The skills gap and what companies can do about it*. Philadelphia, PA: Wharton Digital Press.

Carnevale, A. P., & Desrochers, D. M. (2003). *Connecting education standards and employment: Course-taking patterns of young workers.* Washington, DC: The American Diploma Project.

Cascio, W. F. (1982). *Applied psychology in personnel management* (2nd ed.). Reston, VA: Reston Publishing.

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

Center on Education Policy (2013, October). *Career Readiness across states.* Washington, DC: Georgetown University.

Chinn, D. (2017). *The definition of workplace skills.* Retrieved from http://www.ehow.com/info_7786830_definition-workplace-skills.html

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.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3–17). Hillsdale, NJ: Lawrence Erlbaum.

Cronbach, L. J., Gleser, G. C., Nanda, H. I., & Rajaratnam, N. (1972). *The dependability of behavioral measurement: Theory of generalizability of scores and profiles*. New York, NY: Wiley.

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

Djamasbi, S. (2014). Eye tracking and web experience. *AIS transactions on human-computer interaction, 6*, 16–31.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Erlbaum.

Dunnette, M. D., & Hough, L. M. (Eds.). (1990). *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1). Palo Alto, CA: Consulting Psychologists Press.

Equal Employment Opportunity Commission (EEOC), Civil Service Commission, Department of Labor, & Department of Justice. (2000, revised). *Uniform Guidelines on Employee Selection Procedures (1978)*. Federal Register 43, 38290-38315 (August 25, 1978). Codified in 29 CFR 1607.

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.). Burlingame, CA: Analytics Press.

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*, *32,* 123–158.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests.* New York, NY: Wiley.

Goldin, C., & Katz, L. F. (2008). *The race between education and technology.* Cambridge, MA: The Belknap Press of Harvard University.

Greene, B. B. (2008). *Perceptions of the effects of the WorkKeys system in North Carolina.* (Doctoral dissertation, Western Carolina University). Retrieved from ProQuest.

Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 1–15). New York, NY: Springer.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and applications.* Boston, MA: Kluwever-Nijhoff.

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.

Hendrick, R. Z., & Raspiller, E. E. (2011). Predicting employee retention through pre-employment assessment. *Community College Journal of Research and Practice, 35*, 895–908.

Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (p. 19). New York, NY: Springer.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.

Hunter, J. E. (1980). *Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB).* Washington, DC: U.S. Department of Labor.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*, 28–42.

Infosys (2016). *Amplifying human potential. Education and skills for the Fourth Industrial Revolution.* Retrieved from http://www.experienceinfosys.com/humanpotential

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

Jacob, R. J. K, & Karn, K. S. (2003). Commentary on Section 4. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In R. Hyona, J. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–606). Oxford, England: Elsevier Science.

Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, *20*(2), 146–154.

Johannesson, P., & Perjons, E. (2014). *An introduction to design science*. Cham, Switzerland: Springer International Publishing.

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

Kane, M. T. (2006). Validation. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.

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

Kirsch, I., Braun, H. Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future.* Educational Testing Service Research Paper. Princeton, NJ: Educational Testing Service.

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. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York, NY: 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.

Krugman, P. (2014). Jobs and skills and zombies. *New York Times,* March 30, 2014. Retrieved from http://www.nytimes.com/2014/03/31/opinion/krugman-jobs-and-skills-and-zombies.html?_r=2

Langenfeld, T. E. (2014). *ACT WorkKeys: Awarding college credit through the ACT National Career Readiness Certificate.* Iowa City, IA: ACT, Inc.

Langenfeld, T. E., & Thomas, J. (2016, August). *WorkKeys Graphic Literacy: Applying Design Science to assessment development.* Paper presented at ACT conference on the applications of Design Science, Iowa City, IA.

Langenfeld, T., Thomas, J., Zhu, R., & Morris, C.A. (2019). Integrating multiple sources of validity evidence for an assessment-based cognitive model. *Journal of Educational Measurement*.

