WorkKeys NCRC Assessments Technical Manual

March 2023

Commitment to Fair Testing

ACT endorses and is committed to complying with *The Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). ACT also endorses the Code of Fair Testing Practices in Education (Joint Committee on Testing Practices, 2004), which is a statement of the obligations to test takers of those who develop, administer, or use educational tests and test data in the following four areas: developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. ACT endorses and is committed to complying with the *Code of Professional Responsibilities in Educational Measurement* (NCME Ad Hoc Committee on the Development of a Code of Ethics, 1995), which is a statement of professional responsibilities for those involved with various aspects of assessments, including development, marketing, interpretation, and use.

Table of Contents

Commitment to Fair Testing . . . ii

List of Tables . . . v

List of Figures . . . ix

Chapter 1: Purpose and General Description of the WorkKeys Assessments . . . 1
1.1 WorkKeys NCRC Assessments . . . 1
1.2 Purpose . . . 2
1.3 Foundational Workplace Skills . . . 2
1.4 The Workforce Skills Gap and the WorkKeys Solution . . . 6
1.5 WorkKeys—Assessment Claims . . . 13
1.6 Test Users and Stakeholders . . . 16
1.7 Alignment to the ACT® Holistic Framework® . . . 17
1.8 Alignment of ACT WorkKeys Skills With National Adult Education Academic Standards . . . 19
1.9 Additional WorkKeys Assessments . . . 19

Chapter 2: Test Development . . . 20
2.1 Workplace Documents . . . 20
2.2 Applied Math . . . 31
2.3 Graphic Literacy . . . 44

Chapter 3: Test Specifications . . . 59
3.1 WorkKeys NCRC Specifications—Overview . . . 59
3.2 Content Relevance and Representativeness . . . 60
3.3 WorkKeys NCRC Assessments—Test Blueprints . . . 61
3.4 Graphic Literacy—Evidence Based on Response Processes . . . 70

Chapter 4: Test Administration . . . 76
4.1 Policies and Procedures . . . 76
4.2 Test Administration Personnel and Their Responsibilities . . . 78
4.3 Training Testing Staff . . . 81
4.4 Test Administration Room Requirements . . . 82

Chapter 5: Accessibility . . . 83
5.1 ACT WorkKeys NCRC Assessment Support System . . . 83
5.2 Test Administration and Accessibility Levels of Support . . . 84
5.3 Valid Test Scores and Equal Benefit for All Examinees . . . 87

Chapter 6: Test and Information Security . . . 88
6.1 Test Security . . . 88
6.2 Information Security . . . 89

Chapter 7: Reporting . . . 92
7.1 WorkKeys Reports . . . 92

Chapter 8: Scores and Score Scales . . . 95
8.1 Overview . . . 95
8.2 Selected-Response Item Scoring . . . 96
8.3 Scale Score and Level Score Differences and Rationale . . . 96
8.4 Procedures for Establishing the Score Scale . . . 97
8.5 Procedures for Establishing the Level Scores . . . 111

Chapter 9: Equating and Linking . . . 116
9.1 Equating Method and Procedures . . . 116
9.2 Field Test Studies . . . 117
9.3 Linking . . . 144

Chapter 10: Reliability and Measurement Error . . . 157
10.1 Overview . . . 157
10.2 Reliability Coefficients and Standard Error of Measurement (SEM) . . . 158
10.3 Generalizability Theory . . . 159
10.4 Classification Consistency of Level Scores . . . 162

Chapter 11: Validity . . . 166
11.1 Validation of Test Score Uses and Interpretations . . . 166
11.2 WorkKeys Assessments Claims . . . 167
11.3 Primary Claims and Relevant Findings . . . 169
11.4 Evaluation of Claims . . . 179
11.5 Ongoing Validation . . . 181
12.1 Test Fairness—Overview . . . 182
12.2 Fairness and Test Administration . . . 182
12.3 Fairness in Access to the Measured Construct . . . 183
12.4 Fairness as a Lack of Measurement Bias . . . 184
12.5 Fairness as the Validity of Individual Score Interpretations . . . 190

Chapter 13: Operational Validation . . . 196
13.1 Overview . . . 196
13.2 Workplace Documents . . . 196
13.3 Applied Math . . . 202
13.4 Graphic Literacy . . . 208

Chapter 14: Defining Readiness for Work and Careers . . . 213
14.1 Work and Career Readiness Standards and Benchmarks . . . 214

References . . . 217

List of Tables

Table 1.1. Classroom Reading Versus Workplace Reading 8
Table 2.1. Workplace Documents—Passage Level Complexity Descriptors 22
Table 2.2. Workplace Documents Skills and Subskills. 24
Table 2.3. Workplace Documents—External Subject Matter Experts 31
Table 2.4. Item Stem Characteristics by Level 34
Table 2.5. Skill 1.0—Basic Operations with Numbers Including Decimals 35
Table 2.6. Skill 2.0—Fractions 36
Table 2.7. Skill 3.0—Percentages/Ratios/Proportions 36
Table 2.8. Skill 4.0—Unit Conversions 37
Table 2.9. Skill 5.0—Geometric Measurement 37
Table 2.10. Skill 6.0—Applied Math Reasoning 38
Table 2.11. Applied Mathematics—External Subject Matter Experts 44
Table 2.12. Characteristics of Simple, Low Moderate, High Moderate, and Difficult Graphics . . 46
Table 2.13. Graphic Literacy Cognitive Skills and Subskills 48
Table 2.14. Graphic Literacy—Cognitive Process Model 50
Table 2.15. Graphic Literacy Score or Performance Levels 50
Table 2.16. Graphic Literacy—External Subject Matter Experts 58
Table 3.1. Skill Domain Item Distribution by Level 62
Table 3.2. Skill Subdomain Item Distribution by Level 62
Table 3.3. Number of Passages for Each Document Type 63
Table 3.4. Applied Math Skills Item Distribution by Level 64
Table 3.5. Basic Operations: Subskill Item Distribution 64
Table 3.6. Fractions: Subskill Item Distribution 65
Table 3.7. Percentages/Ratios/Proportions: Subskill Item Distribution 65
Table 3.8. Unit Conversions: Subskill Item Distribution 66
Table 3.9. Geometric Measurement: Subskill Item Distribution 66
Table 3.10. Applied Mathematical Reasoning: Subskill Item Distribution 67
Table 3.11. Number of Items per Level for Applied Math Applications 67
Table 3.12. Interaction of Graphic Complexity Level With Cognitive Skill Levels
With the Overall Graphic Level Definitions 69
Table 3.13. Number of Items by Graphic Complexity and Overall Graphic Literacy Level 69
Table 3.14. Graphic Literacy Skill Distribution by Level 69
Table 3.15. Cognitive Skill: Number of Items by Graphic Complexity 70
Table 4.1. Responsibilities of the Test Coordinator 78
Table 7.1. Various WorkKeys NCRC Reports and Their Functions 93
Table 8.1. Summary of Unrounded and Rounded Scale Scores—Workplace Documents . . . 100
Table 8.2. Summary of Unrounded and Rounded Scale Scores—Applied Math. 105
Table 8.3. Summary of Unrounded and Rounded Scale Scores—Graphic Literacy 110
Table 8.4. Median Cut Scores for the Workplace Documents Assessment 113
Table 8.5. Median Cut Scores for the Applied Math Assessment 114
Table 8.6. Median Cut Scores for the Graphic Literacy Assessment 115
Table 9.1. WorkKeys NCRC Test Population Compared to Field Test Samples of
Workplace Documents by Age, Gender, and Ethnicity 118
Table 9.2. WorkKeys NCRC Test Population Compared to Field Test Samples of
Applied Math by Age, Gender, and Ethnicity 118
Table 9.3. WorkKeys NCRC Test Population Compared to Field Test Samples of
Graphic Literacy by Age, Gender, and Ethnicity 119

Table 9.4. Sample Demographic Information for the Two Delivery Modes of Workplace Documents . . . 122
Table 9.5. Sample Demographic Information for the Two Delivery Modes of Applied Math . . . 123
Table 9.6. Sample Demographic Information for the Two Delivery Modes of Graphic Literacy . . . 124
Table 9.7. Test Summary Statistics for Workplace Documents . . . 125
Table 9.8. Test Summary Statistics for Applied Math . . . 128
Table 9.9. Test Summary Statistics for Graphic Literacy . . . 131
Table 9.10. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Workplace Documents . . . 134
Table 9.11. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Applied Math . . . 135
Table 9.12. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Graphic Literacy . . . 137
Table 9.13. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Workplace Documents . . . 140
Table 9.14. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Applied Math . . . 141
Table 9.15. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Graphic Literacy . . . 142
Table 9.16. Summary Statistics for Total Testing Times (in Minutes) for Reading for Information and Workplace Documents . . . 145
Table 9.17. Test Summary Statistics for Reading for Information and Workplace Documents . . . 146
Table 9.18. Scale Score Summary Statistics for Reading for Information and Workplace Documents . . . 146
Table 9.19. Summary Statistics for Scale Scores for Reading for Information and Workplace Documents . . . 147
Table 9.20. Summary Statistics for Level Scores of Workplace Documents . . . 148
Table 9.21. Summary Statistics for Total Testing Times (in Minutes) for Applied Mathematics and Applied Math . . . 150
Table 9.22. Test Summary Statistics for Applied Mathematics and Applied Math . . . 150
Table 9.23. Scale Score Summary Statistics for Applied Mathematics and Applied Math . . . 150
Table 9.24. Summary Statistics for Scale Scores for Applied Mathematics and Applied Math . . . 152
Table 9.25. Summary Statistics for Level Scores of Applied Mathematics and Applied Math . . . 152
Table 9.26. Summary Statistics for Total Testing Times (in Minutes) for Locating Information and Graphic Literacy . . . 154
Table 9.27. Test Summary Statistics for Locating Information and Graphic Literacy . . . 154
Table 9.28. Scale Score Summary Statistics for Locating Information and Graphic Literacy . . . 155
Table 9.29. Summary Statistics for Scale Scores for Locating Information and Graphic Literacy . . . 156
Table 9.30. Summary Statistics for Level Scores of Locating Information and Graphic Literacy . . . 156
Table 10.1. Reliability Estimates and Standard Error of Measurement for Form W2C_S1 . . . 158

Table 10.2. Reliability Estimates and Standard Error of Measurement for Applied Math Form M2C_S1 . . . 159
Table 10.3. Reliability Estimates and Standard Error of Measurement for Graphic Literacy Form G2C_S1 . . . 159
Table 10.4. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Workplace Documents Form W2C_S1 . . . 161
Table 10.5. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Applied Math Form M2C_S1 . . . 161
Table 10.6. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Graphic Literacy Form G2C_S1 . . . 161
Table 10.7. Estimated Classification Consistency Indices for Level Scores for Form W2C_S1 . . . 163
Table 10.8. Estimated Classification Consistency Indices for Level Scores for Form M2C_S1 . . . 164
Table 10.9. Estimated Classification Consistency Indices for Level Scores for Form G2C_S1 . . . 165
Table 11.1. Initial Assumptions and Chapter Numbers . . . 169
Table 11.2. WorkKeys NCRC Validation Studies . . . 173
Table 11.3. Correlations Between WorkKeys Assessments and GPA . . . 178
Table 12.1. WorkKeys DIF Evaluations—Group Comparisons . . . 185
Table 12.2. WorkKeys NCRC DIF Classification Rules . . . 185
Table 12.3. Workplace Documents—Number and Percent of Field Test Participants by Demographic Group . . . 187
Table 12.4. Identifications of C-Level DIF Items on the Two Workforce Documents Forms . . 187
Table 12.5. Applied Math—Number and Percent of Field Test Participants by Demographic Group . . . 188
Table 12.6. Identifications of C-Level DIF Items on the Two Applied Math Forms . . . 189
Table 12.7. Graphic Literacy—Number and Percent of Field Test Participants by Demographic Group . . . 189
Table 12.8. Identifications of C-Level DIF Items on the Two Graphic Literacy Forms . . . 190
Table 12.9. Comparing the Requirements of the *Uniform Guidelines* to the ACT WorkKeys Job-Profiling Procedure . . . 192
Table 12.10. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Applied Math . . . 194
Table 12.11. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Graphic Literacy . . . 194
Table 12.12. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Workplace Documents . . . 195
Table 13.1. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Workplace Documents (5/1/2021–4/30/2022) . . . 197
Table 13.2. Summary Statistics for Four Workplace Documents Forms . . . 198
Table 13.3. Summary of C-Level DIF Items on Four Workplace Documents Forms . . . 200
Table 13.4. Reliability and SEM Results for Four Workplace Documents Forms . . . 200
Table 13.5. Estimated Classification Consistency Indices for Level Scores for Four Workplace Documents Forms . . . 201
Table 13.6. Eigenvalues and Factor Difference Ratio Index (FDRI) for Workplace Documents—CBT 1 Form . . . 201

Table 13.7. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Applied Math Assessment (5/1/2021–4/30/2022) . . . 202
Table 13.8. Summary Statistics for Four Applied Math Forms . . . 203
Table 13.9. Summary of C-Level DIF Items on Four Applied Math Forms . . . 205
Table 13.10. Reliability and SEM Results for Four Applied Math Forms . . . 206
Table 13.11. Estimated Classification Consistency Indices for Level Scores for Four Applied Math Forms . . . 206
Table 13.12. Eigenvalues and Factor Difference Ratio Index (FDRI) for Applied Math—CBT 1 Form . . . 207
Table 13.13. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Graphic Literacy Assessment (5/1/2021–4/30/2022) . . . 208
Table 13.14. Summary Statistics for Four Graphic Literacy Forms . . . 209
Table 13.15. Reliability and SEM Results for Four Graphic Literacy Forms . . . 211
Table 13.16. Estimated Classification Consistency Indices for Level Scores for Four Graphic Literacy Forms . . . 212
Table 13.17. Eigenvalues and Factor Difference Ratio Index (FDRI) for Graphic Literacy—CBT 1 Form . . . 212

List of Figures

Figure 8.1. Raw Score Distribution for the WD Scaling Study Form W2C_S1 . . . 97
Figure 8.2. Item *p*-Values and *b*-Parameter Estimates by Item Levels for Form W2C_S1 . . . 98
Figure 8.3. Test Characteristic Curve (Left) and Test Information Function (Right) . . . 98
Figure 8.4. CSEM for Raw Scores (Left) and Scale Scores (Right) . . . 100
Figure 8.5. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right) . . . 101
Figure 8.6. Raw Score Distribution for the AM Scaling Study Form M2C_S1 . . . 102
Figure 8.7. Item *p*-Values and *b*-Parameter Estimates by Item Levels for Form M2C_S1 . . . 103
Figure 8.8. Test Characteristic Curve (Left) and Test Information Function (Right) . . . 103
Figure 8.9. CSEM for Raw Scores (Left) and Scale Scores (Right) . . . 105
Figure 8.10. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right) . . . 106
Figure 8.11. Raw Score Distribution for the GL Scaling Study Form G2C_S1 . . . 107
Figure 8.12. Item *p*-Values and *b*-Parameter Estimates by Item Levels for Form G2C_S1 . . . 108
Figure 8.13. Test Characteristic Curve (Left) and Test Information Function (Right) . . . 108
Figure 8.14. CSEM for Raw Scores (Left) and Scale Scores (Right) . . . 110
Figure 8.15. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right) . . . 111
Figure 9.1. Comparison of Item Omit Rates for the Two Delivery Modes of Workplace Documents . . . 123
Figure 9.2. Comparison of Item Omit Rates for the Two Delivery Modes of Applied Math . . . 124
Figure 9.3. Comparison of Item Omit Rates for the Two Delivery Modes of Graphic Literacy . . . 125
Figure 9.4. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Workplace Documents . . . 126
Figure 9.5. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Workplace Documents . . . 126
Figure 9.6. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Workplace Documents . . . 127
Figure 9.7. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Workplace Documents . . . 127
Figure 9.8. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Applied Math . . . 128
Figure 9.9. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Applied Math . . . 129
Figure 9.10. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Applied Math . . . 130
Figure 9.11. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Applied Math . . . 130
Figure 9.12. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Graphic Literacy . . . 131
Figure 9.13. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Graphic Literacy . . . 132
Figure 9.14. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Graphic Literacy . . . 133
Figure 9.15. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Graphic Literacy . . . 133

Figure 9.16. Comparison of Raw Score Distributions for the Two Delivery Modes of Workplace Documents . . . 134
Figure 9.17. Comparison of Scale Score Distributions for the Two Delivery Modes of Workplace Documents . . . 135
Figure 9.18. Comparison of Raw Score Distributions for the Two Delivery Modes of Applied Math . . . 136
Figure 9.19. Comparison of Scale Score Distributions for the Two Delivery Modes of Applied Math . . . 136
Figure 9.20. Comparison of Raw Score Distributions for the Two Delivery Modes of Graphic Literacy . . . 137
Figure 9.21. Comparison of Scale Score Distributions for the Two Delivery Modes of Graphic Literacy . . . 138
Figure 9.22. Workplace Documents Eigenvalue Scree Plot . . . 139
Figure 9.23. Applied Math Eigenvalue Scree Plot . . . 140
Figure 9.24. Graphic Literacy Eigenvalue Scree Plot . . . 141
Figure 9.25. Comparison of Item Omit Rates Between Reading for Information and Workplace Documents . . . 145
Figure 9.26. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Reading for Information and Workplace Documents . . . 146
Figure 9.27. Comparison of Item Omit Rates Between Applied Mathematics and Applied Math . . . 149
Figure 9.28. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Applied Mathematics and Applied Math . . . 151
Figure 9.29. Comparison of Item Omit Rates Between Locating Information and Graphic Literacy . . . 154
Figure 9.30. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Locating Information and Graphic Literacy . . . 155
Figure 11.1. Job Profile Process Designed to Align Job Tasks to Skill Levels . . . 171
Figure 11.2. Average GPA by NCRC Distinction . . . 179
Figure 13.1. Level Score Distributions for Workplace Documents . . . 198
Figure 13.2. Test Characteristic Curves (TCCs) for Workplace Documents—Base Form and Four Operational Forms . . . 199
Figure 13.3. Test Information Function (TIF) Curves for Workplace Documents—Base Form and Four Operational Forms . . . 199
Figure 13.4. Level Score Distributions for Applied Math . . . 204
Figure 13.5. Test Characteristic Curves (TCCs) for Applied Math—Base Form and Four Operational Forms . . . 204
Figure 13.6. Test Information Function (TIF) Curves for Applied Math—Base Form and Four Operational Forms . . . 205
Figure 13.7. Level Score Distributions for Graphic Literacy . . . 210
Figure 13.8. Test Characteristic Curves (TCCs) for Graphic Literacy—Base Form and Four Operational Forms . . . 210
Figure 13.9. Test Information Function (TIF) Curves for Graphic Literacy—Base Form and Four Operational Forms . . . 211
Figure 14.1. General Conceptualizations of College Readiness and Career Readiness . . . 213
Figure 14.2. Summary of Work and Career Readiness . . . 214

Chapter 1: Purpose and General Description of the WorkKeys Assessments

1.1 WorkKeys NCRC Assessments

The ACT® WorkKeys® suite of assessments and the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) provide a comprehensive workforce development solution that gives high school students and job-seeking adults scores that are valid indicators of career readiness. The ACT WorkKeys Assessments and the resulting NCRC are nationally recognized for the comprehensive and holistic evaluation of workforce-ready skills that help job seekers gain employment and help employers find the right candidate.

The WorkKeys cognitive assessments are criterion-referenced assessments. Unlike the more commonly used norm-referenced assessments, WorkKeys test scores are not determined by the relationship of an examinee's score to other examinees within a norm group. In WorkKeys, examinees are measured in terms of their ability to demonstrate competency in identified skill sets. As a result, an individual's scores indicate the skills a person can demonstrate in a given domain. The WorkKeys NCRC is based on the following three WorkKeys Assessments:

ACT® WorkKeys® Workplace Documents is a multiple-choice assessment designed to measure the skills people use when they read and use written documents to do a job. The documents—which include, but are not limited to, messages, emails, letters, directions, signs, notices, bulletins, policies, websites, contracts, and regulations—are based on materials that reflect actual reading demands of the workplace.

ACT® WorkKeys® Applied Math is a multiple-choice assessment designed to measure the extent to which individuals can use the mathematical skills needed in workplaces, where the skills to think problems through to find and evaluate solutions is important. The assessment measures skills that individuals use when they apply mathematical reasoning and problem-solving to work-related problems.

ACT® WorkKeys® Graphic Literacy is a multiple-choice assessment designed to measure the extent to which individuals can 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 assessment measures skills that individuals use to comprehend information presented in graphical format and then to take that information and solve some type of problem.

1.2 Purpose

ACT WorkKeys provides information to examinees, employers, workforce development officials, and educators:

- For examinees, WorkKeys Assessments provide insights about their foundational skills and their career readiness. In some cases, scores on the assessments may assist examinees in finding employment.
- For employers, the assessments provide information that may be used, along with other information, for employment decisions.
- For workforce development officials, the assessments provide information regarding the work-ready status of individuals requesting services and information to guide individuals toward jobs.
- For secondary educators, the assessments provide information related to foundational skills and career readiness that may be used as an accountability measure.
- For postsecondary educators, the assessments provide information related to program readiness or program evaluation.

An additional purpose of the WorkKeys Assessments relates to the issuance of the WorkKeys NCRC. The WorkKeys NCRC is an evidence-based, portable career readiness credential and the foundation of the ACT® Work Ready Communities framework. ACT uses its assessment and certification expertise to help community leaders develop a sustainable mechanism to close skills gaps and collect workforce skills data. The goal is to certify a county, state, or region as "work ready" when specific criteria for building a skilled workforce are met. The collective outcome of ACT Work Ready Communities, powered by ACT® Workforce Solutions, is a collaborative workforce development initiative that gives individuals the verifiable proof they need to show employers that they are ready to succeed. The result is a robust talent pipeline that benefits individuals and employers.

1.3 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 have to be skilled 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 that are taught through formal schooling—such as the academic subjects of reading, writing, mathematics, and science—but they may be learned from other sources. These skills enable individuals to acquire job-specific skills, communicate information with fellow workers, and engage in lifelong learning.

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 et al., 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; National Network of Business and Industry Associations [NNBIA], 2014; Organization of Economic Cooperation and Development [OECD], 2016; Society for Human Resource Management [SHRM], 2012). As the economy has become more technical and global, the pace of change has increased greatly. Successful workers need to have a flexible mindset 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).

In recent years, several business and industry associations have built 21st-century workplace competency models that provide support for the inclusion of reading workplace documents, applied mathematics, and graphic literacy as foundational workplace skills (Infosys, 2016; Association for Career and Technical Education [ACTE], 2018; NNBIA, 2014).