LeFebvre, M. (2015). *Career Readiness in the United States 2015.* Iowa City, IA: ACT, Inc.

LeFebvre, M. (2016). *A summary of ACT WorkKeys validation research*. Iowa City, IA: ACT, Inc.

LeFebvre, M., & Mattern, K. (in press). *The development of an empirical framework linking college readiness and career readiness.* Iowa City, IA: ACT, Inc.

Lewis, D. M., Mitzel, H. C., & Green, D. R. (1996, June). Standard setting: A Bookmark approach. In D. R. Green (Chair), *IRT-based standard setting procedures utilizing behavioral anchoring.* Symposium conducted at the Council of Chief State School Officers National Conference on Large-scale Assessment, Phoenix, AZ.

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, NY: Routledge.

Linn, R. L. (1993). Linking results in distinct assessments. *Applied Measurement in Education, 6*, 83–102.

Liu, J., & Dorans, N. J. (2016). Fairness in score interpretation. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 77–96). New York, NY: Routledge.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement*, *8*, 452–461.

ManpowerGroup® (2015). *Talent shortage survey 2015.* Retrieved from www.manpowergroup.com/talent-shortage-2015

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002, proceedings of the 2002 IEEE 7th conference* (pp. 75–79) Scottsdale, AZ: IEEE.

Marshall, S.P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine, 78* (5), B165-B175.

Masters, G. N., Adams, R., & Lokan, J. (1994). Mapping student achievement. *International Journal of Educational Research, 21*, 595–609.

Matos, R. (2010). *Designing eye tracking experiments to measure human behavior.* Tobil Technology. Retrieved from www.measuringbehavior.org/files/tutorials/Tobil_MB2010_tutorial_handouts.pdf

Mattern, K., Burrus, J., Camara, W., O'Conner, R., Hanson, M., Gambrell, J., Casillas, A., & Bobek, B. (2014). *Broadening the definition of college and career readiness: A holistic approach.* Iowa City, IA: ACT, Inc.

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

Mayo, M. J. (2012). *Evaluation metrics, New Options New Mexico, 2011–2012.* Unpublished report, Albuquerque, NM.

Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive processing*, *13*, 261–265.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: American Council on Education and Macmillan.

Mislevy, R. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects.* Princeton, NJ: Educational Testing Service.

Mislevy, R. (2006). Cognitive psychology and educational assessment. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 257–305). Westport, CT: American Council on Education and Praeger.

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 https://www.yumpu.com/en/document /view/40872213/a-brief-introduction-to-evidence-centered-design-cse-report-632-

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.

Mroch, A. A., Li, D., & Thompson, T. D. (2015). *A framework for evaluating score comparability.* Presented at the annual meeting of National Council on Measurement in Education, Chicago, IL.

National Institute of Standards and Technology. (2017). *Computer security division computer security resource center.* Retrieved from http://csrc.nist.gov/publications/PubsSPs.html

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/

Obersteiner, A., Moll, G., Beitlich, J. T., Cui, C., Schmidt, M., Khmelivska, T., & Reiss, K. (2014). Expert mathematicians' strategies for comparing the numerical values of fractions—evidence from eye movements. In S. Oesterle, C. Nicol, P. Liljedahl, & D. Allan (Eds.), *Proceedings of the joint meeting of PME 38 and PME-NA36* (Vol. 4, pp. 338–345). Vancouver, Canada: PME.

Organization of Economic Cooperation and Development (OECD). (2016). The survey of adult skills: Reader's companion (2nd ed.). Retrieved from http://www.oecd.org/publications/the-survey-of-adult -skills-9789264258075-en.htm

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology, 60*, 211–229.

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). Menlo Park, CA: AAAI Press.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment. *American Psychologist, 63*, 215–227.

Schmidt, F. L, & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schmidt, F. L., Hunter, J. E., Pearlman, K., & Shane, G. S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization procedure. *Personnel Psychology, 32*, 257–281.