ACT identified reading workplace or work-related documents, applied mathematics, and graphic literacy as three facets of foundational workplace skills. ACT based its assumption that these skills are foundational on three sources of evidence: (a) job analysis data has consistently indicated that the skills of reading workplace documents, applied mathematics, and graphic literacy are needed to achieve job success; (b) professional literature and job competency models identify these skills as critical 21st-century skills; and (c) the Programme for the International Assessment of Adult Competencies (PIAAC) assessments describes the skills to understand and interpret information presented in written text as the main component of adult literacy and understanding and solving mathematical problems as a critical element of adult numeracy.

1.3.1 Profiling

The *Principles for the Validation and Use of Personnel Selection Procedures* (2018) 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 such as an assessment. This process is commonly referred to as a job analysis. ACT refers to this job analysis process as ACT® WorkKeys® Job Profiling. WorkKeys Job Profiling focuses on identifying the skills and behaviors present across the WorkKeys Assessments that are required to perform a job. Creating a job profile 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.

Since initiating its WorkKeys Job Profiling services in 1993, ACT has conducted over 22,000 job profiles representing a wide cross-section of U.S. jobs. Job profiles have been conducted on jobs in manufacturing, healthcare, construction, financial services, public administration, leisure and hospitality, agriculture, and other sectors. Analysis of the ACT® WorkKeys® JobPro® database indicates that the skills associated with reading workplace documents were included in 19,482 profiles or 88% of all ACT profiles. When ACT assigned each completed profile to an O*NET job code, reading work-related documents appeared as a required skill for 706 distinct O*NET job codes or 69% of all O*NET job codes. While solving applied mathematics problems is not universally required across all jobs, analysis of the WorkKeys JobPro database indicates that the skills associated with applied math were included in 18,443 profiles or 83% of all ACT profiles. ACT has found that solving applied mathematics problems is used in 665 distinct O*NET job codes or approximately 65% of all O*NET job codes. While graphic literacy, like math and other skills, is not universally required across all jobs, analysis of the WorkKeys JobPro database indicates that the skills associated with graphic literacy were included in 20,336 profiles or 92% of all ACT profiles. ACT has found that graphic literacy skills are used in 706 distinct O*NET job codes or approximately 70% of all O*NET job codes.

When WorkKeys Assessments are 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 WorkKeys Job Profiling service as a valid method for gathering the required evidence to demonstrate both assessment relevancy and score level requirements.

1.3.2 National Network of Business and Industry Associations Competency Model

The competency model developed by Business Roundtable (NNBIA, 2014) defines common employability skills and classifies them into four categories: personal skills, people skills, applied knowledge, and workplace skills. The first skill identified under applied knowledge is reading. Business Roundtable maintains that employees need to be proficient in the following reading and math skills:

- Read and comprehend work-related instructions and policies, memos, bulletins, notices, letters, policy manuals, and government regulations
- Read and comprehend documents ranging from simple and straightforward to more complex and detailed
- Attain meaning and comprehend core ideas from written materials
- Integrate what is learned from written materials with prior knowledge
- Apply what is learned from written material to work situations
- Add, subtract, multiply, and divide whole numbers, fractions, decimals, and percentages
- Convert decimals to fractions; convert fractions to decimals

- Calculate averages, ratios, proportions, and rates
- Take measurement units of time, temperature, distance, length, width, height, and weight; convert one measure to another
- Translate practical problems into useful mathematical expressions

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, Business Roundtable defines specific skills that involve graphic literacy, which they state is required to interpret and understand data to make decisions.

1.3.3 Association for Career and Technical Education

The Association for Career and Technical Education (ACTE) argues that students must be able to apply academic knowledge to real-world situations that they might encounter in their careers. ACTE asserts that students need strong foundational academic knowledge and skills in English language arts and math. ACTE maintains that, because most students will be engaged in more than one career over their working lifetime, core academic skills are critical in helping them to develop new skills and to adjust to new situations. Too often, employers identify deficiencies in employees' abilities to read and communicate effectively as problematic. They find that "most of the written material students will encounter in their careers is informational in nature, such as technical manuals and research articles, and they (students) need to be equipped academically to analyze and use these materials" (p. 1). ACTE also asserts that students will receive workplace communications in varied formats designed to provide information needed to successfully perform tasks. ACTE 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.

1.3.4 Programme for the International Assessment of Adult Competencies

Labeling reading, applied mathematics, and graphic literacy as foundational workplace skills 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). PIAAC defines literacy in a manner that closely aligns to ACT's definition of Workplace Documents. PIAAC emphasizes the ability to read written text to gain information as an important skill to successfully complete a task. In the Numeracy assessment, PIAAC presents examinees with items that require the examinee to first understand the problem, then organize the problem, and then solve the problem. PIAAC expects their examinees to be able to apply mathematical skills to solve problems containing quantitative data. In the Literacy assessment, PIAAC presents reading passages to adult examinees 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. Although the PIAAC adult competencies are defined in terms of the skills required for being a successful adult, and the WorkKeys skills are defined in terms of the skills required for successful job performance, these two skill definitions are closely aligned.

Based on the understandings gained from studying WorkKeys Job Profiling data, the workforce competency models, and the construct definitions developed for the PIAAC assessments, reading workplace documents, applied mathematics, and graphic literacy are necessary foundational workplace skills that contribute to employee success and lifelong learning.

1.4 The Workforce Skills Gap and the WorkKeys Solution

The WorkKeys Assessment program was conceived to mitigate the "skills gap" problem. The term *skills gap* is used to describe the challenges that employers and hiring managers face when, although many well-paying jobs exist, they are unable to find workers to fill these jobs due to the shortage of qualified workers. ManpowerGroup® (2015) surveyed more than 40,000 global employers and found that 75% of companies have experienced problems finding qualified workers—a 16-year high. Seventy-four percent of United States employers reported experiencing problems finding qualified workers.

In response to the skills gap, the willingness of employers to prioritize skills over degrees has increased. In the past, a lack of a degree has shut workers out of many professions. According to the Hechinger Report, 62% of Americans over 25 have no bachelor's degree, and that number rises to 72% for Black adults and 79% for Latinx adults. The report predicts that any shift in the workforce to the advantage of workers without degrees carries obvious implications for economic mobility and equity. However, workers without degrees still need to provide evidence of their skills.

In line with ACT's core mission of seeking to help all people find success, regardless of who they are, where they come from, or where they're going on their education and career journey, ACT created the WorkKeys Assessments to address the discrepancy between foundational skill levels and job requirements (ACT, 2011). Because of the discrepancy, the WorkKeys Assessments provide a solution that is beneficial to both employers and workers. WorkKeys Assessments provide both employers and examinees with 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® WorkKeys® Curriculum program provides workers with the opportunity to improve their skills and achieve the required levels to qualify for jobs. The WorkKeys NCRC assessments provide opportunities for employers to hire the right person for the job, and they provide workers with the opportunity to qualify and demonstrate that they possess the foundational skills required for success.

1.4.1 Reading in the Classroom and the Workplace

To help delineate the construct of Workplace Documents, ACT reviewed the relevant literature on reading skills. In general, it was noted that reading instruction in the classroom does not always align with workplace needs. As highlighted below, reading and workplace research indicates that successful application of reading skills is situation-specific, with reading behaviors dictated by the reader's purpose and circumstances.

1.4.1.1 Workplace Reading

While electronic recordings can sometimes be substituted for live speech or demonstrations, the written word is still the most consistently available communication medium in the workplace. Employees who need to learn or review a procedure, verify previously encountered information, or find answers to job-related questions frequently do so by reading. Whether it is gathering ideas for a presentation, safely using a power tool, or mixing a solution in a lab, good reading skills can be the difference between success and failure. Similarly, good communication skills are frequently cited in surveys of employers as one of the top requirements of today's jobs (NNBIA, 2014). The ability to comprehend and interpret workplace documents is a critical component of workplace communication.

In contrast to classroom reading selections, workplace reading materials are usually written by individuals more qualified by their content knowledge than their writing skills. While these materials may be intended to convey precise meaning, they are not always easy to understand. Such materials may be used to train employees on safety and work procedures or to provide information on employee benefits such as insurance policies and retirement plans. Employees read many of these materials in order to make decisions about some immediate course of action. Other materials describe behaviors or circumstances that may be relevant to their jobs in a more general sense. In both cases, the employees' comprehension of the text and their compliance with what it dictates may be taken for granted.

According to Human Resources and Skills Development Canada (2004):

> A great deal of workplace reading is 'reading to do,' with the reader taking various actions and assuming risks associated with error. The fact that the reader takes various actions as a result of reading materials changes the dynamics of reading considerably. That is why the person with hands-on experience to support the knowledge gained through reading is often the best equipped to carry out the work.

Thus, one important difference between workplace and school reading is the degree to which individuals must directly apply information gathered from texts—often with serious consequences for themselves and their teams.

On the other hand, a primary function of reading in a school environment is to teach widely applicable literary skills. Not surprisingly, there are several foundational reading skills developed in primary and secondary schooling that transfer over into workplace reading situations. Table 1.1 summarizes the essential differences and points of overlap between classroom reading and workplace reading. The differences are likely to be in the purposes for reading, the type of materials that are read, and the amount of help that readers can expect when they approach reading tasks.

Table 1.1. Classroom Reading Versus Workplace Reading

Texts	Reading in School	Reading in the Workplace	Points of Overlap
Typical Text Types	• Literature (fiction and nonfiction) • Informational textbooks on different subjects • Assignments and worksheets • Informational websites	• A range of procedural and informational documents: Instructions, notices, bulletins, policies, and regulations • Email messages, memos, and other communications • Informational websites	• Informational texts and websites • Instructions and procedures
Authors	• Literary authors • Multiple or unspecified authors who contribute to a textbook	• Technical writers, content experts, specialists (e.g., lawyers), coworkers, and customers • Multiple or unspecified authors who contribute to a document	• Multiple or unspecified authors
Text Complexity Features and Readability Levels	• Texts selected and adjusted for grade level • Theoretical, academic language with emphasis on concepts and symbolic meaning • Largely prose, organizational features, and formatting (topic-focused paragraphs, sections, and chapters)	• A wide range of levels related to specific features of a task • Technical, job-specific language with emphasis on concrete tasks • A wide range of organizational features and formatting suited to specific task and purpose (e.g., mixtures of paragraphs, bullets/numbered lists, and other formatting elements)	• Precise terminology • Texts organized into paragraphs, sections, and chapters • Interactive online texts • Online texts with hyperlinks and various navigation features

Applying the above-discussed requirements of the workplace, ACT designed Workplace Documents to assess a wide range of skills related to reading and understanding workplace information, instructions, procedures, and policies. The action-oriented texts found in many workplaces differ from the explanatory and narrative texts on which most academic reading programs are based. In addition, unlike academic texts, which are usually organized to ease understanding and facilitate learning, workplace communication is not necessarily well-written or written with ease of reading as a primary consideration. The reading selections in Workplace Documents are based on actual workplace materials representing a variety of occupations and workplace situations. These selections and their associated test items are designed to the Workplace Documents construct defined in Chapter 2.

1.4.2 Mathematics in the Classroom and the Workplace

To help delineate the construct of Applied Math, ACT reviewed relevant literature on numeracy skills and their application to the workplace. Although classroom instruction in mathematics overlaps in important areas with workplace mathematical applications, it does not account for many workplace uses. A growing body of research has documented the differences between mathematical reasoning as it is taught in the classroom and how it is applied in the workplace. As emphasized below, these studies and evaluations indicate that the successful application of mathematics in the workplace is situational, incorporates problem-solving, and integrates various mathematical and quantitative reasoning skills (Australian Association of Mathematics Teachers Inc., 2014; Smith, 1999).

1.4.2.1 Workplace Math

Changes in workplace technology over the last half century, both large and unpredictable, have been rapidly absorbed and adopted. In the 1970s, the first desktop calculators cost hundreds of dollars and typically performed only the four basic arithmetic operations. Today, handheld graphing calculators selling for under $100 have more capabilities than early mainframe computers. Such developments imply necessary changes in the mathematics skills needed on the job. Where employees used to perform calculations by hand and check the results for accuracy and reasonableness, they now use calculators or spreadsheets from the outset. To be successful on the job, employees need

- problem-solving strategies to set up and run the calculations best suited to answer their needs, and
- sufficient estimation skills to be able to recognize when results are highly unlikely or to determine that incorrect data may have been entered.

Unlike mathematical problems presented in classrooms, workplace problems are seldom clearly defined. In the classroom, mathematical problems are often structured by a textbook and are taught somewhat in isolation. Although classroom mathematical skills tend to progress and build on one another, the student is normally solving mathematical problems as defined by the specific unit.

In applying mathematical skills to workplace problems, employees must utilize their understanding of mathematics and quantitative reasoning to derive the process or procedure for solving the problem. An employee may have a boss or coworker who will help him or her set up and solve the problem, but in many circumstances, the employee will be expected to set up and solve the problem without assistance. In other cases, besides setting up and solving the problem, the employee will need to determine what data is relevant and pertinent to solving the problem. Although the mathematical skills observed in the workplace may appear to be fundamental, it is the application of the skills to the workplace problem that is not straightforward.

To be successful with applying math in the workplace, workers need to be able to blend the following:

- Apply and integrate mathematical concepts, procedures, and skills
- Understand the types of practical tasks that require mathematical solutions
- Identify the strategic mathematical process required to solve the specified problems
- Identify pertinent or relevant information or data for use in solving problems

Each step in solving a workplace Applied Math problem—from defining the problem through evaluating the results—requires a comprehensive understanding of mathematics.

Another critical difference between the classroom and the workplace is the motivation or purpose for using mathematics. In the classroom, the purpose is often to solve an isolated problem or set of problems. In the workplace, context provides the purpose for doing the work and a practical need to know the result exists. Finding the best solution in the workplace can be the difference between an effective and efficient operation or one filled with problems, mistakes, and lost opportunities. Mathematical problem-solving is often intertwined with other issues, where the mathematical result is linked to business success.

Though people may believe they do not use math often, if at all, in their jobs, mathematics is often hidden in tasks as basic as recording hours on a timesheet, compiling an expense report, counting out change to a customer, or taking a patient’s pulse. Mathematics skills and concepts typically used at work include basic arithmetic operations, spatial reasoning, and converting between units of measurement (Nicol, 2002). In some cases, all that is needed is the ability to total a column of numbers, but, in other cases, the ability to analyze data, to move beyond computation to recursive thinking, multiplicative thinking, abstraction, and spatial visualization is essential (Nicol, 2002).

The modern office worker must use technology to solve problems. In this context, mathematics is both more concrete and more intuitive. The need for mathematical literacy and quantitative reasoning skills requires workers to be able to work through multiple-step problems and solve three-dimensional problems using two-dimensional data and elementary data analysis (Lapan, 1999).

Frank Levy and Richard J. Murnane (2004), in their book *The New Division of Labor: How Computers are Creating the Next Job Market*, believe that the increasing use of computers

> has made people into consumers of mathematics. A clothing manager uses a quantitative model to forecast dress demand. A truck dispatcher uses a mathematical algorithm to design delivery routes. A bakery worker monitors production using digital readouts rather than the smell of bread. Employees of all kinds are expected to use web-based tools to help manage their retirement plans. Each of these tasks involves some aspect of mathematical literacy. In most cases, a computerized tool does the actual calculation, but using the model without understanding the math leaves one vulnerable to potential serious misjudgments. (p. 104)

While classroom mathematics may isolate skills and focus on one type of problem at a time, workplace problems may require the application of several different skills to develop a solution. For example, individuals in the workplace may need to know how to select relevant data from a large amount of available information or to recognize that the data are presented in a different metric than the solution requires.

In the ACT® National Curriculum Survey® (NCS), the skills identified as important to postsecondary mathematics teachers are those stressed by the National Council of Teachers of Mathematics. These same skills are also valued in the workplace. However, while students may learn what to do in school, they will need to be able to transfer that knowledge to workplace contexts. The educational efforts of the Council and others are striving to close gaps between the skills learned in school and the skills used at work. The WorkKeys Applied Math assessment provides a standardized method for measuring a person's ability to apply the skills they acquired in the classroom to workplace situations.

1.4.3 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 WorkKeys Job Profiling[1] (ACT, 2023), and its importance has been confirmed through interviews conducted by ACT with outside workplace development professionals (ACT, 2016a). 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 a 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 WorkKeys 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 (ACT, 2016a). 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 examinees' 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 & Pettersson, 2015; Few, 2012; Koomey, 2017). For example, according to Koomey (2017),

> 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. (p. 161)

In evaluating the current Locating Information assessment, the ACT design team concluded that the assessment's title was limiting. At the lower levels, the assessment was designed to assess an examinee'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 & Pettersson, 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, 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 multimedia literacy. Mayer (2009) defined multimedia learning as the presentation of both words and pictures to better facilitate learning and retention. Mayer's cognitive theory of multimedia 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.

1.5 WorkKeys—Assessment Claims

The WorkKeys NCRC assessment claims address workforce development issues including improving worker access to better jobs, improving worker productivity, reducing employee turnover rates, and improving regional business productivity. Each WorkKeys NCRC assessment was designed to measure specific skills and are one part of a suite of assessments designed to measure (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, and (c) work readiness at the individual and community level.

Drawing on its understanding of the skills gap and skills-based hiring practices, ACT has defined the following three claims regarding WorkKeys NCRC 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 WorkKeys NCRC assessments are more likely to successfully perform in more and higher levels of U.S. jobs (in the ACT job taxonomy classified in the WorkKeys JobPro database) than examinees whose scores do not reach that level.

Claim #1 Assumptions:

1. Each of the three WorkKeys NCRC assessments is a component of foundational workplace skills; these skills are required for success in a large number of jobs (based on ACT's WorkKeys job profile database).

2. ACT has developed a professionally valid and appropriate definition of the WorkKeys NCRC assessments construct.

3. The WorkKeys NCRC assessments elicit observable evidence of the construct and provide reliable and interpretable scores that reflect the construct.

4. ACT has defined workplace-appropriate performance level descriptors (PLDs), and ACT has established standards (e.g., cut points (cut scores)) aligned to the PLDs.

5. Cut points (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 WorkKeys NCRC assessments are positively related to measures of employee performance, including productivity and turnover rates.

8. Examinees who score well on the WorkKeys NCRC assessments 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 WorkKeys NCRC assessments are more likely to improve productivity (for example, measured as increased output per day) than if the company had hired examinees whose scores do not reach that level.

Claim #2 Assumptions:

1. These include the first seven Claim 1 assumptions.

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 more effective and productive business.

Claim #3: U.S. companies who hire U.S. examinees of high school or workforce age who demonstrate WorkKeys NCRC assessment 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. These include the first seven Claim 1 assumptions.

2. Employees with higher foundational skill levels are less likely to be terminated in the first six months of employment than employees with lower foundational skill levels.

3. Employees with higher foundational skill levels are less likely to quit in the first six months of employment than employees with lower foundational skill levels.

4. Businesses that utilize scores from the WorkKeys NCRC assessments as part of their hiring process will tend to experience less turnover than businesses who do not use the WorkKeys NCRC assessment as part of their hiring process.

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

1. The skills required in reading workplace documents, in applied mathematics, and in graphic literacy are foundational workplace skills and are required for success in a large number of jobs;

2. ACT has developed a valid and appropriate construct definition of reading workplace documents, applied mathematics, and graphic literacy;

3. ACT's WorkKeys NCRC assessments provide reliable and interpretable scores measuring the construct;

4. ACT has defined appropriate WorkKeys NCRC assessment PLDs, and ACT has established standards aligned to the PLDs; and

5. The 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 chapters present data and analysis related to the five assumptions.

1.6 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.

1.6.1 Examinees

Individuals who take the WorkKeys NCRC assessments are students and workers interested in demonstrating their foundational workplace 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 and/or as a part of state accountability programs. Community college students take the assessment to demonstrate 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 demonstrate that they have the foundational workplace reading 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.

1.6.2 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 WorkKeys NCRC 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 with the opportunity to earn a NCRC. State departments of education use WorkKeys NCRC 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 NCRC assessments provide high schools and school districts with student data regarding the extent to which students can apply foundational reading, applied math, and graphic literacy skills to actual workplace situations. The application of these skills to workplace scenarios differentiates the NCRC assessments from other standardized assessments. The assessments provide students the opportunity to demonstrate their mastery of workplace reading along with the application of their reading, math, and graphic literacy skills to real-world problems.

Community colleges use WorkKeys NCRC assessments in a variety of ways. Many community colleges use the WorkKeys NCRC as part of the process for determining acceptance into career and technical education programs. Other community colleges use the assessments for program evaluation or as a means of assisting their graduates in obtaining employment.

WorkKeys NCRC assessments have the flexibility to assist community colleges in improving their programs in different ways. They can assist a program in identifying students who have the foundational skills required to complete a specific program of study. In this way, the WorkKeys NCRC 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 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 NCRC assessments and then use their test scores as an additional piece of information to determine which candidates to interview.

1.7 Alignment to the ACT® Holistic Framework®

Building on research conducted over the last 50 years, ACT has developed its Holistic Framework (Camara et al., 2015), which provides a more complete description of education and work 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 skills measured by the Workplace Documents assessment fall into three major categories: identify main ideas and details, apply instructions or information, and identify meanings and definitions of words or phrases. These skills align primarily with the first broad domain of the ACT Holistic Framework, which includes domain-specific knowledge and skills necessary for performing essential tasks.

The Workplace Documents assessment uses authentic workplace documents and scenarios in order to determine an examinee's level of proficiency in reading workplace documents and applying the information within these documents to the types of tasks an employee would be expected to perform. The ability to use and interpret entire texts or parts of a text, summarize a text, locate key details, draw conclusions and inferences, and understand vocabulary used in context are skills that are necessary in both academic and workplace settings. As such, these skills are the focus of the Workplace Documents assessment and align this assessment to the skills defined in the Holistic Framework of education and work readiness.

The WorkKeys Applied Math assessment draws on skills defined as part of the core academic skills and skills defined as a part of cross-cutting capabilities. The skills constituting the Applied Math assessment align broadly with the skills defined within the mathematics section of core academic skills. At the same time, because examinees are applying mathematical skills in various ways to make decisions, the assessment construct overlaps with cross-cutting capabilities.

The cross-cutting capabilities that align to the WorkKeys Applied Math skills include troubleshooting (finding and/or correcting errors) and finding an optimal solution from among two or more options (including identifying the correct equation). These skills align to the Thinking and Metacognition capabilities within the Holistic Framework, including critical thinking, problem-solving, decision-making, computational thinking, and metacognition. Based on several workplace competency models, these skills are all identified as critical for work readiness skills (Institute for the Future, 2011; NNBIA, 2014).

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 Interpretation of Data Strand: Substrands Data Analysis; Gathering and Presenting Data)

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

3. Evaluating bias in the use of a graph (HF Interpretation of Data Strand: Substrands Data Analysis; Gathering and Presenting Data; HF Evaluation of Models, Inferences, and Results Strand: Substrands Scientific Reasoning and Argument; Modeling. HF Cross Cutting Science Concepts Strand: Substrands Patterns; Cause and Effect)

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

1.8 Alignment of ACT WorkKeys Skills With National Adult Education Academic Standards

The College and Career Readiness Standards for Adult Education (CCRSAE) were developed in 2014 using a subset of the Common Core State Standards (CCSS) for Mathematics and English Language Arts/Literacy that were deemed most relevant to adult education. In 2020–21, ACT convened a panel of external adult educational subject matter experts to conduct a crosswalk analysis evaluating the alignment between the content on the WorkKeys NCRC assessments and the CCRSAE.

WorkKeys NCRC assessments measure the mathematics, literacy, and critical thinking and problem-solving skills necessary for career success across occupations. Individuals who take the Workplace Documents, Applied Math, and Graphic Literacy assessments can qualify to earn the WorkKeys NCRC, a nationally portable credential that certifies career readiness. The ACT WorkKeys NCRC Crosswalk to College and Career Readiness Standards for Adult Education (CCRSAE) report (ACT, 2021) shows that the WorkKeys NCRC suite of assessments is designed to measure skills and knowledge that directly contribute to employability and work success, and which are integral to the CCRSAE. Using WorkKeys in adult education programs prepares individuals to show measurable skills gains for accountability requirements—while also earning an NCRC to gain employment opportunities.

“The National Reporting System (NRS) is the accountability system for the federally funded, State-administered adult education program. It embodies the accountability requirements of the Workforce Innovation and Opportunity Act (WIOA, the Act) for the adult education and literacy program (Title II) and reporting under WIOA” (Division of Adult Education and Literacy et al., 2021, p. 1). CCRSAE provides the foundation for the NRS Educational Functioning Levels (EFLs). In 2022, ACT convened two panels of external adult educational subject matter experts to participate in a content validity study. One panel matched the items from four WorkKeys Workplace Documents test forms to the individual reading standards used to describe the NRS educational functioning levels for reading and then set an overall EFL for each item. A separate panel completed the same activity using four WorkKeys Applied Math test forms and NRS EFLs for mathematics. WorkKeys Applied Math and Workplace Documents assessments are approved for use by NRS, and authorized by WIOA.

1.9 Additional WorkKeys Assessments

In addition to the three WorkKeys NCRC assessments required for the NCRC (Workplace Documents, Applied Math, and Graphic Literacy), the ACT WorkKeys suite also includes workplace assessments in Applied Technology, Business Writing, Workplace Observation, Fit (interests matched to work environments), and Talent (work-related attitudes and behaviors).

Chapter 2: Test Development

Chapter 2 is divided into three sections describing ACT® WorkKeys® Workplace Documents, ACT® WorkKeys® Applied Math, and ACT® WorkKeys® Graphic Literacy, respectively. Within each section is a construct definition, including a description of the stimuli for each level, the skill domains with subskill definitions, and performance level descriptors (PLDs). Following the construct definition is the item-writing and development process, including content and fairness reviews and pretesting. A list of subject matter experts (SMEs) concludes each section.

2.1 Workplace Documents

2.1.1 Workplace Documents—Overview

WorkKeys Workplace Documents is designed to assess the extent to which individuals can read and comprehend written documents in order to do a job. The documents—which include, but are not limited to, messages, emails, letters, directions, signs, notices, bulletins, policies, websites, contracts, and regulations—are based on materials that reflect the actual reading demands of the workplace. The ability to read and comprehend written information is critical for workplace success. The Workplace Documents assessment measures skills that individuals use when they read workplace documents and use that information to make decisions and solve problems.

To ensure that the Workplace Documents assessment measures useful and relevant skills, a team composed of individuals from within ACT including Test Development, Content, Measurement and Research, Industrial/Organizational Psychology, and Assessment Design was established to design the specifications for the Workplace Documents assessment. The team pooled resources to define the Workplace Documents construct and 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]

[1] Eleven external SMEs reviewed the Workplace Documents test development documentation and provided feedback. The SMEs were provided notebooks that included information on the definition of workplace reading, a description of the difference between reading in the classroom and reading in the workforce, cognitive skill domains and subdomains, sample items, and related questions. The SMEs reviewed the notebooks 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 make comments and notes in their notebooks and return them to ACT. Based on this feedback, the design team made modifications to all related materials. The individuals who served as external SMEs are provided in the table below along with their affiliations.

Through a review of the pertinent empirical and professional literature and through deliberations among team members, the team determined that the Workplace Documents construct was defined through the interplay of three aspects: document level complexity, reading skills, and document types. Although each aspect is defined separately, collectively they interact to provide meaning and interpretability to test scores.

For the Workplace Documents construct, reading and skill progressions are highly relevant. As a result, the team began by defining the characteristics of different levels of reading difficulty and then by identifying the pertinent associated reading skills.

2.1.2 Document Level Complexity

Document level complexity refers to the text complexity of the reading documents examinees are required to read in order to respond to the items. The design team organized document level complexity into five levels. Document (text) complexity for the Workplace Documents assessment is defined by the document's word count, reading level, clarity, amount of detail, and vocabulary level (including the use of technical terms, jargon, and acronyms). Additionally, different document types are permitted at specific levels. Table 2.1 provides the Workplace Documents complexity criteria along with the descriptor for each level.

Table 2.1. Workplace Documents—Passage Level Complexity Descriptors

WD Document Criteria	Level 3	Level 4	Level 5	Level 6	Level 7
Word Count Range	80–150	100–200	150–350	200–450	250–500
Flesch-Kincaid Reading Grade Level*	6	7.5	10	12	13
How complex is the stimulus document?	Short with no extra information and simple sentences	Straightforward with some longer sentences; may contain conditional situations	Mostly clear and direct, but with multiple details; may have complex sentences and/or contain conditional situations	Somewhat complicated sentences; document may be long and/or complex and/or contain conditional situations	Complex sentences with many details; may cover uncommon topics and/or contain conditional situations
Is the information in the document clearly stated?	Yes	Yes, mostly	Not necessarily; may need to make inferences	No; information is often not explicit	No; pieces of information may be spread throughout documents and may be extraneous
How detailed is the document?	Not very; will include a small number of details	There are a number of details	There are many details and some may be extraneous	There are implied and/or extraneous details	There are many implied and extraneous details
How difficult is the vocabulary?	Common, familiar, not difficult	Not too difficult; common vocabulary with some advanced words	Unfamiliar words, professional jargon, and acronyms; may need to use context to determine correct meaning	Difficult words, professional jargon, and technical terms; meanings may need to be determined from context	Advanced, unfamiliar, and/or uncommon words, technical terms, and professional jargon; meanings must be determined from context
Document Type	Informational, Instructional, Policy	Informational, Instructional, Policy	Informational, Instructional, Policy, Contracts, Legal, Multiple Related Documents	Informational, Instructional, Policy, Contracts, Legal, Multiple Related Documents	Informational, Instructional, Policy, Contracts, Legal, Multiple Related Documents

*The Flesch-Kincaid Reading Grade Level is a quantitative measure of the level of readability.

2.1.2.1 Document Classification Evaluation

ACT conducted a study to evaluate the content specialists' ability to consistently classify different reading passages into the five levels by applying the criteria described in Table 2.1.

The study asked four content specialists who regularly worked on the Workplace Documents assessment to discuss how they classified workplace documents and the merits of using the table to determine the level of such documents. Following the discussion, the four content specialists independently evaluated 20 reading passages and classified them into one of the five levels.

ACT utilized Generalizability Theory (Brennan, 2001) to analyze the consistency of the content specialists' categorizing. 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 0.93 and a Phi Coefficient of 0.92. These consistency indices revealed that the four content specialists, using Table 2.1 along with their training, classified workplace documents in a relatively consistent manner.

2.1.3 Workplace Documents Skill Domain Definitions

ACT's reading content specialists reviewed the original list of reading skills measured through the Reading for Information assessment. They determined that several of the defined skills overlapped and caused confusion in identifying the skill that aligns to the item. Consequently, the design team concluded that many of the old Reading for Information skill definitions were confusing and needed to be simplified.

In an effort to achieve greater clarity regarding the skill definitions, the content specialists reviewed the professional literature on reading and the workplace, and they asked the external SMEs for direction and insight. Through this work, they concluded that three primary skill domains exist for reading: comprehending written text, interpreting written text, and applying information and instructions derived from written text to workplace situations. As a result, the design team identified three primary reading-related workplace skills:

- Identify main ideas and details
- Apply instructions or information
- Identify meanings and definitions of words or phases

From these three primary skills, they defined a progression of reading subskills within each primary skill relevant to workplace applications. The workplace reading skills and subskills progression is presented in Table 2.2.

Table 2.2. Workplace Documents Skills and Subskills

Skills and Subskills	
1.0	**Identify Main Ideas and Details**
1.1.a	Identify the main idea
1.1.b	Identify the rationale behind an entire document or a section of a document *Identify an underlying reason for a task or procedure. Often “What is the main reason . . .?”*
1.2.a	Identify specific details
1.2.b	Infer implied details *The details needed to complete a task or procedure are not explicit at all; inferences need to be made to determine the necessary information*
2.0	**Apply Instructions and Information**
2.1	Choose when to perform a step in a series of steps *Often includes questions such as “What should you do first/next/last?”*
2.2.a	Apply information/instructions to a described situation *Identify the necessary information/instructions to complete a task and correctly apply them to a situation described in the document: “You should . . .”*
2.2.b	Apply information/instructions to a situation not directly described or to a completely new situation *Identify the necessary information/instructions to complete a task and correctly apply them to a situation that is not described in the document*
2.2.c	Apply principles inferred from a passage to a situation not directly described or to a completely new situation *Infer the reasons behind instructions/information described in the document and correctly apply them to a situation that is not described*
3.0	**Identify Meanings and Definitions of Words and Phrases**
3.1	Infer the meaning of a word or phrase from context (not jargon or technical terms) *Infer the correct meaning of words and phrases as they are used in a specific workplace scenario from the context of the document*
3.2.a	Identify the meaning of an acronym, jargon, or a technical term *Identify the meaning of words, phrases, acronyms, or jargon that have an exclusive meaning in a particular job or career cluster*
3.2.b	Infer the meaning of an uncommon acronym, jargon, or a technical term from context *Infer the meaning of words, phrases, acronyms, or jargon that have an exclusive meaning in a particular job or career cluster from the context of the document*

After developing the skills and subskills along with having the external SMEs review them, the team concluded that the critical workplace reading skill was “apply instructions and information.” In a workplace context, employers and supervisors are most concerned that workers not only are able to read and understand written texts, but, more importantly, that they understand how and when to apply instructions and information contained in the documents. Being able to apply information appropriately and accurately is critical to being a successful worker, and in today’s workplace, much of the information is presented to workers in written documents.

2.1.4 Workplace Documents—Multiple Related Documents

Reading for Information has traditionally utilized five document types for developing documents: instructions, informational, policies, legal, and contracts. In the redesign of the assessment, ACT has expanded these documents to include multiple related documents. The rationale for this change is that real-world reading situations often require an individual to identify information from multiple documents, make connections and conclusions, and apply this information to accomplish tasks.

The definition of multiple related documents is that they

- consist of two or more documents that are related or cover a common topic and
- have two or more authors.

Examples of these multiple related documents may include:

- an email string
- two webpages on a similar topic
- a company policy followed by a question raised in an email or message by a client or customer
- a formal document followed by an informal document that elaborates or explains

The Workplace Documents team considered it critical to include this type of document in the assessment if it is to accurately represent the type of reading content and context workers use on a daily basis. It provides examinees the opportunity to demonstrate that they are able to read complex text materials, understand and apply differing perspectives, and utilize the information contained in these documents to complete workplace tasks. In many ways, this step not only represents a unique passage type, but is also a step toward the inclusion of authentic and up-to-date reading passages and item tasks (Binkley et al., 2012).

2.1.5 Workplace Documents—Performance Level Descriptors

The Workplace Documents construct is defined through a combination of the text complexity level of a reading passage and the skill elicited by the item. Based on the text complexity level and skill, the design team was able to define the Workplace Documents performance level descriptors (PLDs).

Level 3—Document types include informational, instructional, and policy-related materials.

Examinees scoring at Level 3 are able to read and comprehend relatively short workplace documents which contain no extra information. The document contains short sentences using common, everyday workplace vocabulary. All the information in these documents is clearly and directly stated, and it contains a small number of details. In reading these documents, they are able to do the following:

- Identify the main idea
- Identify specific details
- Choose when to perform a step in a series of short steps
- Apply information/instructions to a situation that is the same as the situation in the reading materials

Level 4—Document types include informational, instructional, and policy-related materials.

Examinees scoring at Level 4 have the skills defined at Level 3 and in addition are able to read and comprehend workplace documents written in straightforward sentences that use familiar vocabulary and the occasional use of conditionals and a few advanced words. In reading these documents, they are able to do the following:

- Identify the main idea
- Identify specific details
- Use the reading materials to figure out the meanings of words that are not defined for them
- Choose when to perform a step in a series of steps
- Apply information/instructions to a situation that is the same as the situation in the reading materials
- Choose what to do when changing conditions call for a different action

Level 5—Document types include informational, instructional, policy-related, contractual, legal, and multiple related document materials.

Examinees scoring at Level 5 have the skills defined at Levels 3 and 4 and in addition are able to read and comprehend longer workplace documents written in more complex sentences that use more advanced vocabulary, including unfamiliar technical words, jargon, and acronyms. The information in Level 5 documents is generally stated directly, but specific details may be more difficult to find because of extraneous information. In reading these documents, they are able to do the following:

- Identify specific details
- Infer the meaning of a word or phrase from context
- Apply information/instructions to a new situation that is similar to the one described in the document while considering changing conditions
- Apply information/instructions that include conditions to situations described in the document
- Identify the appropriate meaning of an acronym, jargon, or technical term defined in the document
- Apply technical terms and jargon to stated situations
- Make some inferences to accomplish a goal

Level 6—Document types include informational, instructional, policy-related, contractual, legal, and multiple related document materials.

Examinees scoring at Level 6 have the skills defined at Levels 3, 4, and 5 and in addition are able to read and comprehend longer workplace documents written in lengthy, complex sentences that use advanced vocabulary, including unfamiliar words, jargon, and acronyms where the meaning is often implied. In reading these documents, they are able to do the following:

- Infer implied details
- Infer the meaning of an acronym, jargon, or technical term from context
- Apply information/instructions to a situation not directly described or to a completely new situation
- Apply principles inferred in a passage to a situation not directly described or to a completely new situation
- Identify the rationale behind an entire document or a section of a document

Level 7—Document types include informational, instructional, policy-related, contractual, legal, and multiple related document materials.

Examinees scoring at Level 7 have the skills defined at Levels 3, 4, 5, and 6 and in addition are able to read and comprehend long workplace documents containing many details and written using lengthy, complex sentences that use advanced vocabulary (including esoteric words, jargon, and acronyms) where meanings must be inferred from context. In reading these documents, they are able to do the following:

- Infer implied details
- Infer the meaning of an acronym, jargon, or technical term from context
- Apply information/instructions to a situation not directly described or to a completely new situation
- Apply principles inferred in a passage to a situation not directly described or to a completely new situation
- Identify the rationale behind an entire document or a section of a document

2.1.6 Designing Items to Elicit Examinee Evidence of Reading Workplace Documents

Workplace Documents uses multiple-choice items to measure examinees' proficiency in reading and comprehending workplace texts to gain information and guidelines to apply in workplace situations. The domain of workplace reading skills measured by the assessment was defined in 2018 by the design team and confirmed by external SMEs with backgrounds in business, industry, and education (see Table 2.3). To properly elicit evidence of the skills in the workplace reading 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 Educational Measurement [NCME], 2014).

2.1.6.1 Item Writing

Item writers qualify to write for the Workplace Documents 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 workplace reading, the training also provides explicit direction in terms of acceptable workplace reading texts. Once an item writer has successfully completed all required training modules, the next step is completing 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 the following:

- All items are linked to a workplace-oriented passage
- Workplace passages are written documents taken from workplace situations and scenarios
- Workplace passages are evaluated by the content team in terms of workplace realism
- Workplace passages are evaluated by the content team and classified into one of the five levels
- Workplace passages must be designed for one of the following purposes: (a) convey information to workers; (b) instruct workers on a procedure, process, or other activity; (c) convey a workplace policy; (d) convey contractual and/or legal information; and (e) convey information in multiple related documents written by two or more authors
- Multiple items should be developed for each workplace passage
- Each item is aligned to one of the skills defined as part of the construct

2.1.6.2 Item Review

After items have been developed, edited, and tentatively finalized by the Content Assessment team, they are submitted to external consultants with backgrounds in workplace reading and documents 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.

A separate review requires reviewers to evaluate the item and the reading passage on the basis of fairness and cultural bias. The reviewer is asked to evaluate the item and passage 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 Black, Latinx, 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.

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.

2.1.6.3 Item Pretesting

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

ACT conducts statistical analyses to determine if each pretest item meets required statistical criteria. ACT analyzes the items using both classical and item response theory (IRT-based) statistics to evaluate their 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 Workplace Documents content team reviews it and considers whether it should be edited, modified, or removed from the pool. When items are edited, 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 Workplace Documents 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.)

Table 2.3. Workplace Documents—External Subject Matter Experts

Name	Institution	Qualifications
Beverly Deal	S.B. Phillips	Workforce Readiness Director
Ana Gilbertson	Kirkwood Community College	Advanced Manufacturing Department Coordinator
Julia Holdridge	Sedgwick Industries	Director, Colleague Resources
Randy Lane	Eastman Chemical	ACT Job Profiler; Industrial Engineer
Chris Manheim	Manheim Solutions (Independent Consultant)	President and ACT Job Profiler
Scott Oppler	Vice President of Exam Development & Research, Society for Human Resource Management	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
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

Note. SME information current as of 2016.

2.2 Applied Math

2.2.1 Applied Math—Overview

WorkKeys Applied Math is designed to assess the extent to which individuals can use mathematical skills needed in workplaces. The ability to think problems through to find and evaluate solutions is critical for workplace success (Australian Association of Mathematics Teachers Inc., 2014; Smith, 1999). The Applied Math assessment measures skills that individuals use when they apply mathematical reasoning and problem-solving to work-related problems.

To ensure that the Applied Math assessment measures useful and relevant skills, a team composed of individuals from within ACT, including Test Development, Content, Measurement and Research, Industrial/Organizational Psychology, and Assessment Design was established to design the specifications for the Applied Math assessment. The team pooled resources to define the Applied Math 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.[2]

Through a review of the pertinent empirical and professional literature and thorough deliberations among team members, the team determined that the Applied Math construct was defined through a combination of the test item characteristics and the mathematics skill elicited by the item. (This conclusion was a modification of the current Applied Math definition that defined the construct as an interaction of mathematics skills, applications, and level of complexity.) For example, a Level 5 item must meet the content criteria (identified in Table 2.1) and assess a mathematics skill identified as a Level 5 skill (see Tables 2.5 through 2.10). Both the item/stem characteristics and the mathematics skills were aligned to a level of difficulty for the assessment.

2.2.2 Applied Math Domain

The design team carefully reviewed information and research assessing the uses of workplace Applied Mathematical skills. Through multiple discussions and reviews, the team decided that six general Applied Mathematics skills constituted the domain. Each of the six skills was further defined into a set of subskills within the skill domain.

The six general Applied Mathematics skills are

- basic operations including decimals,
- fractions,
- percentages/ratios/proportions,
- unit conversions,
- geometric measurement, and
- applied math reasoning.

More information regarding these dimensions is provided throughout Chapter 2.
Tables 2.5–2.10 provide the subskills defined within each skill.

[2] Eleven external SMEs reviewed the Applied Mathematics test development documentation and provided feedback. The SMEs were provided notebooks that included information on the definition of workforce applied mathematics, a description of the difference between mathematics in the classroom and mathematics in the workforce, cognitive skill domains and subdomains, sample items, and related questions. The SMEs reviewed the notebooks 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 make comments and notes in their notebooks and return them to ACT. Based on this feedback, the design team made modifications to all related materials. The individuals who served as external SMEs are provided in the table below along with their affiliations.

2.2.3 Revisions to Applied Math Domain as a Result of Review

Consultation with SMEs revealed technology (particularly spreadsheets, calculators, and scanning devices) removed many of the computational demands from the workplace. Despite these advanced tools, employees still needed mathematical and quantitative reasoning skills. For example, employees utilize spreadsheets to do calculations, but they must be capable of troubleshooting and finding errors in cells that are automatically calculated. Furthermore, in production situations, employees need to be able to understand and interpret measures of central tendency, spread, and tolerances, particularly as they relate to quality control. Given the evaluation and feedback, the following skills were included in the revised Applied Math assessment:

1. Troubleshooting was expanded to include identifying whether an error occurred. In these cases, examinees must identify where values are incorrect.

2. Basic statistical concepts were expanded beyond calculating means and medians to include calculating weighted means and tolerances, as well as interpreting measures of central tendency and spread. Examinees might be asked to interpret or make a decision based on statistical values, but they are not required to calculate the values. Interpreting these values is considered within the construct of the Applied Math assessment; calculating these values is considered outside of the construct.

3. Identifying the correct equation was added to assess examinees’ skill in creating and comprehending equations used to produce automated calculations.

2.2.4 Applied Math—Item Stem Characteristics

The Applied Math assessment team found that the successful application of mathematics in a workplace situation incorporates problem-solving and integrates mathematical and quantitative reasoning skills (Australian Association of Mathematics Teachers Inc., 2014; Smith, 1999). While the proliferation of spreadsheets and calculators in the workplace has reduced the need for computational skills, employees still need to be able to apply quantitative reasoning skills to solve complex problems. Item levels are determined by the situational and problem-solving complexity along with the mathematical skill and reasoning required.

Because the Applied Math assessment presents realistic workplace problems, each item is defined, in part, by its context. To assist in creating realistic workplace problems, each test item is presented in the context of money, time, measurement, or quantity.

Additionally, WorkKeys defines Applied Math items as having varying degrees of complexity. The complexity of each item is determined by the following dimensions:

- Presentation of quantitative information (Is the quantitative information presented in the order required to set up the problem?)

- Amount of language that must be understood to translate it into a mathematics problem

- Whether extraneous information is included

- Whether the item contains a graph
- Whether solving the problem requires multiple steps

The design team developed Table 2.4 to provide guidelines on how item complexity influences item level.

Table 2.4. Item Stem Characteristics by Level

Item Stem Characteristics	Level 3	Level 4	Level 5	Level 6	Level 7
Presentation of Quantitative Information	Presented in logical order	May not be in logical order	May not be in logical order	May not be in logical order	May have incomplete information or require an assumption
Amount of Language to Translate to Math Expression	Minimal	Some	Some	Considerable translation	May have unusual format
Extraneous Information	None	May have some extraneous information	May have some extraneous information	May have some extraneous information	May have some extraneous information
Contains Simple Graph	No	May be included	May be included	May be included	May be included
Setup/Planning	Minimum setup	Some setup required	May require complicated setup	May require complicated setup	May require complicated setup
Calculations	One operation	One or two operations	May have several operations	May have several operations	May have several operations
Solving for Unknowns	Solve for one unknown	Solve for one or two unknowns	May solve for one unknown and then use to solve the problem to answer the question	May solve for one unknown and then use to solve the problem to answer the question	May solve for one unknown and then use to solve the problem to answer the question

Using this table: The table is intended as a guide describing the general characteristics of the item/stem for each given level.