Schmidt, F. L., Oh, I. S, & Shaffer, J. A. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings* (working paper)*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2853669

Schmidt, F. L., & Sharf, J. C. (2010). *Review of ACT's WorkKeys program relative to the Uniform Guidelines and more current professional standards.* Unpublished report, Iowa City, IA.

Schultz, D., & Stern, S. (2015). *How do WorkKeys assessment affect college and career readiness perceptions of Alaska high school students?* Retrieved from http://www.iser.uaa.alaska.edu/CAEPR /home/docs/2015_02-WorkKeysPolicyBrief.pdf

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1997). A study of modified-Guttman and scale scores using IRT. *Journal of Educational Measurement*, *33*, 129–140.

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1999). A rationale for defining achievement levels using IRT-estimated domain scores. *Applied Psychological Measurement*, *23*, 347–362.

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.

Schulz, E. M., & Mitzel, H. C. (2005). *The mapmark standard setting method.* Portions of this research paper have been presented to the National Assessment Governing Board for the National Assessment of Educational Progress. ERIC Number: ED490643.

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

Shepard, L. A., Camilli, G., & Williams, D. M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement, 22*, 77–105.

Society for Human Resource Management (SHRM) (2010). *Critical employee skills for the changing workforce.* Retrieved from https://www.shrm.org/search/pages/default.aspx?k=Critical%20employee%20skills%20for%20the%20changing%20workforce&filters=site:www.shrm.org/hr-today/public-policy

Society for Human Relations Management (SHRM). (2015). *Avoiding adverse impact in employment practices.* Retrieved from https://www.shrm.org/resourcesandtools/tools-and-samples/toolkits/pages/avoidingadverseimpact.aspx

Society for Industrial and Organizational Psychology, Inc. (SIOP). (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measure*, *7*, 201–210.

Stone, E. A., & Cook, L. L. (2016). Testing individuals in special populations. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 135–156). New York, NY: Routledge.

Subkoviak, M. J. (1984). Estimating the reliability of mastery-nonmastery classifications. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 267–290). Baltimore, MD: The Johns Hopkins University Press.

Symonds, W. (2011). *Pathways to prosperity.* Cambridge, MA: Harvard Graduate School of Education.

Thomas, J. & Langenfeld, T. E. (2017, April). *Analyzing think-aloud and eye-tracking data to support score interpretations.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Antonio, TX.

U.S. Bureau of Labor Statistics. (2013), *Occupational employment projections to 2022.* Retrieved from https://www.bls.gov/opub/mlr/2013/article/overview-of-projections-to-2022.htm

Van Aken, J. E. & Romme, G. L. (2012). A design science approach to evidence-based management. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 140–184). London, England: Oxford University Press.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think-aloud method: A practical guide to modelling cognitive processes*. San Diego, CA: Academic Press.

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

Wollack, J. A., & Case, S. A. (2016). Maintaining fairness through test administration. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 33–53). New York, NY: Routledge.Society for Human Resource Management (SHRM) (2010). *Critical employee skills for the changing workforce.* Retrieved from https://www.shrm.org/search/pages/default.aspx?k=Critical%20employee%20skills%20for%20the%20changing%20workforce&filters=site:www.shrm.org/hr-today/public-policy

Society for Human Relations Management (SHRM). (2015). *Avoiding adverse impact in employment practices.* Retrieved from https://www.shrm.org/resourcesandtools/tools-and-samples/toolkits/pages/avoidingadverseimpact.aspx

Society for Industrial and Organizational Psychology, Inc. (SIOP). (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measure*, *7*, 201–210.

Stone, E. A., & Cook, L. L. (2016). Testing individuals in special populations. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 135–156). New York, NY: Routledge.

Subkoviak, M. J. (1984). Estimating the reliability of mastery-nonmastery classifications. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 267–290). Baltimore, MD: The Johns Hopkins University Press.

Symonds, W. (2011). *Pathways to prosperity.* Cambridge, MA: Harvard Graduate School of Education.

Thomas, J. & Langenfeld, T. E. (2017, April). *Analyzing think-aloud and eye-tracking data to support score interpretations.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Antonio, TX.

U.S. Bureau of Labor Statistics. (2013), *Occupational employment projections to 2022.* Retrieved from https://www.bls.gov/opub/mlr/2013/article/overview-of-projections-to-2022.htm

Van Aken, J. E. & Romme, G. L. (2012). A design science approach to evidence-based management. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 140–184). London, England: Oxford University Press.

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think-aloud method: A practical guide to modelling cognitive processes*. San Diego, CA: Academic Press.

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

Wollack, J. A., & Case, S. A. (2016). Maintaining fairness through test administration. In N. J. Dorans & L. L. Cook (Eds.), *Fairness in educational assessment and measurement* (pp. 33–53). New York, NY: Routledge.

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.

Appendix

Professional Literature Review

As part of the effort to update the Locating Information assessment, the Design Team conducted a thorough review of the professional literature on graphics, including their development and use in education and work. The Graphic Literacy construct was defined, in large part, through the integration of the ideas and findings of numerous researchers. The following section lists the literature the Design Team reviewed in developing the construct. These references can also serve as a resource for educators and workforce development officials to better understand graphics, and the ways in which they facilitate communication and learning.

Graphic Literacy Resources

Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching, 43*(1), 43–62. Retrieved January 25, 2016 from http://doi.org/10.1002/tea.20087

Abilock, D. (2008). Visual information literacy: Reading a documentary photograph. *Knowledge QUEST*, *36*(3), 7–13.

ACRL Visual Literacy Competency Standards for Higher Education Association of College & Research Libraries (ACRL). (n.d.). Retrieved from http://www.ala.org/acrl/standards/visualliteracy

American College and Research Libraries. (2000). *Information literacy competency standards for higher education*. American Library Association. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl /files/content/standards/standards.pdf

Anderson, D. (2008). The low bridge to high benefits: Entry-level multimedia, literacies, and motivation. *Computers and Composition*, *25*(1), 40–60. Retrieved from http://doi.org/10.1016/j .compcom.2007.09.006

Anirban Mukherjee, U., & Biswas, A. (2014). Experimenting with automatic text-to-diagram conversion: A novel teaching aid for the blind people. Retrieved from http://www.ifets.info/others/download_pdf .php?j_id=64&a_id=1489

Arslan, R., & Nalinci, G. Z. (2014). Development of visual literacy levels scale in higher education. *Turkish Online Journal of Educational Technology-TOJET*, *13*(2), 61–70.

Attali, Y., & Goldschmidt, C. (1996). The effects of component variables on performance in graph comprehension tests. *Journal of Educational Measurement, 33*, 93–105.

Avgerinou, M. D. (2007). Towards a visual literacy index. *Journal of Visual Literacy*, *27*(1). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl =1051144X&AN=37702159&h=sjGHt4dTRcjib3KhBHzzjeYlHohNUDpOuTk8Ef%2Fr66PPUzc BZLrCnX1Pv9%2BnAh1gSMm4QU1k7l5EF9XSMaLC1g%3D%3D&crl=c

Avgerinou, M. D. (2009). Re-viewing visual literacy in the "bain d'images" era. *TechTrends*, *53*(2), 28–34.

Avgerinou, M. D., & Pettersson, R. (2011). Toward a cohesive theory of visual literacy. *Journal of Visual Literacy*, *30*(2), 1–19.

Bamford, A. (2003). *The visual literacy white paper* (White Paper commissioned by Adobe Systems). Uxbridge, UK. Retrieved from https://www.aperture.org/wp-content/uploads/2013/05/visual-literacy -wp.pdf

Bell, A., & Janvier, C. (1981). The interpretation of graphs representing situations. *For the Learning of Mathematics*, *2*(1), 34–42.