2.2.5 Applied Math—Skill Definitions

The Applied Math assessment strives to measure the most relevant and consequential foundational mathematical skills that are widely used in the workplace. To determine these skills, the design team drew upon information from the ACT® WorkKeys® JobPro® database, professional literature, and feedback from external SMEs.

The Applied Math domain was defined through six critical skills. Each of the six skills was subdivided into subskills. The skill and subskill definitions collectively constitute the workplace applied math construct. Tables 2.5–2.10 provide the subskills that constitute each Applied Math skill.

Table 2.5. Skill 1.0—Basic Operations with Numbers Including Decimals

1.0		Basic Operations with Numbers Including Decimals
Subskill	1.1	Add positive numbers
	1.2	Add with negative number(s)
	1.2.1	Add more than four numbers, some of which may be negative
	1.2.2	Add two negative numbers
	1.3	Subtract positive numbers
	1.3.1	Subtract positive numbers where the result is positive
	1.3.2	Subtract positive numbers where the result is negative
	1.4	Subtract with negative number(s)
	1.4.1	Positive minus negative
	1.4.2	Negative minus positive
	1.4.3	Negative minus negative
	1.5	Multiply positive numbers
	1.6	Divide positive numbers (result could be a fraction)
	1.7	Two or more basic operations

Table 2.6. Skill 2.0—Fractions

2.0		Fractions
Subskill	2.1	Add/Subtract fractions
	2.1.1	Add/Subtract fractions (limited to halves and fourths) No more than two fractions
	2.1.2	Add/Subtract fractions that share a common denominator (such as $\frac{1}{8} + \frac{3}{8} + \frac{7}{8}$)
	2.1.3	Add/Subtract fractions with unlike denominators
	2.2	Multiply fractions
	2.2.1	Multiply fractions (none are mixed numbers)
	2.2.2	Multiply a mixed number (such as $12\frac{1}{8}$) by a whole number or a decimal
	2.2.3	Multiply more than 1 mixed number
	2.3	Divide fractions
	2.4	Change between fractions and decimals

Table 2.7. Skill 3.0—Percentages/Ratios/Proportions

3.0		Percentages/Ratios/Proportions
Subskill	3.1	Convert between decimals and percentages
	3.2	Calculate a given percentage of a given number (e.g., what is 4% of 10? Tax, commission, discount, markup, raise)
	3.3	Calculate the percentage one number is of another (e.g., 6 is what percentage of 15?)
	3.4	Calculate percent change
	3.5	Calculate reverse percent (e.g., you have discounted a coat by 15% and now the sale price is $30; what was the original price?)
	3.6	Set up and/or manipulate simple ratios/proportions/rates
	3.6.1	Figure out simple ratios
	3.6.2	Figure out simple proportions
	3.6.3	Figure out simple rates (such as 10 mph)
	3.7	Set up and/or manipulate ratios, rates, or proportions (at least one of the quantities related is a fraction)
	3.8	Rates, production rates, rate × time (e.g., 15 cups over 40 mins = x cups per minute; at 59 units per hour, how many made in 8 hours?)

Table 2.8. Skill 4.0—Unit Conversions

4.0		Unit Conversions
Subskill	4.1	Convert between familiar units (e.g., between hours and minutes, dollars and cents)
	4.2	Convert where the conversion factor is given in the problem
	4.3	Convert where you must select the conversion factor (e.g., from the formula sheet)
	4.4	Two or more step conversions (e.g., inches to feet to yards, kilometers to meters to feet)
	4.5	Two or more separate conversions (e.g., problem that has minutes to hours and pounds to ounces)
	4.6	Operations with mixed units (e.g., add 6 feet 4 inches and 3 feet 8 inches, 3.5 hours + 4 hours 30 minutes, etc.)
	4.7	Convert the unit of measurement using fractions, mixed numbers, decimals, or percentages

Table 2.9. Skill 5.0—Geometric Measurement

5.0		Geometric Measurement
Subskill	5.1	Calculate perimeter or circumference
	5.2	Calculate area
	5.2.1	Find the area of one rectangle with dimensions given
	5.2.2	Find the area of other polygons with dimensions given
	5.2.3	Find the area of a circle given radius or diameter
	5.2.4	Find the area of multiple shapes
	5.2.5	Find the area of a composite shape
	5.2.6	Find the area when it may be necessary to rearrange the formula, convert units of measurement in the calculations, or use the result in further calculations
	5.3	Calculate volume
	5.3.1	Calculate volume of a rectangular solid
	5.3.2	Calculate volume of spheres, cylinders, and cones
	5.3.3	Find the volume when it may be necessary to rearrange the formula, convert units of measurement in the calculations, or use the result in further calculations

Table 2.10. Skill 6.0—Applied Math Reasoning

6.0		Applied Math Reasoning
Subskill	6.1	Troubleshooting
	6.1.1	Identify where a mistake occurred (e.g., in the spreadsheet, identify the row where the problem occurred)
	6.1.2	Identify the reason for the mistake
	6.2	Best Deal
	6.2.1	Find the best deal using one- or two-step calculation that meets the stated conditions
	6.2.2	Find the best deal from a group and then do something with the answer
	6.2.3	Determine the better economic value of several alternatives by using graphics, or determining the percentage difference, or by determining unit cost
	6.3	Basic Statistical Concepts
	6.3.1	Calculate the average (mean)
	6.3.2	Calculate the weighted average
	6.3.3	Interpret measures of central tendency
	6.3.4	Interpret measures of spread and tolerances
	6.4	Identify the Correct Equation

2.2.6 Applied Math—Performance Level Descriptors

Individuals taking the assessment earn a scale score and a level score. Scale scores range from 65 to 90. The scale scores are transformed to level scores ranging from Level 3 to Level 7. Most examinees focus on their level scores because they have interpretability related to job skills. Consistent with the other WorkKeys Assessments required for the NCRC, Level 3 is defined as the lowest level at which an employer would be willing to hire and pay an employee to perform those skills in a job requiring applied mathematics. (An individual may perform poorly and not earn a Level score of less than 3; in this case, the individual has not achieved a WorkKeys level.) Level 7 is defined as the highest skill level that an employee could be expected to hold without specialized formal training.

Applied Mathematics score levels are interpreted as a progression in that a test taker who holds skills at a specific level will be able to do the skills defined for each lower level. For example, a test taker who scores at Level 5 not only possesses the skills defined as Level 5 skills, but also possesses the skills defined at Levels 3 and 4.

The following section identifies performance level descriptors for examinees who earn scores at each level.

2.2.6.1 Applied Math Level 3

Level 3 problems can easily be translated from a word problem to a math equation requiring a single type of math operation. All the needed information is presented in logical order, and there is no extra information given. When test takers use Level 3 Applied Math skills, they are able to do the following:

- Solve problems that require one type of mathematical operation. They add or subtract either positive or negative numbers (such as 10 or –2). They multiply or divide using only positive numbers (such as 10).
- Convert a familiar fraction (such as $\frac{1}{2}$ or $\frac{1}{4}$) to a decimal and convert from a decimal to a common fraction; OR convert between decimals to percentages (such as 0.75 to 75%).
- Convert between familiar units of money and time (such as one hour equals 60 minutes or $\frac{1}{2}$ of a dollar equals $0.50).
- Add the prices of several products together to find the total and calculate the correct change for a customer.

2.2.6.2 Applied Math Level 4

In Level 4 problems, tasks may present information out of order and may include extra, unnecessary information. One or two operations may be needed to solve the problem. A chart, diagram, or graph may be included. When test takers use Level 4 Applied Math skills, they use the skills described at Level 3, and they also are able to do the following:

- Solve problems that require one or two mathematical operations. They can add, subtract, or multiply using positive or negative numbers (such as 10 or –2), and they can divide positive numbers (such as 10).
- Calculate the average or mean of a set of numbers (such as $\frac{(10+11+12)}{3}$). For this, they may use whole numbers and decimals.
- Figure out simple ratios (such as $\frac{3}{4}$), simple proportions (such as 10/100 cases), or rates (such as 10 mph).
- Add commonly known fractions, decimals, or percentages (such as $\frac{1}{2}$, 0.75, or 25%).
- Add or subtract fractions with a common denominator (such as $\frac{1}{4}+\frac{3}{4}+\frac{1}{4}$).
- Multiply a mixed number (such as $12\frac{1}{8}$) by a whole number or a decimal.
- Put the information in the right order before they perform calculations.

2.2.6.3 Applied Math Level 5

In Level 5 problems, the information may not be presented in a logical order, the item may contain extraneous information, it may contain a graph or diagram, and the mathematical setup may be complicated. In solving, the test taker may need to perform multiple operations. (For example, at this level, examinees may complete an order form by totaling an order and then calculating sales tax.) When test takers use Level 5 Applied Math skills, they use the skills described at Levels 3 and 4, and they also are able to do the following:

- Decide what information, calculations, or unit conversions to use to find the answer to a problem.
- Add and subtract fractions with unlike denominators (such as $\frac{1}{2} - \frac{1}{4}$).
- Convert units within or between systems of measurement (e.g., time, measurement, and quantity) where the conversion factor is given either in the problem or in the formula sheet.
- Solve problems that require mathematical operations using mixed units (such as adding 6 feet and 4 inches to 3 feet and 10 inches or subtracting 4 hours and 30 minutes from 3.5 hours).
- Identify the best deal using one- or two-step calculations that meet the stated conditions.
- Calculate the perimeter or circumference of a basic shape or calculate the area of a basic shape.
- Calculate a given percentage of a given number and then use that percentage to find the solution to a problem (e.g., find the percentage and then use it to find the discount, markup, or tax).
- Identify where a mistake occurred in a calculation (such as identifying the row in a spreadsheet where a problem occurred).

2.2.6.4 Applied Math Level 6

Level 6 problems may require considerable translation from verbal form to mathematical expression. They generally require considerable setup and involve multiple-step calculations. When test takers use Level 6 Applied Mathematics skills, they use the skills described at Levels 3, 4, and 5, and they also are able to do the following:

- Use fractions with unlike denominators and calculate reverse percentages.
- Convert units within or between systems of measurement (e.g., time, measurement, and quantity) where multiple-step conversions are required and the formulas are provided such as converting from kilometers to meters to feet.
- Identify why a mistake occurred in a solution.

- Find the best deal from a group of solutions and then use the result for another calculation.

- Find the area of basic shapes when it may be necessary to rearrange a formula, convert units of measurement in the calculations, or use the result in further calculations.

- Calculate the volume of rectangular solids (e.g., cubes).

- Calculate rates, production rates, rate by time (such as: production rate is 59 cups produced per hour, so how many will be produced in an 8-hour shift).

- Identify the correct equation for solving a problem.

2.2.6.5 Applied Math Level 7

Level 7 problems may be presented in an unusual format, and information presented may be incomplete or require the test taker to make an assumption. Problems often involve multiple steps of logic and calculation. When test takers use Level 7 Applied Math skills, they use the skills described at Levels 3, 4, 5, and 6, and they also are able to do the following:

- Solve problems that include ratios, rates, or proportions where at least one of the quantities is a fraction.

- Identify the reason for a mistake.

- Convert between units of measurement using fractions, mixed numbers, decimals, and percentages.

- Calculate volumes of spheres, cylinders, or cones.

- Calculate the volume when it may be necessary to rearrange the formula, convert units of measurement in calculations, or use the result in further calculations.

- Set up and manipulate ratios, rates, or proportions where at least one of the quantities is a fraction.

- Determine the better economic value of several alternatives by using graphics, determining the percentage difference, or determining unit cost.

- Apply basic statistical concepts. For example, calculate the weighted mean, interpret measures of central tendency, or interpret measure of spread and tolerance.

2.2.7 Designing Items to Elicit Evidence of Applied Math

Applied Math uses multiple-choice items to measure examinees' proficiency in various mathematical skills necessary for workplace success. The domain of mathematical skills measured by the assessment was defined in 2018 by the design team and confirmed by external SMEs with backgrounds in business, industry, and education (see Table 2.11). To properly elicit evidence of the skills in the Applied Mathematics domain, ACT follows an item-design model aligned with both evidence-centered assessment design (Mislevy et al., 1999) and the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

2.2.7.1 Item Writing

Item writers qualify to write for the Applied Math 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. Once an item writer has successfully completed all required training modules, the next step is completing an item-writing assignment that details the number of items to be developed at specific levels. The assignment may also include other item specifications such as Career Cluster alignment, the required level of stem complexity, the presence or absence of particular data displays, or other item-defining characteristics. What follow are other requirements that are universal to all Applied Math items:

- The context must be work-related and realistic, and the mathematics should be authentic to the work presented in the item.
- Prices, rates, and procedures in the item should be authentic and realistic for the next few years. Moreover, the source of the information regarding the prices, rates, and procedures should be documented.
- Avoid overlap with the Graphic Literacy assessment. That is to say, while graphics are allowed in an item, the mathematical skill must be the emphasis of the item rather than reading and interpreting the graphic.

2.2.7.2 Item Review

After items have been developed, edited, and tentatively finalized by the Content Assessment team, they are submitted to external consultants with backgrounds in workplace math assessment 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.

A separate review requires the reviewers to evaluate the item on the basis of fairness and cultural bias. The reviewer is asked to evaluate the item in terms of how members of different demographic groups would respond to the item. (ACT asks the item reviewer to evaluate the item from the perspective of men and women examinees, and from the perspective of Black, Latinx, 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.

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 content specialists conclude that all relevant issues are resolved.

2.2.7.3 Item Pretesting

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

ACT conducts statistical analyses to determine if each pretest item meets required statistical criteria. ACT analyzes the items using both classical and item response theory (IRT-based) 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 Applied Math content team reviews it and considers whether it should be edited, modified, or removed from the pool. When items are edited, 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 Applied Math 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.)

Table 2.11. Applied Mathematics—External Subject Matter Experts

Name	Institution	Qualifications
Beverly Deal	S.B. Phillips	Workforce Readiness Director
Ana Gilbertson	Kirkwood Community College	Advanced Manufacturing Department Coordinator
Julia Holdridge	Sedgwick Industries	Director, Colleague Resources
Randy Lane	Eastman Chemical	ACT Job Profiler; Industrial Engineer
Chris Manheim	Manheim Solutions (Independent Consultant)	President and ACT Job Profiler
Scott Oppler	Vice President of Exam Development & Research, Society for Human Resource Management	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
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

Note. SME information current as of 2016.

2.3 Graphic Literacy

2.3.1 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 measures 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.[3]

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.

2.3.2 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.

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,

[3] 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.

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 phenomena (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, Care, & McGaw, 2012; Koomey, 2017).

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 2.12 presents a description of the characteristics for each of the four categories of graphic complexity.

Table 2.12. 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 complexity 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.

2.3.2.1 Graphic Complexity Classification Evaluation

ACT conducted a study to evaluate the content specialists' ability to consistently classify different graphic materials into the four categories by applying the principles described in Table 2.12.

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 2.12. 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 using 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 2.12 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 2.12 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.

2.3.3 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 subdivided into subskills. A total of 13 subskills define the cognitive processes used by examinees to solve the graphic literacy problems. Table 2.13 presents the Graphic Literacy domain divided into four skills with each skill subdivided into subskills.

Table 2.13. 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 et al., 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 2.13) were integrated with the three-step cognitive skill model to derive the graphic literacy cognitive process model. Table 2.14 presents the Graphic Literacy Cognitive Process Model.

Table 2.14. Graphic Literacy—Cognitive Process Model

Cognitive skills	Cognitive steps		
	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	NA
Assess trends/ patterns/ relationships	NA	• 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	NA	• 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	NA	• 2.R.1 Identify the graphic that represents the data	• 3.R.1 Identify the most effective graphic given a defined purpose • 3.R.2 Justify the most effective graphic given a defined purpose

2.3.4 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 2.15.

Table 2.15. 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 one-step process is applied to a graphic of low moderate complexity, the performance level overall is defined as Level 3. However, if that same one-step process is applied to a graphic of high moderate complexity, the overall performance level is then defined as Level 4. In effect, when the same cognitive skill is applied to a more complex graphic, the task elicited by the item is at a higher performance level.

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

2.3.5 Graphic Literacy—Performance Level Descriptors

The Graphic Literacy construct is defined through Tables 2.12, 2.13, 2.14, and 2.15 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 skills:

- 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

2.3.6 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 2.16). 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 et al., 1999) and the *Standards for Educational and Psychological Testing* (AERA et al., 2014).

2.3.6.1 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, the next step is completing 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 the following:

- 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 they 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 that the program also includes an Applied Mathematics assessment.

The Graphic Literacy assessment is a measure of an examinee's ability to find information and solutions by applying graphic literacy skills, and 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.

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

Graphic Literacy items may have a limited number 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 × 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 × 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 × 100, 900 – 500, 70% + 10%)
- Use of fractions less than or equal to one; limit the denominator to 2, 3, 4, or 10 ($\frac{1}{2}, \frac{2}{3}, \frac{3}{4}$)
- 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.

2.3.6.3 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.

A separate review requires reviewers 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 Black, Latinx, 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 for their sensitivity to issues of cultural diversity and fairness. Reviewers' performance is regularly evaluated by ACT staff.

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

2.3.6.4 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-based (IRT-based) 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.)

Table 2.16. 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	Vice President of Exam Development & Research, Society for Human Resource Management	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

Note. SME information current as of 2016.

Chapter 3: Test Specifications

Chapter 3, like Chapter 2, is organized into three sections that describe the process used to define the test specifications for each exam. These specifications include the sources of information reviewed, content relevance and representativeness, and the test blueprint.

3.1 WorkKeys NCRC Specifications—Overview

The purpose of the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessments is to assist workers, students, employers, and workforce development leaders by providing a system to measure and improve individuals' skills. Chapter 1 provided evidence demonstrating that foundational skills required for success in the modern economy include the ability to read, comprehend, and apply information conveyed through written workplace documents; the ability to solve applied mathematical problems; and the ability to comprehend, interpret, apply, and construct information conveyed through graphics.

This chapter provides test specifications for the Workplace Documents, Applied Math, and Graphic Literacy assessments. An assessment's test specifications are created first by developing the assessment's claims and score interpretations, then 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 WorkKeys NCRC design 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 constructs on the three assessments and that provide evidence supporting score interpretations (Kane, 2013; Messick, 1989).

The design team used a variety of reputable source materials to identify relevant content that should constitute a measure of workplace reading, mathematics, and graphic literacy skills. Over the past 25 years, through its job-profiling services, ACT has gathered information related to job skills from the manufacturing, health care, construction, transportation, financial, and sales sectors. The team reviewed these findings and used the information to determine what types of reading materials, applied math problems, and graphics should be included on the assessments and which skills were most frequently required. To further support content-related decisions, the team reviewed professional literature around workplace reading (Binkley et al., 2012; Smith et al., 2000), workplace applied math (Australian Association of Mathematics Teachers, 2014; Binkley et al., 2012; Smith, 1999), and workplace graphic literacy (Binkley et al., 2012; Brumberger, 2011; Few, 2012). They also reviewed workplace competency models (NNBIA, 2014). Lastly, the team consulted a group of external subject matter experts (SMEs) to obtain their perspectives on workplace texts and reading skills, applied mathematics, and graphics. (See the list of participating SMEs in the Chapter 2 tables.)

Based on the findings from the review of these resources, ACT formulated test specifications for each assessment. 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 being measured (Messick, 1989; Mislevy et al., 1999). In the context of Workplace Documents, construct relevance requires that the examinee demonstrate the ability not only to read and comprehend a workplace document but also to apply the information conveyed by the document to a job task. For Applied Math, the examinee must demonstrate the ability not only to solve mathematical problems but also to use the tools commonly found in the workplace for solving quantitative problems. For Graphic Literacy, the examinee must demonstrate the ability not only to comprehend and interpret workplace graphics but also to apply the information conveyed by the graphic to a job task. Construct representativeness, meanwhile, refers to a range of reading passages, math problems, and graphics and the various reading, mathematical, and interpretive skills needed in the workplace. Materials must represent the full range of job sectors, from manufacturing to construction to office work and beyond, because the WorkKeys NCRC assessments are designed to measure skills that are applicable to a large number of jobs. Materials must also represent appropriate ranges of difficulty.

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

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

1. Content is construct relevant and representative.

2. Content representation is identical across alternate forms.

3.3 WorkKeys NCRC Assessments—Test Blueprints

3.3.1 Workplace Documents—Test Blueprint

ACT developed detailed blueprints defining the content attributes of each test item. The content specifications were developed by clearly specifying the attributes of a reading passage at each of five levels (see Chapter 2: Test Development). They were further defined by specifying the workplace reading skills and subskills. Within the test specifications table, each subskill was evaluated and aligned to one or more levels. Following the alignment of subskills, weights were determined based on the overall importance of the subskill to the construct of workplace reading (Allen & Yen, 2002).

The Workplace Documents construct was based on three critical facets:

- the reading complexity level of the passage
- the reading skill elicited by the item
- the document type

The reading complexity level was divided into five levels, each of which was defined based on word count, reading level, clarity, amount of detail, and vocabulary level (see Table 2.1). With these factors in mind, ACT content specialists evaluated each passage and determined its level.

Workplace reading skills were divided into three primary skills: identifying main ideas and details, applying instructions or information, and identifying meanings and definitions of words or phrases. According to professional literature on workplace reading and data from ACT's job profiling, workplace documents are used not just to communicate information but also to direct people toward specific actions. As a result, the skill of applying instructions and information received greater weighting as a measure of an individual's workplace reading skills.

Five document types were identified as relevant to workplace reading: informational, instructional, policy, legal, and multiple related. At the lowest two levels (Level 3 and Level 4), only informational, instructional, and policy documents were considered relevant for workplace reading. At the intermediate level (Level 5), informational, instructional, and policy were considered most relevant, although Level 5 passages may include a legal or multiple related document. At the two highest levels, all five document types were considered relevant.

Tables 3.1, 3.2, and 3.3 present the Workplace Documents test specifications. The test specifications provide a blueprint for form development and also represent the relative importance of the reading skills and subskills in the workplace.