Bell, J. C. (2014). Visual literacy skills of students in college-level biology: Learning outcomes following digital or hand-drawing activities. *Canadian Journal for the Scholarship of Teaching and Learning*, *5*(1), 1–16. Retrieved from http://doi.org/10.5206/cjsotl-rcacea.2014.1.6

Blajenkova, O., Kozhevnikov, M., & Motes, M. A. (2006). Object-spatial imagery: a new self-report imagery questionnaire. *Applied Cognitive Psychology*, *20*(2), 239–263. Retrieved from http://doi.org/10.1002/acp.1182

Bowen, G. M., & Roth, W. M. (2002). Why students may not learn to interpret scientific inscriptions. *Research in Science Education*, *32*(3), 303–327.

Brumberger, E. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19–47.

Bryant, P. E., & Somerville, S. (1986). The spatial demands of graphs. *British Journal of Psychology*, *77*, 187–197.

Callow, J. (2005). Literacy and the visual: Broadening our vision. *English Teaching: Practice and Critique*, *4*(1), 6–19.

Callow, J. (2006). Images, politics and multiliteracies: Using a visual metalanguage. *Australian Journal of Language & Literacy*, *29*(1), 7–23.

Callow, J. (2008). Show me: Principles for assessing students' visual literacy. *The Reading Teacher*, *61*(8), 616–626. Retrieved from http://doi.org/10.1598/RT.61.8.3

Callow, J. (2011). When image and text meet. Retrieved from https://www.researchgate.net/publication /281069602_When_image_and_text_meet_-_teaching_with_visual_and_multimodal_texts

Capraro, M., Kulm, G., & Capraro, R. (2005). Middle grades: Misconceptions in statistical thinking. *School Science and Mathematics*, *105*(4), 165–174.

Castera, J., Sarapuu, T., & Piksoot, J. (2012). Developing students' visual literacy in a virtual modeling environment. In *Conference proceedings of "eLearning and Software for Education" (eLSE)* (pp. 532–537). Retrieved from http://www.ceeol.com/aspx/getdocument .aspx?logid=5&id=74156709cd964561aaa509657399b675

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

Danos, X., & Norman, E. W. L. (2009). The development of a new taxonomy for graphicacy. Retrieved from https://dspace.lboro.ac.uk/dspace/handle/2134/5100.

Danos, X., & Norman, E. W. L. (2011). Continuity and progression in graphicacy. In *IDATER online conference: graphicacy and modeling 2010* (pp. 103–119). Loughborough Design Education Group. Retrieved from https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/9023

Debes, J. (1969). The loom of visual literacy. *Audiovisual Instruction*, *14*, 25–27.

Emanuel, R., & Challons-Lipton, S. (2013). Visual literacy and the digital native: Another look. *Journal of Visual Literacy*, *32*(1), 7–26.

English, L. D. (2012). Young children's metarepresentational competence in data modeling? (pp. 266–273). MERGA Inc. Retrieved from http://eprints.qut.edu.au/55617/

Ericson, M. D. (2009). *Bibliography/webliography for visual literacy: Visual imagery, databases, and a contextual approach to understanding images*. Retrieved from http://digitalcommons.colby.edu/cgi/viewcontent.cgi?article=1011&context=faculty_scholarship

Farrell, T. (2013). *Measuring visual literacy ability in graduate level pre-service teachers*. Retrieved from http://scholarworks.umt.edu/etd/4130/

Felten, P. (2008). Visual literacy. *Change: The Magazine of Higher Learning*, *40*(6), 60–64.

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

Friedman, H. (2003). Methodolatry and graphicacy. *American Psychologist*, *58*(10), 817–818.

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*, *32,* 124–158.

Garcia, G. G., & Cox, R. (2010). An interactive educational diagrammatic system for assessing and remediating the graph-as-picture misconception. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 224–226). New York, NY: Springer.

Harris, B. R. (2006). Visual information literacy via visual means: Three heuristics. *Reference Services Review, 34*(2), 213–221.

Hattwig, D., Bussert, K., Medaille, A., & Burgess, J. (2013). Visual literacy standards in higher education: New opportunities for libraries and student learning. *Portal: Libraries and the Academy*, *13*(1), 61–89.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, *91*(4), 684–689.

Holloway, S. M. (2012). Visual literacies: An ecology arts-based pedagogical model. *Language and Literacy*, *14*(3), 150–168.

Hug, T. (2011). Visual competence, media literacy and "New Literacies"–Conceptual considerations in a plural discursive landscape. *Media, Technology, and Life-long Learning, 7* (1), 1–17. Retrieved from http://www.seminar.net/index.php/volume-3-issue-3-2007-previousissuesmeny-121/172-visual -competence-media-literacy-and-qnew-literaciesq-conceptual-considerations-in-a-plural-discursive -landscape

Jacobs, G. E. (2013). Designing assessments: A multiliteracies approach. *Journal of Adolescent & Adult Literacy*, *56*(8), 623–626.

Jones-Kavaler, B. R., & Flannagan, S. L. (2006). Connecting the digital dots: Literacy of the 21st Century. *Educause Quarterly*, *29*(2), 8–10.

Kelly-Jackson, C., & Delacruz, S. (2014). Using visual literacy to teach science academic language: Experiences from three preservice teachers. *Action in Teacher Education*, *36*(3), 192–210. Retrieved from http://doi.org/10.1080/01626620.2014.917364

Kragten, M., Admiraal, W., & Rijlaarsdam, G. (2013). Diagrammatic literacy in secondary science education. *Research in Science Education*, *43*(5), 1785–1800. Retrieved from http://doi.org/10.1007 /s11165-012-9331-0

Kress, G., & van Leeuwen, T. (2006). *Reading images the grammar of visual design* (2nd ed.). London, England: Routledge. Retrieved from http://www.slideshare.net/praetece/reading-images-the -grammar-of-visual-design-van-leeuwen-and-kress

Krufka, S. E., & Barner, K. E. (2006). A user study on tactile graphic generation methods. *Behaviour & Information Technology*, *25*(4), 297–311. Retrieved from http://doi.org/10.1080/01449290600636694

Kuntze, M., Golos, D., & Enns, C. (2014). Rethinking literacy: Broadening opportunities for visual learners. *Sign Language Studies*, *14*(2), 203–224.

Lesy, M. (2007). Visual literacy. *The Journal of American History*, *94*(1), 143–153.

LeVine, P., Scollon, R. (Eds.) (2005). Discourse and technology: Multimodal discourse analysis. *Applied Linguistics, 26*(4). Washington, DC: Georgetown University Press.

Long, V. (2010). *An agent-based approach to table recognition and interpretation.* (Doctoral dissertation, Macquarie University, Sydney, Australia). Retrieved from http://web.science.mq.edu.au/~rdale/ students/VanessaLong/dissertation_revised_2010-05-23.pdf

Lowrie, T., Diezmann, C., & Logan, T. (2011). Understanding graphicacy: Students' making sense of graphics in mathematics assessment tasks. *International Journal for Mathematics Teaching and Learning*. Retrieved from http://eprints.qut.edu.au/48802/

Mackey, T. P., & Jacobson, T. E. (2010). Reframing information literacy as a metaliteracy. *College & Research Libraries*. Retrieved from http://crl.acrl.org/index.php/crl/article/view/16132/17578

Marshall, S.P. (2007). Identifying cognitive state from eye metrics. *Aviation, Space, and Environmental Medicine*, *78*(5), B165-B175

Matusitz, J. (2005). The current condition of visual communication in colleges and universities of the United States. *Journal of Visual Literacy*, *25*(1), 97–112. Retrieved from http://search.ebscohost.com /login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=1051144X&AN=17604479 &h=xhNjVj3w5QqTmVZfpshF3wf06EQJzoM5IczA1OK%2F%2Flej5gRpm%2FIefs%2FxmZBTUwt 5XNgQ%2BAM4yoTiLoIXq%2BvYbQ%3D%3D&crl=c.