Table 3.1. Skill Domain Item Distribution by Level

Domain	Number per Level					
	Level 3	Level 4	Level 5	Level 6	Level 7	Total
1.0 Identify the main idea and details	2	3	1	2	2	10
2.0 Apply instructions or information	2	4	4	3	2	15
3.0 Identify meanings and definitions of words and phrases	0	1	2	1	1	5
Total	4	8	7	6	5	30

Table 3.2. Skill Subdomain Item Distribution by Level

Subdomain	Number per Level					
	Level 3	Level 4	Level 5	Level 6	Level 7	Total
1.1.a Identify the main idea	1	1	0	0	0	2
1.1.b Identify the rationale behind an entire document or a section of a document	0	0	0	1	1	2
1.2.a Identify specific details	1	2	1	0	0	4
1.2.b Infer implied details	0	0	0	1	1	2
2.1 Choose when to perform a step in a series of steps	1	1	0	0	0	2
2.2.a Apply information/instructions to a described situation	1	3	2	0	0	6
2.2.b Apply information/instructions to a situation not directly described or to a completely new situation	0	0	2	1	0	3
2.2.c Apply principles inferred from a passage to a situation not directly described or to a completely new situation	0	0	0	2	2	4
3.1 Infer the meaning of a word or phrase from context (nonprofessional)	0	1	1	0	0	2
3.2.a Identify the meaning of an acronym, jargon, or technical term defined in a document	0	0	1	0	0	1
3.2.b Infer the meaning of an acronym, jargon, or technical term from context	0	0	0	1	1	2
Total	4	8	7	6	5	30

Table 3.3. Number of Passages for Each Document Type

Document Type	Number of Passages per Form	Max Number of Passages (Includes Pretest)
Instructional (INS)	3–5	5
Informational (INF)	2–4	4
Policy (POL)	2–4	4
Legal (LEG)	1–3	3
Multiple Related (MUL)	1–3	3

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

3.3.2 Applied Math—Test Blueprint

ACT developed detailed blueprints defining the content attributes of each test item. The content specifications were developed by clearly specifying the attributes of types of math problems that workers need to solve. Using this information, the team identified six primary applied mathematical skills, then defined subskills within each of the primary skills. Using the job-profiling data, the team weighted the criticality and frequency of use of each subskill (Allen & Yen, 2002). Doing this resulted in some subskills being removed. The weightings were then reviewed by the external SMEs, and based on the feedback, the team made final adjustments to the blueprint. Lastly, the team evaluated the problem context and set up a table recommending the problem context distribution.

The workplace Applied Math construct was based on three critical facets:

- the applied mathematical complexity level
- the applied mathematical skills and subskills
- the applied mathematical problem context

The applied mathematical complexity level was divided into five levels, each defined based on the presentation of quantitative information, the amount of language used to translate to a mathematical expression, the amount of extraneous information, the presence of a graphic, the planning and mathematical setup, and the number of unknowns (see Table 2.4).

Applied mathematical skills were divided into six primary categories: basic operations with numbers including decimals; fractions; percentages/ratios/proportions; unit conversions; geometric measurement; and applied mathematics reasoning. According to professional literature on applied mathematics and data from ACT's job profiling, workplace applied mathematics is conducted using tools (e.g., calculators and spreadsheets). Determining whether workers can effectively use tools to apply their mathematical skills and find the correct solution was thus deemed a critical component of the assessment.

Four contexts were identified as relevant to workplace applied mathematics: quantity, money, time, and measurement. The overwhelming majority of foundational applied mathematics workplace tasks involved one of these four contexts.

Tables 3.4 through 3.11 present the Applied Math test specifications. The test specifications provide a blueprint for form development and also represent the relative importance of the applied mathematics skills and subskills in the workplace.

Table 3.4. Applied Math Skills Item Distribution by Level

Domain	Number per Level					
	Level 3	Level 4	Level 5	Level 6	Level 7	Total
1.0 Basic operations with numbers including decimals	4	2	0	0	0	6
2.0 Fractions	1	2	1	0	0	4
3.0 Percentages/ratios/proportions	0	1	1	2	1	5
4.0 Unit conversions	1	0	2	1	1	5
5.0 Geometric measurement	0	0	1	1	1	3
6.0 Applied mathematics reasoning	0	1	2	2	3	8
Total	6	6	7	6	6	31

Table 3.5. Basic Operations: Subskill Item Distribution

1.0 Basic Operations With Numbers Including Decimals	Level 3	Level 4	Level 5	Level 6	Level 7	Total
1.1 Add positive numbers OR 1.2 Add negative numbers	1	0	0	0	0	1
1.3 Subtract positive numbers OR 1.4 Subtract negative numbers	1	0	0	0	0	1
1.5 Multiply positive numbers	1	0	0	0	0	1
1.6 Divide positive numbers	1	0	0	0	0	1
1.7 Perform two or more basic operations	0	2	0	0	0	2
Total	4	2	0	0	0	6

Table 3.6. Fractions: Subskill Item Distribution

2.0 Fractions	Level 3	Level 4	Level 5	Level 6	Level 7	Total
2.1 Add/subtract fractions with a common denominator	0	1	0	0	0	1
2.1.3 Add/subtract fractions with unlike denominators	0	0	1	0	0	1
2.2.2 Multiply a mixed number (such as $12\frac{1}{8}$) by a whole number	0	1	0	0	0	1
2.4 Change between fractions and decimals OR 3.1 Convert between decimals and percentages	1	0	0	0	0	1
Total	1	2	1	0	0	4

Table 3.7. Percentages/Ratios/Proportions: Subskill Item Distribution

3.0 Percentages/Ratios/Proportions	Level 3	Level 4	Level 5	Level 6	Level 7	Total
3.2 Calculate a given percentage of a given number (e.g., what is 4% of 10? Tax, commission, discount, markup, raise) OR 3.3 Calculate the percentage one number is of another number	0	0	1	0	0	1
3.4 Calculate percent change OR 3.5 Calculate reverse percent	0	0	0	1	0	1
3.6 Set up and/or manipulate simple ratios, proportions, or rates (includes 3.6.1, 3.6.2, and 3.6.3)	0	1	0	0	0	1
3.7 Set up and/or manipulate ratios, rates, or proportions (at least one of the quantities is related to a fraction)	0	0	0	0	1	1
3.8 Rates, production rates, rate × time (e.g., 15 cups over 40 minutes = x cups per minute; at 59 units per hour, how many made in 8 hours?)	0	0	0	1	0	1
Total	0	1	1	2	1	5

Table 3.8. Unit Conversions: Subskill Item Distribution

4.0 Unit Conversions	Level 3	Level 4	Level 5	Level 6	Level 7	Total
4.1 Convert between familiar units (e.g., between hours and minutes, dollars and cents)	1	0	0	0	0	1
4.2 Convert where the conversion factor is given in the problem OR 4.3 Convert where you must select the conversion factor (from the formula sheet)	0	0	1	0	0	1
4.4 Conversions requiring two or more steps (e.g., inches to feet to yards, kilometers to meters to feet) OR 4.5 Two or more separate conversions	0	0	0	1	0	1
4.6 Operations with mixed units (e.g., add 6 feet and 4 inches to 3 feet and 8 inches; add 3.5 hours to 4 hours and 30 minutes)	0	0	1	0	0	1
4.7 Convert the unit of measurement using fractions, mixed numbers, decimals, or percentages	0	0	0	0	1	1
Total	1	0	2	1	1	5

Table 3.9. Geometric Measurement: Subskill Item Distribution

5.0 Geometric Measurement	Level 3	Level 4	Level 5	Level 6	Level 7	Total
5.1 Calculate perimeter or circumference OR 5.2 Calculate area (includes 5.2.1, 5.2.2, and 5.2.3)	0	0	1	0	0	1
5.2.6 Find the area of basic shapes when it may be necessary to rearrange the formula, convert units of measurement in the calculations, or use the result in further calculations OR 5.3 Calculate volume (includes 5.3.1)	0	0	0	1	0	1
5.2.4 Find the area of multiple shapes OR 5.2.5 Find the area of a composite shape OR 5.3.2 Calculate volume of spheres, cylinders, and cones OR 5.3.3 Find the volume when it may be necessary to rearrange the formula, convert units of measurement in the calculations, or use the result in further calculations	0	0	0	0	1	1
Total	0	0	1	1	1	3

Table 3.10. Applied Mathematical Reasoning: Subskill Item Distribution

6.0 Applied Mathematical Reasoning	Level 3	Level 4	Level 5	Level 6	Level 7	Total
6.1 Troubleshooting	—	—	—	—	—	—
6.1.1 Identify where a mistake occurred (e.g., identify the row in a spreadsheet where the problem occurred)	0	0	1	0	0	1
6.1.2 Identify the reason for a mistake	0	0	0	1*	1*	1
6.2.1 Find the best deal using a one- or two-step calculation that meets the stated conditions	0	0	1	0	0	1
6.2.2 Find the best deal from a group and then do something with the answer	0	0	0	1*	0	1*
6.2.3 Determine the better economic value of several alternatives by using graphics, determining the percentage difference, or determining unit cost	0	0	0	0	1*	1*
6.3 Basic statistical concepts	—	—	—	—	—	—
6.3.2 Calculate the weighted mean OR 6.3.3 Interpret measures of central tendency OR 6.3.4 Interpret measures of spread and tolerance	0	0	0	0	1	1
6.3.1 Calculate the average (mean)	0	1	0	0	0	1
6.4 Identify the correct equation	0	0	0	1	1	2

*For Troubleshooting (6.1) and Best Deal (6.2.1), if a form contains a troubleshooting item at Level 6, then it must not have a best deal item at Level 6 but should have one at Level 7; if a form contains a best deal item at Level 6, then it must not have a troubleshooting item at Level 6 but should have one at Level 7.

Table 3.11. Number of Items per Level for Applied Math Applications

Application	Number per Level					
	Level 3	Level 4	Level 5	Level 6	Level 7	Total
Quantity (QUA)	0–4	0–4	0–4	0–4	0–4	4–9
Money (MON)	0–4	0–4	0–4	0–4	0–4	4–9
Time (TIM)	0–4	0–4	0–4	0–4	0–4	4–9
Measurement (MEA)	0–4	0–4	0–4	0–4	0–4	4–9
Total	6	6	7	6	6	31

Each form of the Applied Math assessment is built to conform to the test specifications defined in Tables 3.4 through 3.11. ACT's test development and psychometric staff members thoroughly review each form to ensure that it meets the specifications and is parallel in terms of content to all other Applied Math forms.

3.3.3 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:

- the graphic complexity category of the stimulus
- the graphic skill elicited by the item
- the interaction between the graphic complexity of the stimulus and the graphic skill of the item

The graphic complexity category was defined based on the stimulus's number of variables, data levels, number of axes, graphic type, and total number of graphics (see Table 2.4). ACT content specialists evaluated each stimulus and, based on these characteristics, determined its category.

Graphic literacy skills were divided into four primary skills: locating information; assessing trends, patterns, and relationships; making inferences or decisions; and selecting a graphic to represent information. According to professional literature on workplace graphic literacy and data from ACT's job profiling, 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 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 the frequency of their use in the workplace.

Tables 3.12 through 3.15 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 3.12. Interaction of Graphic Complexity Level With Cognitive Skill Levels With the Overall Graphic Level Definitions

Cognitive Skill Levels	Graphic Complexity			
	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 3.13. 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 3.14. 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 a 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 3.15. Cognitive Skill: Number of Items by Graphic Complexity

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

Each form of the Graphic Literacy assessment is built to conform to the test specifications defined in Tables 3.12 through 3.15. ACT's test development and psychometric staff members thoroughly review each form to ensure that it meets the specifications and is parallel in terms of content to all other Graphic Literacy forms.

3.4 Graphic Literacy—Evidence Based on Response Processes

ACT conducted two cognitive laboratory studies to analyze the cognitive processes that examinees use to solve graphic literacy tasks. In the first study, ACT implemented a think-aloud protocol (Van Someren et al., 1994) to gather data for the purpose of better understanding the construct. In the second study, ACT used eye-tracking software to identify the item features participants focused on as they solved the graphic literacy tasks (Beatty, 1982; Marshall, 2002; Porter et al., 2007). In the second study, participants answered items constituting a complete form of the Graphic Literacy assessment built to the content specifications defined earlier in this chapter. The purpose of both studies was to elicit evidence to support the interpretation and use of Graphic Literacy scores as indicators of ACT's graphic literacy construct.

Findings From the Think-Aloud Study. ACT's cognitive labs used think-aloud protocols in the initial stages of assessment development to gain a greater understanding of graphic literacy. Twenty-one individuals participated in the think-aloud study: 10 high school students, three college students, and eight adults currently in the workforce. The group included 16 White participants, two Black participants, one Latinx participant, and two multiracial participants. The study included 16 women and five men. Six participants were employed full-time, and eight were employed part-time (one adult, three college students, and four high school students). The remaining seven participants were not in the workforce.

Participants were recruited from two high schools, one university, and several places of employment. When participants came into ACT's cognitive lab, a lab assistant provided instruction on testing using the think-aloud protocol. The assistant explained that each participant would be working through a series of problems dealing with graphics. The assistant and the participant would be seated in a small room where the participant would be videotaped while working through the problems. Each participant had to agree to be videotaped during the session. (Participants under the age of 18 had a parent or guardian agree to have the session videotaped.) As the participants worked through the problems, they were asked to vocalize their thoughts—that is, to speak aloud whatever thoughts they were having while they tried to answer the items. The assistant then demonstrated this think-aloud process on a few graphic literacy problems. During testing, the assistant remained silent except when the participant stopped verbalizing. When this occurred, the assistant reminded the participant to keep talking.

ACT staff members reviewed the videotapes and coded the participants' cognitive processes to specific graphic literacy skills. Based on the verbalized cognitive processes and the subsequent coding, the initial set of 40 graphic literacy skills was condensed to the final set of 17. Through the analysis, ACT found that several skills in the original set were item task descriptions and were indistinguishable from one or two other skills based on the cognitive coding.

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 the problem-solving exercise by focusing on finding or extracting the needed information from the graphic. Some participants were capable of taking that information and performing additional cognitive tasks with it. For example, they might find several data points and then analyze them to determine whether a trend was present. A few participants were able to take this information and work beyond the data to predict or justify a decision.

Participants reported feeling greater comfort with a graphic set as they answered multiple items associated with it. In particular, individuals who tended to score well would first make sense of the graphic set before looking at the questions associated with it. This enabled them to answer questions in the set more quickly because they understood the entire graphic. This finding provided further evidence that including sets of questions associated with a single graphic would allow examinees to answer more items without having the assessment become speeded.

Additionally, participants generally agreed with the graphic complexity labels (see Chapter 2); however, participants' responses revealed a particularly strong context effect for low performers. For relatively simple graphics, low performers would become discouraged or give up if the graphic or its labels included scientific subject matter or jargon. From this finding, ACT determined that scientific subject matter and jargon should not be used on any of the lower level graphic resources.

From the findings of the think-aloud study, ACT concluded that participants' verbalized cognitive processes supported (a) the 17 graphic literacy cognitive skills, (b) the proposed three-step cognitive model that partially defines item complexity, and (c) the inclusion of multiple items with each graphic on the assessment as a means of increasing score reliability. Additionally, the findings indicated that the subject matter associated with the graphics could potentially produce construct-irrelevant variance. As a result, ACT determined that at lower levels, graphic content should not include scientific subject matter and jargon.

Findings From the Eye-Tracking Study. After Graphic Literacy items had been developed and pretested, eye-tracking research was conducted on a newly constructed form of the Graphic Literacy assessment. The form was administered in both paper and online modes, with eye-tracking measured using the SMI RED250 and ETG-2 mechanisms, respectively. The data were then analyzed using the BeGaze^{TM} software suite.

Eye-tracking research has been used to gain insight into problem-solving, reasoning, and search strategies (Jacob & Karn, 2003; Mele & Federici, 2012). The purpose of the eye-tracking study was to analyze gaze data collected from examinees to determine how they interacted with the graphics and the questions to solve test items. A secondary purpose of the study was to analyze differences in how high-performing examinees interacted with the graphic items compared to low-performing examinees. Since differences in eye movements have been observed between experts and non-experts in problem-solving (Jarodzka et al., 2010; Obersteiner et al., 2014), ACT proposed that there would be differences between the gaze patterns of individuals with high and low graphic literacy skills.

A total of 38 individuals participated in the eye-tracking study. Twenty participants took the paper version while using the ETG-2 goggles. Eighteen participants took the online version using the RED250 collector. The participants included 16 high school students, four college students, and 18 workforce-age adults. Twenty-two of the participants were women, and 16 were men. Of the participants, 29 were White, six were Black, two were Asian, and one was Latinx. Of the workforce-age adults, 10 were employed full-time, five were employed part-time, and three were not in the workforce. Of the 20 high school and college students, 12 were employed part-time.

Participants were instructed to answer each item as they would if they were taking the test to obtain a usable score. The equipment was calibrated with SMI Experiment Center^{TM} 3.7; for the paper administration, the ETG-2 goggles were used with a 3-point calibration, and for the online administration, the RED250 was used with a 5-point calibration. Both data streams were collected at 60 Hz. Participants were given the Graphic Literacy assessment and either the Applied Math or Workplace Documents assessment. A proctor monitored each participant's progress to make certain that the eye measurements were in frame. All participants were able to finish the assessments in less than the allotted time.

The eye-tracking study collected gaze data as the participants worked through the test form. The gaze data included information on fixations, saccades, sequence, heat maps, and pupillometry. Fixation refers to the amount of time an individual's eyes are focused on a single point. Fixation lengths tend to vary from about 100 to 160 milliseconds. It is during this time that the brain starts to process the visual information received by the eyes. The length of fixation is an indication of information processing or cognitive activity (Matos, 2010). Saccades are extremely fast jumps from one fixation point to another. The average length of a saccade is 20 to 40 milliseconds, and when reading English, an individual's mean saccade size is 7–9 letter spaces. Saccade patterns can also reveal either confusion or understanding (Matos, 2010). Heat maps for a page or screen show the areas individuals looked at, the order in which they looked at them, and how long they spent looking at each area. By using data gathered from heat maps, it is possible to infer the thought patterns used by examinees to arrive at a response. For example, a heat map may indicate that an examinee initially spent time studying the graphic and then proceeded to the question. Heat maps also provide information on the aspects of graphics on which examinees focused. For example, did they focus on the critical information, or were they distracted by other information (Djamasbi, 2014)? Pupillometry is the measure of pupil size and reactivity. Originally, pupillometry provided a metric to assess the cognitive functioning of individuals who had suffered neurological injury. From this research, it was discovered that, as humans become more highly engaged in cognitive activities, their pupils enlarge. As a result, cognitive research now uses pupillometry as a measure of cognitive engagement and effort (Marshall, 2002; Porter et al., 2007).

Eye-tracking data for each item were analyzed using SMI BeGaze^{TM} 3.7 software, and the findings provided support for several components of the graphic literacy construct. Heat map and gaze sequence differences existed between individuals who answered an item correctly and those who did not. For example, one item required examinees to extract information from one graphic in order to locate information in a second graphic. Individuals who answered correctly spent an average of 6.59 seconds (17.6% of their response time) on the second graphic, which contained the needed information. Conversely, individuals who responded incorrectly spent an average of 0.44 seconds (2.7% of their response time) on the second graphic.

Individuals who answered an item correctly had hot zones (areas that indicated a large amount of total gaze time) on the areas that were defined as critical by the item skill description. An analysis of sequence maps and areas of interest (AOI) showed that individuals who answered an item correctly not only looked where predicted based on the graphic literacy skills but also generally followed the skill list path describing the optimal way to solve the task (Thomas & Langenfeld, 2017). Conversely, individuals who did not answer a question correctly tended to either miss hot zones on one or more key pieces of data or have hot zones on irrelevant information. When a second graphic was required, individuals who did not answer correctly often did not look at the second graphic at all. For each item, key differences were observed between the groups who answered an item correctly and those who did not. The differences were qualitatively evident in both the heat maps and in AOI quantitative data such as gaze time, time to first fixation, and returns to critical information.

The data were then analyzed to compare the gaze patterns of individuals who achieved high scores (Levels 6 and 7) to those of individuals who achieved low scores (Level 3 or below). ACT was interested in investigating whether these two groups used qualitatively different approaches to solve graphic literacy tasks. The eye-tracking data supported the position that high performers used different strategies and interacted with the graphics differently than low performers.

The gaze data indicated that each time a new item set was presented, high performers tended to spend a significant initial amount of time working to understand each new graphic. In contrast, low performers tended to spend little to no initial time on the graphic; instead, they tended to spend their initial time on the questions. On subsequent items associated with the same graphic, high performers tended to need less time studying the graphic. ACT concluded that, because of their initial effort to understand the graphic, high performers were able to quickly move on to subsequent questions and focus only on the needed information in the graphic. On some sets, low performers would have hot zones on the introductory contextual information on the follow-up questions. This finding confirmed the finding from the think-aloud study that examinees with high graphic literacy skills invest time to understand a graphic initially, and this understanding facilitates solving problems later.

Low performers struggled to fully use information when multiple graphics were required to answer a question. They tended not to be able to take information from one graphic and apply it to the relevant information in a second graphic. This finding provided support for the graphic complexity rating system, in which a graphic literacy stimulus is considered more complex when it contains multiple graphics. When an item required the use of two or more graphics, low performers tended to have hot zones on only the graphic containing information directly related to the item stem and responses. It appeared that low performers selected responses based on the first available information rather than analyzing the complete graphic to determine the relevant information. High performers tended to have hot zones in the necessary AOI of two or more graphics, demonstrating the ability to analyze the full graphic set and then focus on the relevant information for answering the question.

An additional difference between high and low performers was in the eye-tracking findings for items requiring examinees to analyze trends. High performers tended to have broad hot zones that traced a trend line. On the other hand, low performers tended to have hot zones that jumped from point to point. One of the purposes of converting data from a table (individual data points) to a graph is to illustrate trends (Few, 2012), and it appeared that high performers were looking for these trends while low performers were looking at individual data points. From a skills-description perspective, this finding indicated that low performers were able to extract information (one-step cognitive process) and could do so several times, especially from a single graphic; however, they struggled to identify a trend, which was defined as a two-step cognitive process.

An additional difference between the two groups was found in how they perceived graphics with more than two axes. ACT defined graphics that contain more than two axes as having greater graphic complexity. For example, a graph might contain a left and right *y*-axis, with each axis representing a different dependent variable. High performers tended to read, in their initial gazes, the labels defining both axes, while low performers tended to ignore the labels. Low performers tended to focus on the incorrect axis more frequently than high performers, indicating that they had difficulty understanding that the two *y*-axes represented two separate variables. This difference in being able to discern a third variable in a two-dimensional graphic provided additional evidence that the addition of a third axis to a graphic increases graphic complexity.

The findings from the eye-tracking analysis have implications both for supporting ACT's construct definition and for assisting individuals interested in improving their graphic literacy skills. ACT defined the Graphic Literacy assessment construct as an interaction between a task's graphic complexity and its cognitive complexity. The findings from the eye-tracking study provide support for this definition. In terms of graphic complexity, two features appeared to greatly increase the complexity of a graphic—multiple graphics in a stimulus and two variables represented on the *y*-axis. In terms of cognitive complexity, the findings supported ACT's three-step process model of cognitive complexity.

Additionally, ACT found that high performers tend to start by studying a graphic to gain an understanding of its overall purpose. They are then able to focus on the critical information that is needed for solving the problem. On the other hand, low performers tend to start with the question and then search the graphic for something that appears relevant. This strategy frequently has them focusing on sections of the graphic that are not pertinent to solving the problem. High performers also recognize the importance of the labels that are included on graphics and work to understand their implications for the problem. Low performers tend not to use the information contained within the labels.