McDonald, C., McDonald, S., & Maddox, A. (2011). *Strategies for developing 21st century skills through visual literacy.* Retrieved from

Merhar, V. K., Planinsic, G., & Cepic, M. (2009). Sketching graphs—an efficient way of probing students' conceptions. *European Journal of Physics*, *30*(1), 163–175. Retrieved from http://doi.org/10.1088/0143-0807/30/1/017

Milner, M. M., & Ying Hill, W. (2008). Support for graphicacy: A review of textbooks available to accounting students. *Accounting Education*, *17*(2), 173–185. Retrieved from http://doi.org/10.1080/09639280701612150

Moodley, V. (2013). In-service teacher education: Asking questions for higher order thinking in visual literacy. *South African Journal of Education*, *33*(2), 1–17.

Moore-Russo, D., & Shanahan, L. E. (2014). A broader vision of literacy: Including the visual with the linguistic. *Journal of Adolescent & Adult Literacy*, *57*(7), 527–532. Retrieved from http://doi.org/10.1002/jaal.282

National Council of Teachers of English (NCTE). (2008). *NCTE framework for 21st century curriculum and assessment*. Retrieved from http://www.ncte.org/positions/statements/21stcentframework

Nemirovsky, R. (1994). On ways of symbolizing: The case of Laura and the Velocity Sign. *Journal of Mathematical Behavior*, *13*(4), 389–422.

Newfield, D. (2011). From visual literacy to critical visual literacy: An analysis of educational materials. *English Teaching: Practice and Critique*, *10*(1), 81–94.

Palmer, M. S., & Matthews, T. (2014). Learning to see the infinite: Measuring visual literacy skills in a 1st-year seminar course. *Journal of the Scholarship of Teaching and Learning*, *15*(1), 1–9. Retrieved from http://doi.org/10.14434/josotl.v15i1.13089

Palmquist, N. N. (2008). Creating images to understand visual literacy. *Knowledge Quest*, *36*(3), 20–23.

Pauwels, L. (2008). Visual literacy and visual culture: Reflections on developing more varied and explicit visual competencies. *Open Communication Journal*, *2*, 79–85.

Pettersson, R. (1988). *Visuals for information*. Stockholm, Sweden: Esselte Forlag. Retrieved from http://files.eric.ed.gov/fulltext/ED304085.pdf

Piksoot, J., & Castera, J. (2012). Developing students' visual literacy in a virtual modeling environment. *eLearning & software for education*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=2066026X&AN=87676669&h=xtyKmMEFn9%2B5clVqLGoiHmmoNzhUhqzNayaalAVwfN7%2BUiy8crsniBE5SqFET6mgCFh%2BzFECbKubO1GSi%2FvV2w%3D%3D&crl=c.

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2008). Thinking graphically: Connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology Applied*, *14*(1), 36–49.

Rosier, J., & Dyer, P. (2010). Visual literacy: A necessary skill for planning graduates? *Social Alternatives*, *29*(4), 64–68.

Roth, W. M., & Bowen, G. M. (1994). Mathematization of experience in a grad 8 open-inquiry environment: An introduction to the representational practices of science. *Journal of Research in Science Teaching*, *31*(3), 293–318.

Rourke, A. J., & O'Connor, Z. (2009). *Recommendations for intervention in the visual domain: A study of undergraduate design students' visual literacy skills and predominant learning modalities*. Paper presented at the ACUADS 2009 Conference in Queensland, Australia.

Rowsell, J., McLean, C., & Hamilton, M. (2012). Visual literacy as a classroom approach. *Journal of Adolescent & Adult Literacy*, *55*(5), 444–447. http://doi.org/10.1002/JAAL.00053

Scheuermann, A., & van Garderen, D. (2008). Analyzing students' use of graphic representation: Determining misconceptions and error patterns for instruction. *Mathematics Teaching in the Middle School*, *13*(8), 471–477.

Serafini, F. (2011). Expanding perspectives for comprehending visual images in multimodal texts. *Journal of Adolescent & Adult Literacy*, *54*(5), 342–350.