The eye-tracking data also captured pupillometry measurements for the online test takers. This information will be analyzed using Workload RT V3 Academic software from EyeTracking Inc., which is based on the Index of Cognitive Activity (Marshall, 2002; Bartels & Marshall, 2012). Analyses are under way to evaluate the interaction of the cognitive skill and graphic complexity to determine overall difficulty using measurements of cognitive load.

Chapter 4: Test Administration

Instructions for administering the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessments are in the *ACT WorkKeys Administration Manual: Paper Testing* and the *ACT WorkKeys Administration Manual: Online Testing*. Staff members of approved testing sites are responsible for securely administering the WorkKeys NCRC assessments.

In addition to the administration manuals, ACT WorkKeys has additional resources available online.[4]

4.1 Policies and Procedures

It is important that all staff involved in administering the WorkKeys NCRC assessments precisely follow the instructions 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 NCRC assessments:

- accessing or obtaining a test booklet or test questions prior to the test for any reason (exception 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 help prepare examinees for testing
- 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

[4] ACT WorkKeys provides test administrators with multiple support materials through the ACT website: http://www.act.org/content/act/en/products-and-services/workforce-solutions/act-workkeys/administer.html#techspecs

- 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.

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 NCRC tests within the next 12 months
- anyone involved in ACT WorkKeys NCRC test preparation activities at any time during the current testing year (September 1 through August 31) because of potential conflict of interest[5]

In addition, if any relative or ward of yours will test at your site or any school in the state during the testing window, please follow these guidelines:

- 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 or after testing.
- You **may** serve as a room supervisor or proctor, provided that the examinee who is your relative or ward is not assigned to test in a room where you are working.

[5] 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.

- 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 the examinee will be canceled 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 school.

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 administration manual appropriate for the test location (both manuals if necessary) and ensuring compliance with the policies and procedures • viewing and participating in training provided by ACT • evaluating and approving requests for ACT WorkKeys NCRC accommodations • ordering paper materials for standard time tests • ordering paper alternate testing formats for examinees needing accommodations • receiving, checking in, and securely storing test materials • arranging for the application of bar-coded labels on the answer documents by testing staff if required by your contract • 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, 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 the test responses of examinees who have been approved for alternate response modes to be transferred to their answer document and for examinees who have been approved locally to mark their answers directly on their test booklet • serving as room supervisor as needed

4.2.2 Backup Test Coordinator

The test coordinator should have a qualified backup test coordinator available if the test coordinator becomes ill or is otherwise unable to be present on test day. Backups are encouraged to assist test coordinators prior to, during, and after testing.

Backup test coordinators are also expected to participate in training conducted by ACT (if previously untrained by ACT) prior to the test date. Backups can serve at only one school.

4.2.3 Test Accommodations Coordinator (Optional)

The test coordinator may appoint a test accommodations coordinator. The test accommodations coordinator is responsible for the following tasks:

- assisting with the test coordinator’s responsibilities as needed
- reading the appropriate administration manual (both manuals if necessary) and ensuring compliance with the policies and procedures
- evaluating and approving requests for ACT WorkKeys NCRC accommodations
- notifying the test coordinator of any examinees needing alternate format test materials from ACT
- viewing and participating in accommodations training provided by ACT

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 the room supervisor if only one room is used.

The room supervisor’s specific responsibilities include the following:

- reading the appropriate administration manual (both manuals if necessary) and ensuring compliance with the policies and procedures
- 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 identification (ID) or personally recognizing and admitting examinees
- marking attendance on the roster and indicating whether IDs were checked
- 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
- using two timepieces to properly time the tests and recording the start, five-minutes-remaining, and stop times listed in the relevant administration manual
- completing all information on the appropriate administration forms found in the *ACT WorkKeys Administration Manual: Paper Testing* or the *ACT WorkKeys Administration Manual: Online Testing*
- being attentive to examinees and attending to 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 assist a room supervisor or the test coordinator if fewer than 30 examinees are testing (20 examinees for accommodations/supports examinees). As test rooms increase in size, proctors are required to assist the room supervisor or test coordinator.

The proctor's responsibilities include the following:

- reading the appropriate administration manual (both manuals if necessary) and ensuring compliance with the policies and procedures
- attending both the training and briefing sessions conducted locally by the test coordinator
- helping admit examinees and marking attendance and whether IDs were checked 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 attending to materials at all times
- walking around the test room during testing to replace defective materials, ensure all examinees are working on the correct test, and 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 detect and discourage 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 may arise as well as concerns staff members may have.

4.3.2 Administration Manuals

ACT provides administration manuals that communicate our expectations for paper testing and online testing, and every staff member is expected to read the manual appropriate for the test location (both manuals if necessary). The manuals are proprietary information copyrighted by ACT. They are to be used only for the purpose of administering the ACT WorkKeys NCRC assessments and are not to be copied or shared for any other purpose.

Each testing staff member is to be given complete copies of the administration manuals 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 in the following list. If these requirements are not met, scores may be canceled.

- **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; 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 reach every seat during testing without disturbing examinees.
- Seat the 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.

Chapter 5: Accessibility

Over the last decade of educational research and practice, ACT has come to understand that all examinees need and use certain tools every day to engage in the classroom and to communicate effectively what they have learned and can do. There are distinct levels of support that examinees may need in order to demonstrate what they know and can do on academic tests. ACT makes several levels of support available for the three assessments aligned to the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®). 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 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 intended; and
- remove unnecessary barriers to the content knowledge measured on the WorkKeys NCRC assessments.

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

Accommodations for the WorkKeys NCRC assessments are determined and documented by personnel responsible for accommodations coordination at the school level. In most cases, a current Individualized Education Program (IEP) or Section 504 plan prepared by appropriate academic or psychological staff for an examinee will be acceptable documentation to support the use of accommodations on the WorkKeys NCRC assessments. The use of accommodations results in WorkKeys NCRC assessment scores that are fully reportable. Information on reportable accommodation support can be found in the WorkKeys NCRC assessment administration manuals and in the tables below. Currently, certain accommodations are available only with paper testing.

5.1 ACT WorkKeys NCRC Assessment Support System

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 WorkKeys NCRC assessment support continuum recognizes that the need for personalized support is not restricted to any one group of examinees. It encompasses the needs of the entire testing population, including those with disabilities, those who are English learners, and those who have no diagnostic label at all. All these individuals have a shared need to be able to communicate fairly and effectively what they know and can do when they take a test.

The WorkKeys NCRC assessment accessibility supports are structured along a continuum of increasingly intensive supports designed to meet the needs of all learner populations. Three levels of accessibility support are offered: 1) embedded universal supports, 2) designated supports, and 3) accommodation supports.

5.2 Test Administration and Accessibility Levels of Support

As part of ACT's commitment to providing a fair testing experience for all examinees, the WorkKeys NCRC assessments provide an integrated system of accessibility supports that include both accommodations and less intensive accessibility supports. At times, 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: Examinees using the paper format sometimes also take advantage of certain online options. Regardless of test format, all examinees who use designated supports or accommodations must have this use documented by appropriate school (or test site) personnel. For this reason, we have provided a general description of WorkKeys NCRC assessment accessibility supports here in one chapter. Full procedural requirements and instructions for using permitted supports during test administration are provided in the *ACT WorkKeys Accessibility Supports Guide*.

The WorkKeys NCRC assessments permit the use of only those accessibility supports that validly preserve the skills and knowledge that the assessment claims to measure while removing construct-irrelevant barriers to examinee performance. The three levels of support in the WorkKeys NCRC assessment accessibility system represent a continuum of supports, from least intensive to most intensive, with the assumption that all users have communication needs that fall somewhere on this continuum. This support continuum results in every examinee having a personalized performance opportunity.

5.2.1 Support Level 1: Universal Support

The first level of support is called the universal support. These supports meet the common, routine accessibility needs of the most typical test takers. All examinees are provided with these tools as appropriate, even examinees who have no documented support plan. System tools include, but are not limited to, the following examples:

- Magnifier tool (online or provided by examinee)
- Browser zoom magnification (online)
- Answer eliminator (online or provided by examinee)
- Test directions available on demand (online or printable)
- Highlighter (online or provided by examinee)
- Keyboard navigation (online)

- Scratch paper (online or provided by examinee)
- Mark item for review (online or provided by examinee)

These tools are either embedded in the basic computer test delivery platform or provided locally as needed. No advance request is needed for universal supports.

5.2.2 Support Level 2: Designated Supports

Designated supports are available to all users but must be identified in advance, planned for, and then either selected from the menu inside the test to be activated (online) or provided locally.

Many examinees' unique sensory and communication 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, designated supports provide just the tools needed by an individual user, allowing true personalization of the test experience.

Designated supports are slightly more intensive than universal supports but can be delivered in a fully standardized manner that is valid, appropriate, and tailored to the specific needs of an individual examinee. Some of these supports require the use of tool-specific administration procedures. In the WorkKeys NCRC assessments, designated supports include, but are not limited to, the following examples:

- Color contrast
- Line reader
- Translated verbal instructions (locally provided)
- Signed Exact English (SEE), American Sign Language (ASL), or Cued Speech for directions and verbal instructions (locally provided)
- Answer masking
- Dictate responses
- Respond in test booklet or on locally provided separate papers
- Audio indicator of time remaining
- Individual administration
- Special seating and grouping

Designated supports 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 become comfortable using these types of tools both individually and in combination with other tools.

5.2.3 Support Level 3: Accommodations

Accommodations are high-level accessibility tools needed by relatively few examinees. The accommodation must be identified in advance, planned for, and either selected from the menu inside the test to activate it (online) or provided locally. Accommodations may require advance ordering of specialized 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 level of support have a formally documented need and have therefore been identified as qualifying for resources or specialized supports. In order for such supports to be selected, administered, and used effectively and securely, the examinee must have expertise or special training in their use or be closely monitored. These supports can include, but are not limited to, the following examples:

- Braille EBAE (contracted, includes tactile graphics)
- Braille UEB with Nemeth (contracted, includes tactile graphics)
- Braille UEB Math/Science (contracted, includes tactile graphics)
- Cued Speech: test items
- Text-to-speech
- English prerecorded audio
- English audio reader’s script
- Signed Exact English (SEE): test items
- Screen reader software compatible (JAWS or NVDA)
- Abacus
- Extra time

Decisions about accommodations are typically made by an educational team on behalf of and including the examinee. Accommodation decisions are based on a formal, documented evaluation of specialized need and require the examinee to be familiar enough with the tools to use them fluidly and effectively during the test experience. For accommodations to be delivered successfully and securely, additional local resources or specialized knowledge is often required.

Accommodations are available to users who have been qualified to use them by the local governing school or employment authority (e.g., a school district, a work training agency, an employer, or a branch of the 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. The WorkKeys NCRC assessments require examinees who use accommodations to have a formally documented need, as well as relevant knowledge of and familiarity with the required tools.

Accommodations must be requested through the local test site according to WorkKeys NCRC assessment procedures, as defined in the administration manual. Before testing, appropriate documentation of accommodations as specified in the manual must be provided to the administering agency by either the examinee or a local governing educational authority.

5.3 Valid Test Scores and Equal Benefit for All Examinees

ACT aims to ensure that all examinees benefit equally from the WorkKeys NCRC assessments. Assessments administered with accommodations and other accessibility supports under standardized conditions result in valid and fully reportable scores. The use of any accessibility supports that are not specified by ACT or not properly administered violates what the test is designed to measure and results in a score that is invalid and incompatible with the stated purposes of the assessment.

For a full listing of accessibility support options for the WorkKeys NCRC assessments, please see Table 1 in the *Accessibility Supports Guide for ACT WorkKeys Testing* linked below. This table includes presentation supports, interaction and navigation supports, response supports, and general test conditions.

https://www.act.org/content/dam/act/unsecured/documents/WorkKeysAccessibilitySupportsGuide.pdf

Chapter 6: Test and Information Security

6.1 Test Security

To ensure the validity of scores on the assessments leading to the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®), all examinees, individuals who 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 its Test Security Principles and test security requirements. Those principles and 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 examinees and testing staff.

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

Testing staff must protect the confidentiality of the WorkKeys NCRC 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 the WorkKeys NCRC assessments.

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, misconduct (including copying answers or using an answer key), questionable or prohibited test-taking behavior (such as using prohibited electronic devices during testing or communicating 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, and test proctor or test administrator misconduct (such as providing answers or questions to examinees or permitting them to engage in prohibited conduct during testing).

In addition to establishing these security-related administration protocols, ACT engages in additional test security practices designed to protect the WorkKeys NCRC assessments and the validity of scores. These practices include (a) using a reporting hotline through which individuals with information about misconduct on a WorkKeys NCRC assessment can anonymously report such information to ACT, (b) using data forensics in support of WorkKeys NCRC–related investigations, and (c) monitoring the web to detect testing misconduct, possible unauthorized disclosure of secure WorkKeys NCRC test content, and any other activity that might compromise the security of the WorkKeys NCRC assessments or the validity of scores.

6.2 Information Security

ACT's information security program framework is based on the widely recognized International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 27001:2013 standard and National Institute of Standards and Technology (NIST) Computer Security Resource Center (CSRC) best practices. This framework was selected because it covers a range of information security categories that comprehensively match the broad perspective that ACT takes in safeguarding its information assets. These are the categories covered by the framework and brief statements of their importance to ACT:

1. Information Security Program Management: This is overseen by the information security officer at ACT. The information security officer and the information security management steering team have the 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/IEC 27001:2013 standard 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/IEC 27005:2018 standard. The identification, management, and mitigation of information security risks are managed using the information security management system (ISMS) guidelines defined in the ISO/IEC 27005:2018 standard. ACT also follows the NIST Special Publication (NIST SP) 800-37.

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

    a. 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 or media with appropriate access controls.

        - Only a limited number of authorized users may have access to this information.

        - Physical records must be locked in drawers or cabinets while not being used.

    b. To form a system of control to protect examinee data, ACT also has standards in place related to access management, business continuity, clear desk/clear screen policies, end user storage, external authentication, information security incident management, malware protection, mobile device use, network security management, payment card security, secure application development, secure system configuration, security event logging and monitoring, system vulnerability and patch management, and web content.

4. Information and Technology Compliance: The systems that store, maintain, and process information are designed to protect data security through all life cycle 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 ensure 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 grant access only 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 virtual private network (VPN) remote access are 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 that is presented to all ACT employees and reinforced with training. ACT is held accountable for 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, including monthly vulnerability scanning, to identify vulnerabilities on networks, servers, and desktops. ACT has always maintained a compliant status in accordance with the Payment Card Industry Data Security Standard (PCI DSS) version 3.2.1 requirements. In addition to performing scans for PCI DSS compliance, ACT has a suite of vulnerability scanning tools that are coordinated with a log management and event monitoring tool to provide reporting and alerting.

10. Boundary Defense: ACT uses 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 demilitarized zone (DMZ) systems and networks that provide alerts and procedures for review and response. Procedures include security review and approval of changes to configurations and semiannual 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 used to ensure that a secure environment is maintained at the level of the end user device. This includes segmentation within the ACT network, antivirus programs, and programs to prevent data loss. 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 that 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 policies outline response procedures, notification protocols, and escalation procedures. Forensic investigations are performed at the direction of the information security officer. In the event of a declared incident, ACT maintains a subscription service with a third party specializing in computer forensics.

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 that is 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 ISIRP tasks an ACT information security incident response team (ISIRT) with providing an around-the-clock (24/7) coordinated security incident response throughout ACT. Information security management has the responsibility and authority to manage the ISIRT and implement necessary ISIRP actions and decisions during an incident.

Chapter 7: Reporting

7.1 WorkKeys Reports

ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) reports are designed to provide detailed information to examinees, testing staff, 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 testing staff. This information is available through the portal whether an examinee takes an assessment online or on paper.

The WorkKeys NCRC reports fulfill these objectives:

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

The WorkKeys NCRC assessments are criterion-referenced tests. Unlike norm-referenced tests, criterion-referenced tests interpret scores on the basis of the skills demonstrated through testing. The WorkKeys NCRC 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 WorkKeys NCRC PLDs.)

The Individual Examinee Score Report summarizes examinee performance. A separate Individual Examinee Score Report is generated for each WorkKeys NCRC assessment that a person takes. This report 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 the examinee’s score means, a section that includes the PLD for the specified level score
- how the examinee can use their 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 to either examinees or institutions. Table 7.1 lists the available WorkKeys NCRC reports.

Table 7.1. Various WorkKeys NCRC Reports and Their Functions

Report	Function
Data Export Report	• Information about the examinee including demographic information, date tested, test titles, and scores received • Source codes: WKIV for online tests or WKPP for paper tests • Online test status codes: complete (C), incomplete (IC), and inactive (IA) *Note.* Report does not include User IDs and passwords.
Individual Score Report (by Group)	• Examinee’s score for one test • Information on how to interpret the score
Individual Score Report (by Examinee)	• Examinee’s score for one test • Information on how to interpret the score
Individual Summary Score Report	• Examinee’s scores • Information on how to interpret the scores
Summary Score Report	• Examinee’s scores • No information on how to interpret the scores

Report	Function
Roster Score Report	List of examinees, the tests they took, test dates, and the scores they received, according to the selected date range *Note.* Because a single score cannot be shown for Personal Skills assessments, the report only indicates whether the examinee has taken the test.
Individual Score vs. Profile Report	Scores an examinee achieved compared to the scores that are required for a job that has been profiled for WorkKeys NCRC score levels
Group vs. Profile Report	Scores that a group of examinees achieved compared to scores that are required for a job that has been profiled for WorkKeys NCRC score levels
Registered to Test Report	List of all examinees who are registered for online tests but have not yet taken the tests
Test Usage Report	Count of the tests taken at the selected site for a given date range
Local Scan Instant Score Report	Examinee’s score and score interpretation, one per page, for one test taken (if specified) or for all tests taken (if not specified) *Note.* Similar to Individual Score Report (by Examinee)
Invoice Report Paper-Based Testing	• Candidate and test details for only the paper-based tests on the invoice • Contains a row for each candidate’s test along with the additional details of that test
Certificate Data Export	WorkKeys NCRC details along with the WorkKeys NCRC test scores that qualify the candidate for the specific certificate level for the selected agency and/or testing location
Test Data Export	WorkKeys NCRC test scores, including a row for each test taken at the selected testing location for the selected time period
Certificates by Certificate ID	WorkKeys NCRC, including the certificate level achieved and the WorkKeys NCRC test scores that qualify the candidate for the specific certificate level for a given candidate ID
Certificates by Testing Location	WorkKeys NCRC for the selected testing location and time period, including the certificate level achieved and the WorkKeys NCRC test scores that qualify the candidate for the specific certificate level
Certificate Count by Testing Location	Counts of WorkKeys NCRCs for the selected testing location and time period, including the counts by certificate type and level
Qualification Letter by Certificate ID	ACT WorkKeys Qualification Letter, including the WorkKeys NCRC level achieved and the log-in information for the candidate
Qualification Letter by Testing Location	ACT WorkKeys Qualification Letters for the selected testing location and time period, including the certificate level achieved and the WorkKeys NCRC test scores that qualify the candidate for the specific certificate level

Chapters 8–11 of the technical manual describe in detail the WorkKeys NCRC assessments’ scores, metrics, and interpretations.

Chapter 8: Scores and Score Scales

8.1 Overview

This chapter describes the rationales, procedures, and outcomes related to scoring the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessment items, establishing scale scores, and defining level scores for the assessments.

Raw and scale scores are two types of scores used to facilitate score interpretation and use. The *Standards for Educational and Psychological Testing* define 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. 95). Raw scores are frequently transformed to scale scores to facilitate and standardize score interpretations. Producing scale scores for a new assessment requires a scaling analysis, which 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. 95). For the WorkKeys NCRC assessments, an item response theory (IRT) approach with arcsine transformation was applied to produce a scale with nearly equal conditional standard error of measurement (CSEM) for most score points.

Any WorkKeys foundational skill assessment, including Applied Math, Graphic Literacy, and Workplace Documents, places 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 WorkKeys NCRC assessment skills demonstrated by the examinee. For this alignment to be achieved, cut scores are established on the reported scale score to support level score interpretations. A cut score is “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. 100). For the WorkKeys NCRC assessments, 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 WorkKeys NCRC assessments are selected-response items (i.e., multiple-choice items). Selected-response items require examinees to select a correct answer from a set of alternative choices. For the WorkKeys NCRC assessments, each Workplace Documents and Applied Math selected-response item has five choices, and each Graphic Literacy item has four. Each item that an examinee answers correctly provides the examinee with a score value of one raw point. Incorrect responses, missing responses (items that an examinee did not answer), or multiple responses yield a value of zero points. The examinee's raw score is calculated by summing the correct responses.

ACT strives to write each WorkKeys NCRC 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, which includes item writing, editing, reviewing, and pretesting. After these steps have been followed, an item may be selected for inclusion on a WorkKeys NCRC form. ACT psychometricians and content specialists regularly conduct preliminary item analyses 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 certain skill at a specified level defined by the WorkKeys NCRC assessment construct. WorkKeys NCRC skills associated with each of the five levels (Levels 3 to 7) were defined through the design process described in Chapter 2. Each WorkKeys NCRC form is composed of items that assess the skills defined by each level, and each form is built to the test specifications described in Chapter 3. When examinees complete a WorkKeys NCRC assessment, they receive a report that includes their scale and level scores. The scale and level scores serve two distinct purposes in facilitating score interpretations and uses.

Scale scores provide finer-grained score distinctions than level scores, and they are designed to assist in (a) analyzing growth or improvement over time, (b) evaluating group comparisons on outcome measures, and (c) providing evidence of the benefits derived 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 removes the need for ACT to report the SEM at the different score points.

Level scores provide examinees with information about 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 were 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

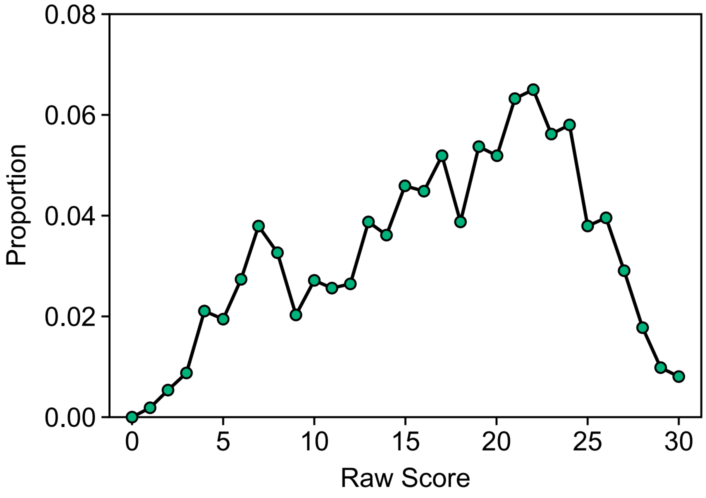
A scaling study was conducted in spring 2017 as part of a series of field test studies to establish the score scale for the updated WorkKeys NCRC assessments. ACT recruited examinees from various regions in the United States to participate in the field test studies. The sampling plan was designed to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

8.4.1 Workplace Documents

After data cleaning, the scaling study included a sample of 1,136 examinees. Forty sites participated in the scaling study. These sites included 13 high schools and 27 adult testing centers across 22 states. Examinees were 44% male and 53% female. Approximately 6% of the examinees were Latinx, 17% were Black, and 61% were White. ACT concluded that the sample was representative of the current WorkKeys NCRC test-taking population.

The examinees in the scaling study took the Workplace Documents assessment form W2C_S1. 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 in previous administrations of the Workplace Documents assessment.

Figure 8.1. Raw Score Distribution for the WD Scaling Study Form W2C_S1



Note. The mean and standard deviation are 17.24 and 6.82, respectively.

Figure 8.2 illustrates the item *p*-values (ranging from 0.2 to 0.9) and *b*-parameter estimates by corresponding levels for form W2C_S1, where each purple dot represents the average item *p*-value or *b*-parameter estimate for that level. As expected, the item *p*-values tend to decrease as the item difficulty increases. The plot of *b*-parameter estimates shows a similar trend (i.e., the average *b*-parameter values increase 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 W2C_S1

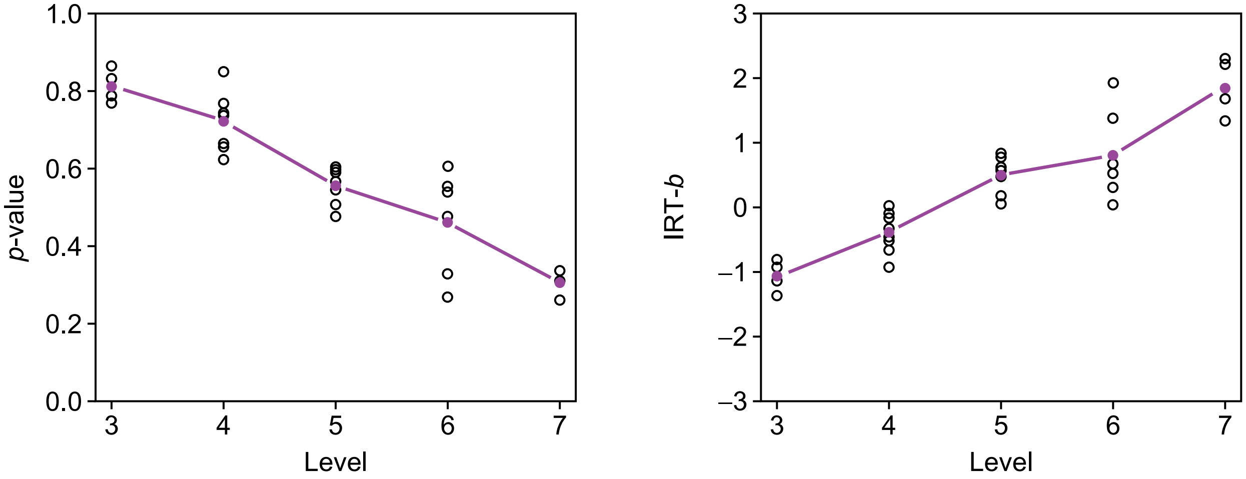
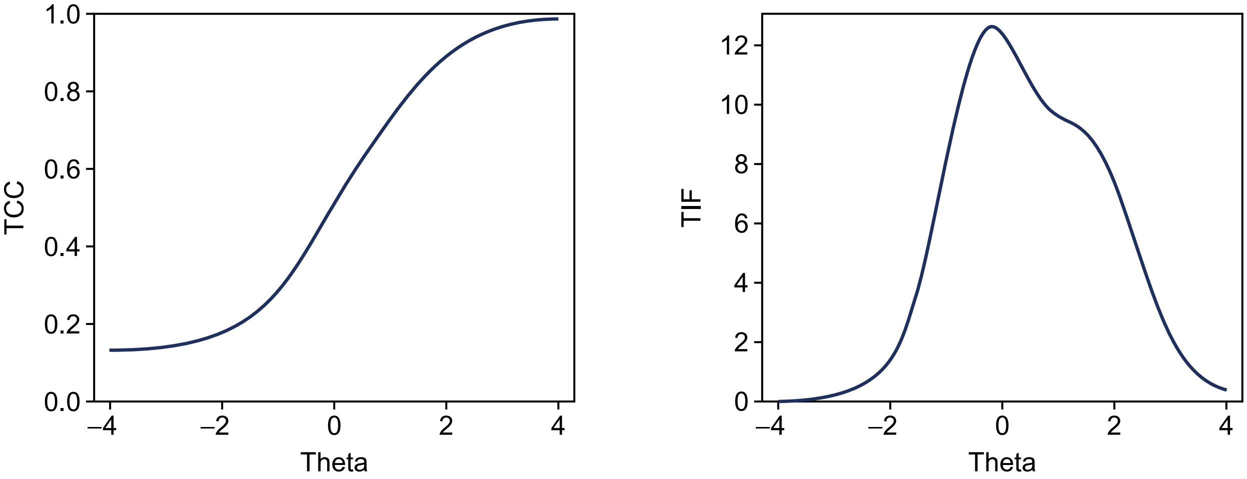


Figure 8.3. Test Characteristic Curve (Left) and Test Information Function (Right)



To be consistent with the Reading for Information assessment (currently called Workplace Documents) and the other NCRC assessments, the average scale score was set to about 78, and the scale score CSEM was set to less than 2. In addition, the scale score range was defined as 65 to 90, which is identical to the NCRC 1.0 assessment scale score. In order for scaling to be conducted, the target scale score mean and the target scale score SEM were required. IRT was used to derive the raw-to-scale score conversion (Ban & Lee, 2007), and the arcsine transformation (Kolen, 1988; Kolen & Brennan, 2014) was used to equalize the conditional standard error of measurement (CSEM) along the score scale. The following five steps were implemented to derive the raw-to-scale score conversion:

1. Item parameters were calibrated based on the 3-PL IRT model.
2. Theta estimate (ability estimate) for each examinee was calculated using the item scoring vector data and the item parameter estimates calibrated in Step 1.
3. The expected raw score distribution was estimated based on the item parameter estimates from Step 1 and theta estimates from Step 2, 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 4 were linearly transformed to the scale scores using the target scale score mean and target scale score SEM. The slope of the linear transformation is $A = \frac{\sigma(E_S)}{\sigma(E_g)}$ and its intercept is $B = \mu(S) - \frac{\sigma(E_S)}{\sigma(E_g)} \times \mu[c(X)]$, where $\mu(S)$ and $\sigma(E_S)$ are the targeted mean and SEM of the scale scores, respectively, and $\mu[c(X)]$ and $\sigma(E_g)$ are the mean and SEM of the *g*-scores, respectively.

To transform raw scores to scale scores using the previous five steps, ACT applied the following requirements:

- 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, there were few scale scores with a first decimal place of 0.5.
- CSEM was as similar as possible across the score scale.

The target scale score mean and target scale score SEM were specified to be 77.3 and 1.7, respectively. These values were obtained through several explorations using the data from the scaling study and the requirements previously defined.

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

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

- As shown in Figure 8.4, the scale score CSEMs are flat below 2.0 along the scale scores, except for the 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 (77.4) is very close to the target scale score mean (77.3) used as the input for the arcsine transformation. Table 8.1 presents a summary of the unrounded scale scores (USS) and rounded scale scores (RSS) for form W2C_S1. 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)

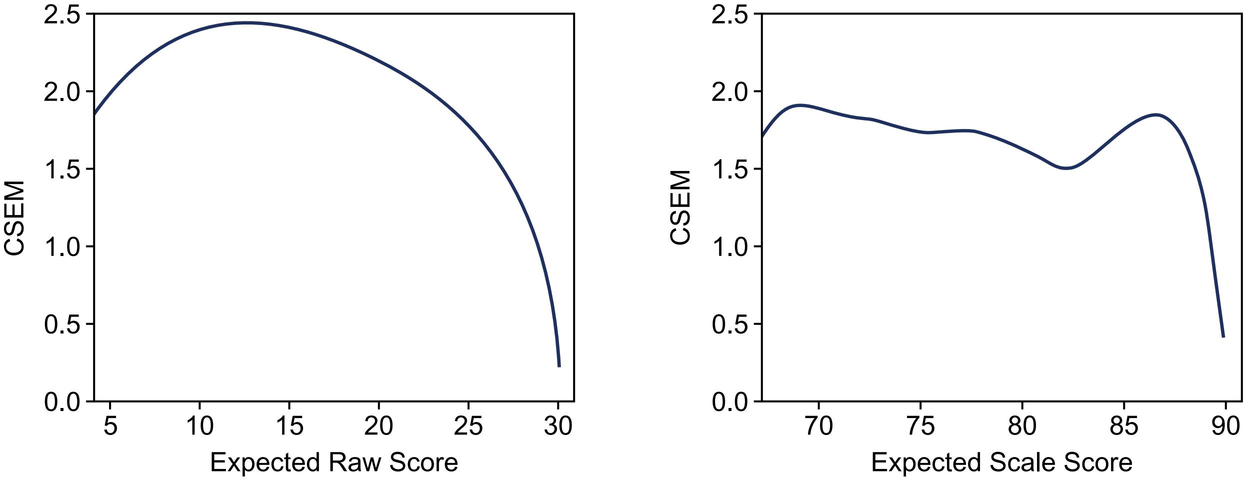
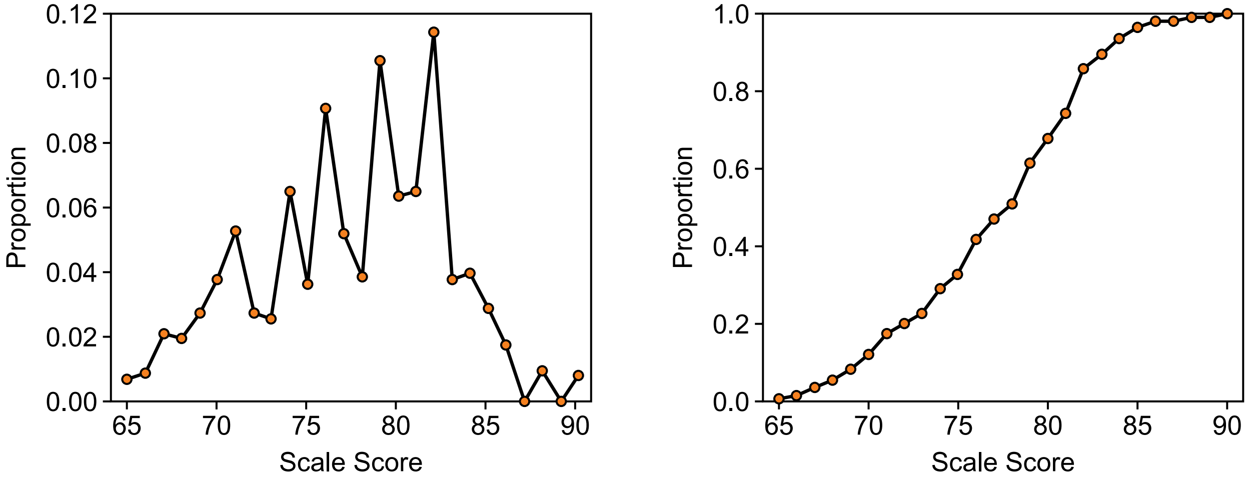


Table 8.1. Summary of Unrounded and Rounded Scale Scores—Workplace Documents

Scale score	Mean	SD	Min	10th	25th	50th	75th	90th	95th	Max
USS	77.42	5.21	63.59	69.91	73.63	77.80	81.52	84.17	85.20	90.58
RSS	77.41	5.20	65	70	74	78	82	84	85	90

Figure 8.5. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right)

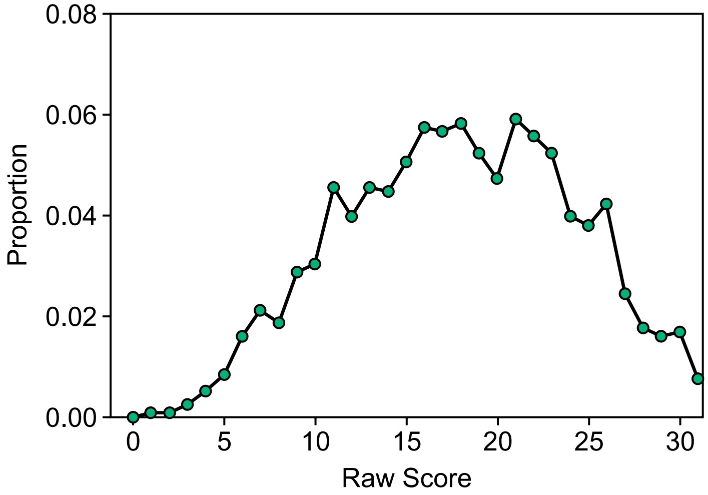


8.4.2 Applied Math

After data cleaning, the scaling study included a sample of 1,185 examinees. Forty sites participated in the scaling study. These sites included 13 high schools and 27 adult testing centers across 22 states. Examinees were 46% male and 51% female. Approximately 7% of the examinees were Latinx, 17% were Black, and 60% were White. ACT concluded that the sample was representative of the current WorkKeys NCRC test-taking population.

The examinees in the scaling study took the Applied Math assessment form M2C_S1. ACT analyzed examinee data from the scaling study, applying a three-parameter logistic (3-PL) IRT model to calibrate item parameters. Figure 8.6 presents the raw score distribution from the sample. The distribution appears to be slightly left-skewed, which is consistent with distributions observed in previous administrations of the Applied Math assessment.

Figure 8.6. Raw Score Distribution for the AM Scaling Study Form M2C_S1



Note. The mean and standard deviation are 17.88 and 6.25, respectively.

Figure 8.7 illustrates the item *p*-values (ranging from 0.2 to 0.95) and *b*-parameter estimates by corresponding levels for form M2C_S1, where each purple dot represents the average item *p*-value or *b*-parameter estimate for that level. As expected, the item *p*-values tend to decrease as the item difficulty increases. The plot of *b*-parameter estimates shows a similar trend (i.e., the average *b*-parameter values increase as the level increases). Figure 8.8 shows the TCC and TIF for the scaling study form.

Figure 8.7. Item *p*-Values and *b*-Parameter Estimates by Item Levels for Form M2C_S1

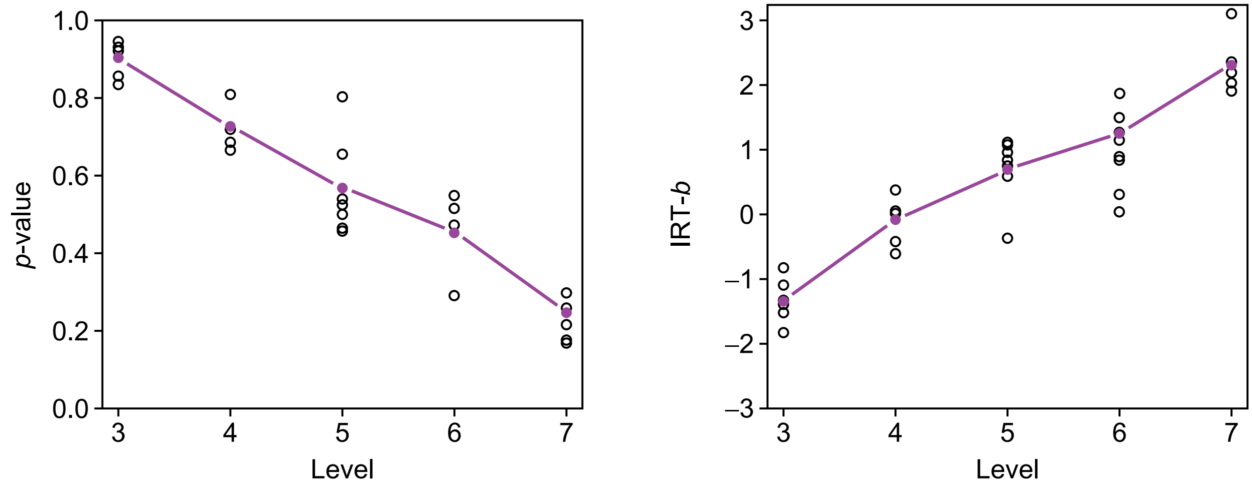
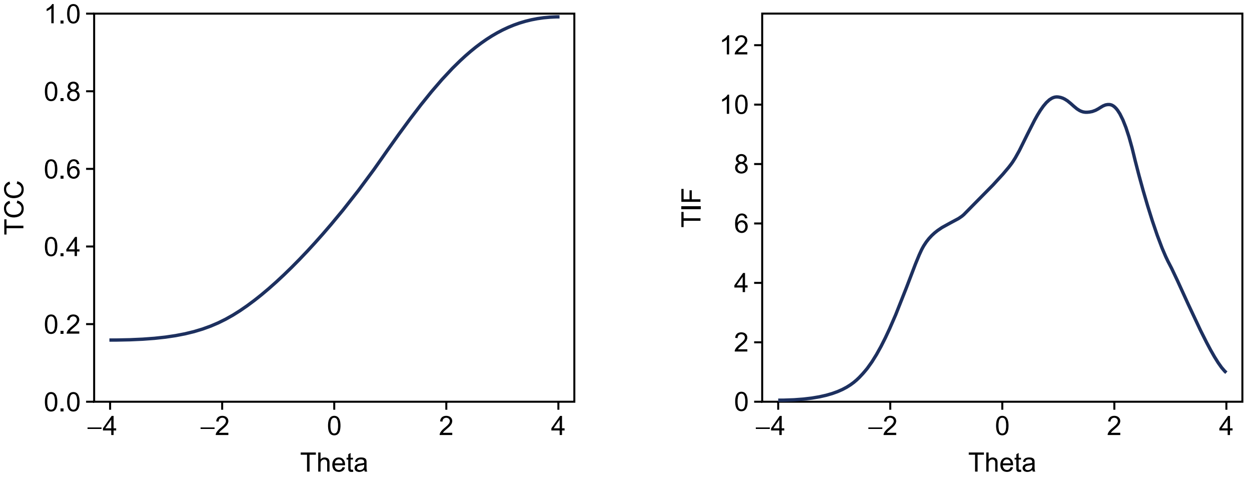


Figure 8.8. Test Characteristic Curve (Left) and Test Information Function (Right)



To be consistent with the Applied Mathematics (currently called Applied Math) assessment and the other NCRC assessments, the average scale score was set to about 78, and the scale score CSEM was set to less than 2. In addition, the scale score range was defined as 65 to 90, which is identical to the NCRC 1.0 assessment scale score. In order for scaling to be conducted, the target scale score mean and the target scale score SEM were required. IRT was used to derive the raw-to-scale score conversion (Ban & Lee, 2007), and the arcsine transformation (Kolen, 1988; Kolen & Brennan, 2014) was used to equalize the CSEM along the score scales. The following five steps were implemented to derive the raw-to-scale score conversion:

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

2. Theta estimate (ability estimate) for each examinee was calculated using the item scoring vector data and the item parameter estimates calibrated in Step 1.

3. The expected raw score distribution was estimated based on the item parameter estimates from Step 1 and theta estimates from Step 2, 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 4 were linearly transformed to the scale scores using the target scale score mean and target scale score SEM. The slope of the linear transformation is $A = \frac{\sigma(E_S)}{\sigma(E_g)}$ and its intercept is $B = \mu(S) - \frac{\sigma(E_S)}{\sigma(E_g)} \times \mu[c(X)]$, where $\mu(S)$ and $\sigma(E_S)$ are the targeted mean and SEM of the scale scores, respectively, and $\mu[c(X)]$ and $\sigma(E_g)$ are the mean and SEM of the *g*-scores, respectively.

To transform raw scores to scale scores using the previous five steps, ACT applied the following requirements:

- 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, there were few scale scores with a first decimal place of 0.5.

- 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 1.6, respectively. These values were obtained through several explorations using the data from the scaling study and the requirements previously defined.

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

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

- As shown in Figure 8.9, the scale score CSEMs are flat below 2.0 along the scale scores, except for the 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 8.2 presents a summary of the USS and RSS for form M2C_S1. Figure 8.10 illustrates the relative and cumulative frequency distributions of the scale scores.