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

Shah, P., & Hoeffner, J. (2002a). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47–69.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, *91*(4), 690–702.

Shoemaker, J., Schau, E., & Ayers, R. (2008). Making information visual. *Knowledge Quest*, *36*(3), 28–32.

Sibanda, J., & Sibanda, L. (2013). Visual literacy development through the mediation of Grade 4 English textbooks. *Journal of Visual Literacy*, *32*(2), 39–66.

Smith, D. W., & Smothers, S. M. (2012). The role and characteristics of tactile graphics in secondary mathematics and science textbooks in braille. *Journal of Visual Impairment & Blindness*, *106*(9), 543–554.

Snavely, L. (2005). Visual images and information literacy. *Reference & User Services Quarterly*, 27–32.

Spalter, A. M., & van Dam, A. (2008). Digital visual literacy. *Theory into Practice*, *47*(2), 93–101. Retrieved from http://doi.org/10.1080/00405840801992256

Stein, M. K., Baxter, J. A., & Leinhardt, G. (1990). Subject-matter knowledge and elementary instruction: A Case from functions and graphing. *American Educational Research Journal*, *27*(4), 639–663. Retrieved from http://doi.org/10.2307/1163104

Stokes, S. (2002). Visual literacy in teaching and learning: A literature perspective. *Electronic Journal for the Integration of Technology in Education*, *1*(1), 10–19.

Takemura, R., Shimojima, A., & Katagiri, Y. (2014). Logical investigation of reasoning with tables. In T. Dwyer, H. Purchase, & A. Delaney (Eds.), *Diagrammatic Representation and Inference* (pp. 261–276). New York, NY: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-662-44043-8_27

Tang, G. M. (1997). From graphic literacy across languages to English and content teaching in vocational settings. *Hong Kong Journal of Applied Linguistics*, *2*(1), 97–114.

Taylor, H. A., Renshaw, C. E., & Choi, E. J. (2004). The effect of multiple formats on understanding complex visual displays. *Journal of Geoscience Education*, *52*(2), 115–121.

Tillmann, A. (2012). *What we see and why it matters: How competency in visual literacy can enhance student learning*. Retrieved from http://digitalcommons.iwu.edu/education_honproj/9/

van Leeuwen, T. (2004). Ten reasons why linguists should pay attention to visual communication. In P. Levine & R. Scollon (Eds.), *Discourse and technology: Multimodal discourse analysis* (pp. 7–19). Washington, DC: Georgetown University Press.

Vasquez, J. A., Comer, M., & Troutman, F. (2010). *Developing visual literacy in science, K–8*. Arlington, VA: National Science Teachers Association Press.

Vukicevic, M., & Fitzmaurice, K. (2009). Eccentric viewing training in the home environment: Can it improve the performance of activities of daily living. *Journal of Visual Impairment and Blindness*, *103*, 277–290.

Wainer, H. (1980). A test of graphicacy in children. *Applied Psychological Measurement*, *4*(3), 331–340.

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

Waller, R. (2012). Graphic literacies for a digital age: The survival of layout. *The Information Society*, *28*(4), 236–252. Retrieved from http://doi.org/10.1080/01972243.2012.689609

Werts, B. C. E. (2008). One specific method inquiry and education (visual and media literacy): Interview with Tamara Moats, speaker on visual thinking and use of art data. *Education Libraries*, *31*(2), 39–41.

Yeh, H., & Lohr, L. (2010). Towards evidence of visual literacy: Assessing pre-service teachers' perceptions of instructional visuals. *Journal of Visual Literacy*, *29*(2), 183–197.

Zambo, D. M. (2009). Using visual literacy to help adolescents understand how images influence their lives. *Teaching Exceptional Children*, *41*(6), 60–67.

Zebehazy, K. T., & Wilton, A. P. (2014). Quality, importance, and instruction: The perspectives of teachers of students with visual impairments on graphics use by students. *Journal of Visual Impairment & Blindness*, *108*(1), 5–16.