Figure 8.9. CSEM for Raw Scores (Left) and Scale Scores (Right)

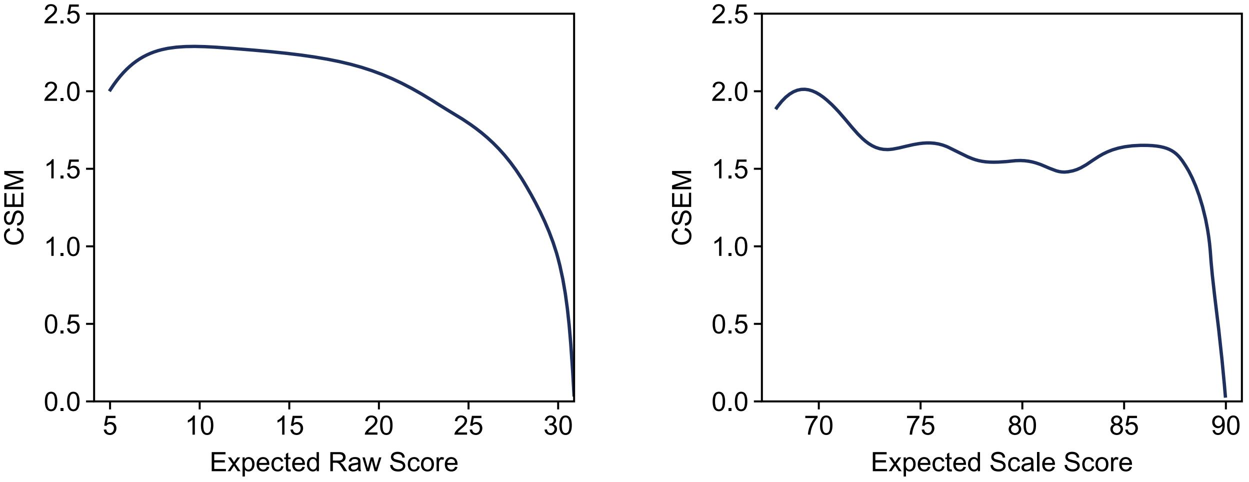
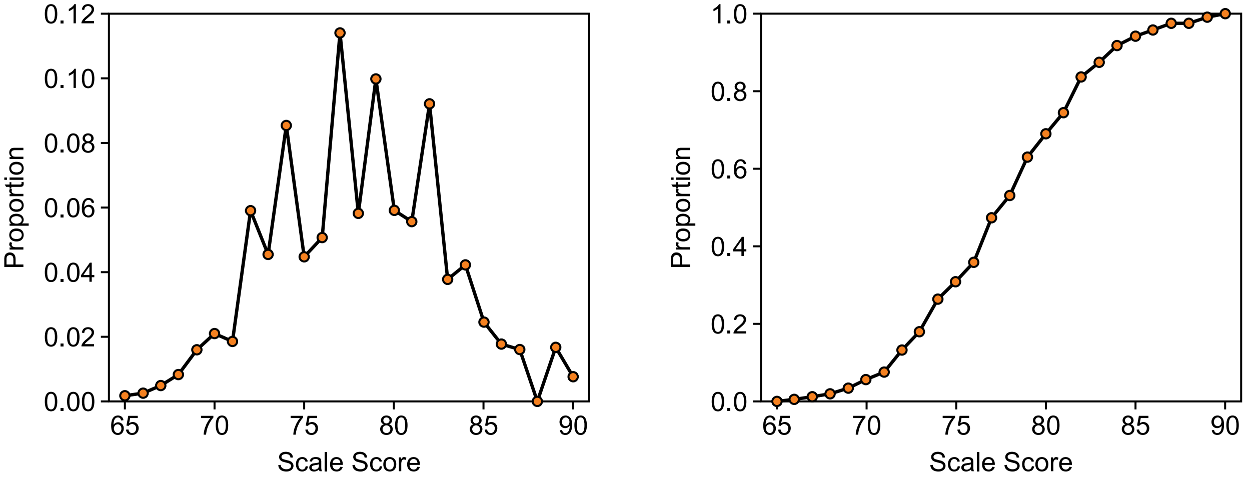


Table 8.2. Summary of Unrounded and Rounded Scale Scores—Applied Math

Scale score	Mean	SD	Min	10th	25th	50th	75th	90th	95th	Max
USS	78.03	4.77	63.65	71.60	74.47	77.93	81.57	84.08	86.08	91.53
RSS	78.02	4.75	65	72	74	78	82	84	86	90

Figure 8.10. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right)

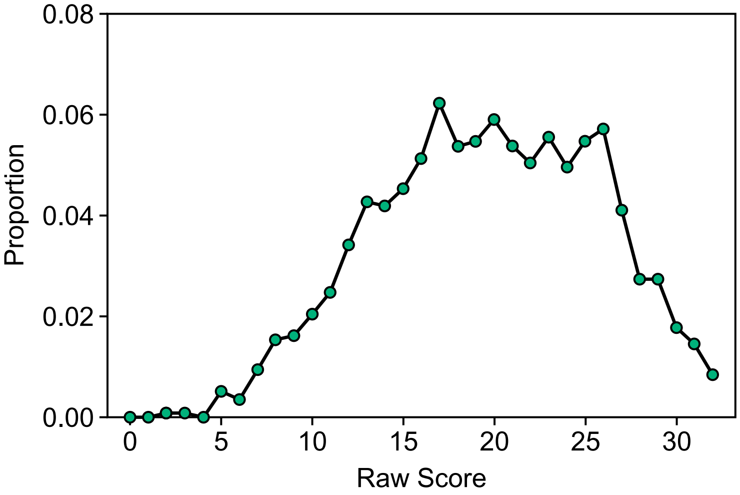


8.4.3 Graphic Literacy

After data cleaning, the scaling study included a sample of 1,170 examinees. Forty sites participated in the scaling study. These sites included 13 high schools and 27 adult testing centers across 22 states. Examinees were 46% male and 51% female. Approximately 7% of the examinees were Latinx, 18% were Black, and 61% were White. ACT concluded that the sample was representative of the current WorkKeys NCRC test-taking population.

The examinees in the scaling study took the Graphic Literacy assessment form G2C_S1. ACT analyzed examinee data from the scaling study, applying a three-parameter logistic (3-PL) IRT model to calibrate item parameters. Figure 8.11 presents the raw score distribution from the sample. The distribution appears to be slightly left-skewed, which is consistent with distributions observed in previous administrations of the Graphic Literacy assessment.

Figure 8.11. Raw Score Distribution for the GL Scaling Study Form G2C_S1



Note. The mean and standard deviation are 19.66 and 6.03, respectively.

Figure 8.12 illustrates the item *p*-values (ranging from 0.2 to 0.9) and *b*-parameter estimates by corresponding levels for this form, where each purple dot represents the average item *p*-value or *b*-parameter estimate for that level. As expected, the item *p*-values tend to decrease as the item difficulty increases. The plot of *b*-parameter estimates shows a similar trend (i.e., the average *b*-parameter values increase as the level increases). Figure 8.13 shows the TCC and TIF for the scaling study form.

Figure 8.12. Item *p*-Values and *b*-Parameter Estimates by Item Levels for Form G2C_S1

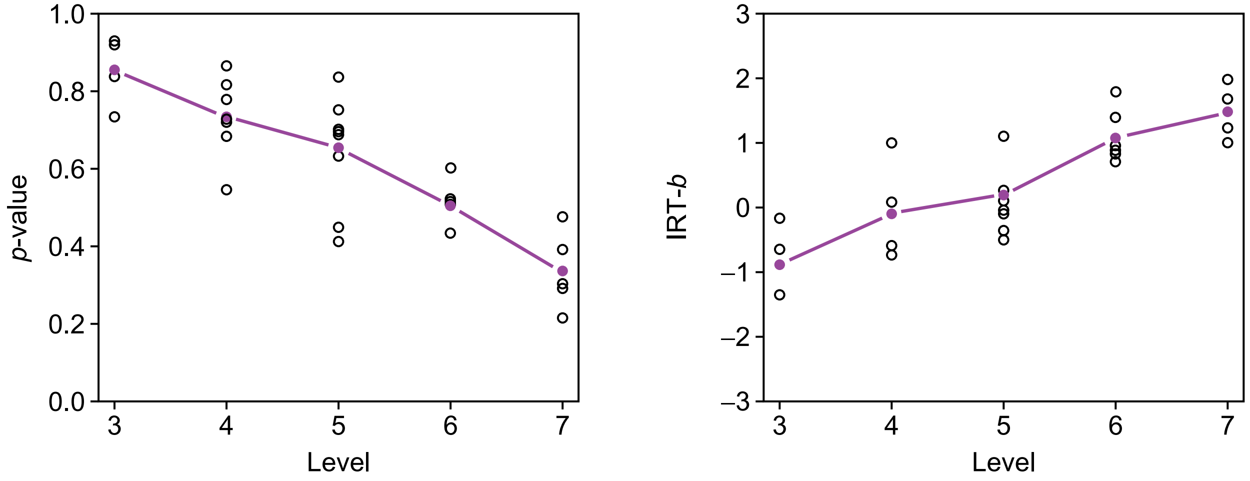
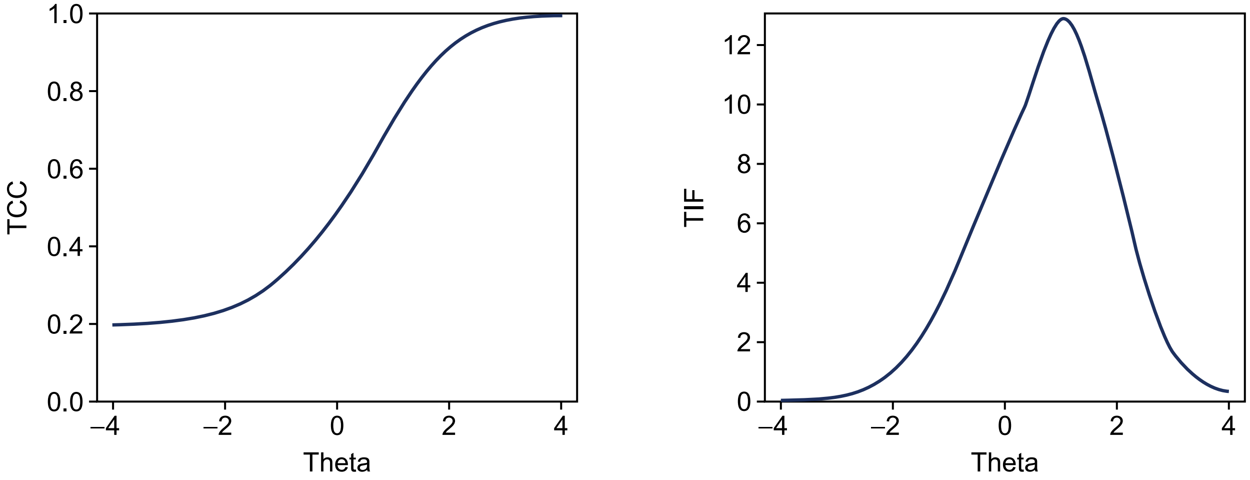


Figure 8.13. Test Characteristic Curve (Left) and Test Information Function (Right)



To be consistent with the Graphic Literacy assessment and the other NCRC assessments, the average scale score was set to about 78, and the scale score CSEM was set to less than 2. In addition, the scale score range was defined as 65 to 90, which is identical to the NCRC 1.0 assessment scale score. In order for scaling to be conducted, the target scale score mean and the target scale score SEM were required. IRT was used to derive the raw-to-scale score conversion (Ban & Lee, 2007), 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 to derive the raw-to-scale score conversion:

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

2. Theta estimate (ability estimate) for each examinee was calculated using the item scoring vector data and the item parameter estimates calibrated in Step 1.

3. The expected raw score distribution was estimated based on the item parameter estimates from Step 1 and theta estimates from Step 2, 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 4 were linearly transformed to the scale scores using the target scale score mean and target scale score SEM. The slope of the linear transformation is $A = \frac{\sigma(E_S)}{\sigma(E_g)}$ and its intercept is $B = \mu(S) - \frac{\sigma(E_S)}{\sigma(E_g)} \times \mu[c(X)]$, where $\mu(S)$ and $\sigma(E_S)$ are the targeted mean and SEM of the scale scores, respectively, and $\mu[c(X)]$ and $\sigma(E_g)$ are the mean and SEM of the *g*-scores, respectively.

To transform raw scores to scale scores using the previous five steps, ACT applied the following requirements:

- 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, there were few scale scores with a first decimal place of 0.5.

- 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, respectively. These values were obtained through several explorations using the data from the scaling study and the requirements previously defined.

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

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

- As shown in Figure 8.14, the scale score CSEMs are flat below 2.0 along the scale scores, except for the 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 8.3 presents a summary of the USS and RSS for form G2C_S1. Figure 8.15 illustrates the relative and cumulative frequency distributions of the scale scores.

Figure 8.14. CSEM for Raw Scores (Left) and Scale Scores (Right)

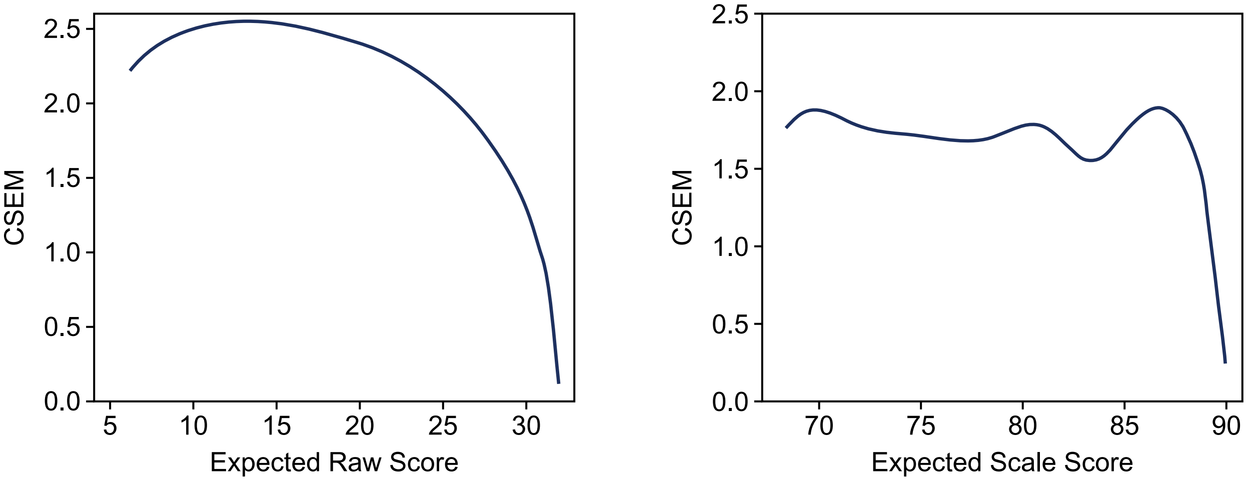
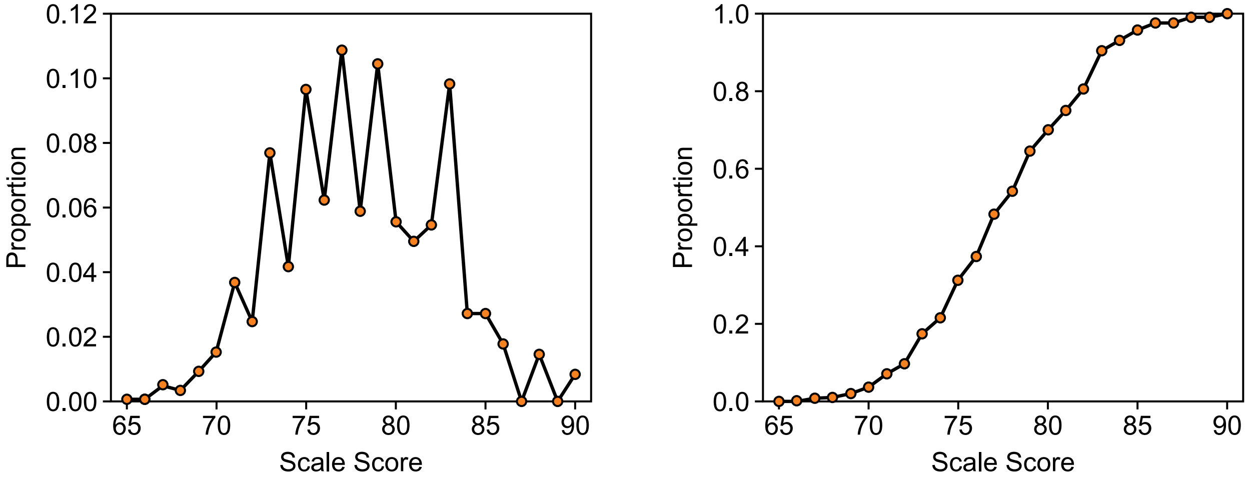


Table 8.3. Summary of Unrounded and Rounded Scale Scores—Graphic Literacy

Scale score	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 8.15. Relative Frequency Distribution (Left) and Cumulative Frequency Distribution (Right)



8.5 Procedures for Establishing the Level Scores

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

The Mapmark standard-setting method with Whole Booklet Feedback is a variation of the popular Bookmark procedure (Lewis et al., 2012). Both methods employ an ordered item booklet (OIB), which contains a sample of items from the item pool ordered from easiest to hardest. The key difference between Bookmark and Mapmark is that the Mapmark OIB includes an item map, which maps the difficulty of each item onto the actual scale value. The item map, therefore, shows how much more difficult one item is than another, thereby providing additional information on item difficulty.

ACT conducted a standard-setting study for each of the three WorkKeys NCRC assessments—Workplace Documents, Applied Math, and Graphic Literacy—with a special 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 in the three assessments. Because the three assessments are criterion-referenced measures, the reported scores on each assessment are aligned to the PLDs (see Chapter 2) that an examinee has demonstrated by 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 cut point have demonstrated the ability to perform the skills corresponding to that skill level, and examinees who score below the cut 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 the following are true:

- Examinees are expected to correctly respond to at least 67% of the items that belong to their reported levels.
- Examinees are expected to have demonstrated mastery for all levels below their reported levels.
- Examinees are NOT expected to correctly respond to more than 67% of the items that belong to levels higher than their reported levels.

The Mapmark standard setting included a three-round process with Whole Booklet Feedback. For each of the three rounds, the SMEs set cut scores for each level. In Round 1, the SMEs (a) took the relevant assessment (i.e., Workplace Documents, Applied Math, or Graphic Literacy), (b) reviewed the relevant assessment's PLDs, (c) reviewed test items and their associated scale scores, (d) linked the 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: (a) those items that the panelists felt were easy enough for an examinee who was minimally qualified in the skill level to have mastered and (b) those items that were too difficult for a minimally qualified examinee to have mastered. In this context, mastery was defined as having a two-in-three 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 (Levels 3–7).

In Round 2, the panelists received feedback regarding their bookmark placement relative to the recommended scale scores on the item map scale and to the group's median cut score. The group was then provided with the Whole Booklet Feedback. Specifically, the panelists were provided with data showing how 16 examinees (two test takers in each level, one test taker between each pair of adjacent levels, one test taker above the highest level, and one test taker below the lowest level) answered each of the items on each scaling form: Workplace Documents form W2C_S1, Applied Math form M2C_S1, and Graphic Literacy form G2C_S1. For each skill level, data were provided for two examinees that scored at or near the Round 1 cut score and for one borderline examinee. The purpose was to help the panelists understand what examinees at the Round 1 cut scores can do and help the panelists consider whether this is what examinees should be able to do according to the PLD for each skill level. Using all of this information, the 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 the 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 because the assessment is criterion referenced. With that, the panelists 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, the panelists 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 the impact data on the basis of the final cut scores. After these discussions, the panelists approved the final median cut score for each of the five performance levels.

8.5.1 Workplace Documents

A total of 83 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 (i.e., the item difficulty was converted to a scale score) to form the OIB.

ACT reviewed the work of the standard-setting panelists and evaluated whether it achieved the desired result of a criterion-referenced assessment with the level scores aligned to the PLDs. After reviewing the panelists' work and recommendations, ACT approved the cut scores for the five levels for the Workplace Documents assessment. The final median cut scores presented in Table 8.4 are used to define each performance level in the Workplace Documents assessment.

Table 8.4. Median Cut Scores for the Workplace Documents Assessment

Final Scale Score Cut Points			
		Range of Median Cut	
Level	Median Cut	Minimum	Maximum
3	72	71	74
4	77	77	77
5	81	80	81
6	83	82	84
7	86	85	89

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

8.5.2 Applied Math

A total of 77 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 (i.e., the item difficulty was converted to a scale score) to form the OIB.

ACT reviewed the work of the standard-setting panelists and evaluated whether it achieved the desired result of a criterion-referenced assessment with the level scores aligned to the PLDs. After reviewing the panelists' work and recommendations, ACT approved the cut scores for the five levels for the Applied Math assessment. The final median cut scores presented in Table 8.5 are used to define each performance level on the Applied Math assessment.

Table 8.5. Median Cut Scores for the Applied Math Assessment

Final Scale Score Cut Points			
		Range of Median Cut	
Level	Median Cut	Minimum	Maximum
3	72	72	72
4	76	74	78
5	80	78	81
6	83	83	84
7	86	86	89

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

8.5.3 Graphic Literacy

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 (i.e., the item difficulty was converted to a scale score) to form the OIB.

ACT reviewed the work of the standard-setting panelists and evaluated whether it achieved the desired result of a criterion-referenced assessment with level scores aligned to the PLDs. After reviewing the panelists' work and recommendations, ACT approved the cut scores for the five levels for the Graphic Literacy assessment. The final median cut scores provided in Table 8.6 are used to define each performance level on the Graphic Literacy assessment.

Table 8.6. Median Cut Scores for the Graphic Literacy Assessment

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

With established scale scores and cut scores, new forms will be built to be parallel to other forms based on the test specifications (see Chapter 3) and will be equated to the base form to achieve score comparability. As a result, the 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 for each ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessment. The first section describes the equating methods used for the ACT WorkKeys NCRC assessments. Because multiple alternate forms of the assessments are required, ACT applies equating methods to ensure that scores from different forms are interchangeable and comparable. The second section reports the findings of the following field test studies: testing time, dimensionality, model fit, and mode comparability. ACT administers the WorkKeys NCRC assessments both on paper and online. The mode comparability study was conducted to learn whether scores earned by examinees using the paper mode are interchangeable with and comparable to scores earned by examinees using the online mode. The third section presents the findings of a linking study that was conducted to provide concordance scale scores for the current assessment. To obtain concordance scale scores of the current assessment and help WorkKeys NCRC examinees understand the relationship between scores earned on previous assessments and scores earned on current assessments, ACT linked the scale scores of the current assessment to the scale scores of the previous assessment. Although scores earned on current assessments are not interchangeable with scores earned on previous assessments, the linking study will help users understand the relationship between the updated assessment and the earlier assessment.

9.1 Equating Method and Procedures

New WorkKeys NCRC test forms are developed regularly to ensure the fairness and security of the test scores. Though each form is constructed to meet the same statistical and content specifications (see Chapter 3 for the detailed content blueprint), the forms may differ slightly in difficulty. Equating is the process of making statistical adjustments to achieve score interchangeability across 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, ACT either pre-equates or post-equates the WorkKeys NCRC forms to produce scale scores and level scores. Pre-equating is the process by which raw scores are converted to scale scores prior to test delivery. Pre-equating enables examinees to receive their score reports shortly after testing. When a new WorkKeys NCRC form is being constructed, items that meet the content classification specifications and the item statistical specifications are selected from an item pool. Test development content specialists and research psychometric specialists review the proposed form to ensure that the form 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 additional details about skill levels 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 test administrations if enough examinees have taken the assessment.

Pre-equating cannot be applied to a newly developed form unless all the items on the form have IRT-calibrated parameter estimates that have been placed on the same scale. For WorkKeys NCRC assessments, ACT is continually developing new items. When new 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 (Hambleton & Swaminathan, 1985). 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 deemed acceptable are added to the WorkKeys NCRC item pool, which is continually expanded and maintained.

In addition, ACT periodically reviews the item pool to archive 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 Field Test Studies

ACT engaged in a series of three field test studies to evaluate the psychometric properties of the initial Workplace Documents, Applied Math, and Graphic Literacy forms. The field testing was designed to (a) determine an acceptable time allotment for testing, (b) develop a standardized score scale that is interpretable and that could be applied to the development of subsequent WorkKeys NCRC forms, (c) evaluate model-data fit for the three-parameter logistic (3-PL) IRT model, and (d) evaluate the mode effect on test scores (paper vs. online administration).

For each of the field test studies, ACT attempted to recruit samples that were representative of the WorkKeys NCRC test population. In recruiting for the field test studies, ACT was cognizant of the need to recruit a sufficient number of adult examinees, because adults make up the major population in the workforce. For Workplace Documents, Applied Math, and Graphic Literacy, Tables 9.1, 9.2, and 9.3 (respectively) compare the percentages of examinees from the WorkKeys NCRC test population (2013–2014) to the examinee percentages from the three field test samples.

Table 9.1. WorkKeys NCRC Test Population Compared to Field Test Samples of Workplace Documents by Age, Gender, and Ethnicity

Group	Group Subcategory	WorkKeys NCRC Test Population	Field Test #1 Sample	Field Test #2 Sample	Field Test #3 Sample
Age	High school age*	40.6%	66.6%	59.4%	45.9%
	Adult	59.4%	33.4%	40.6%	54.1%
Gender	Female	46.0%	48.6%	54.2%	56.4%
	Male	54.0%	48.2%	45.8%	43.6%
Ethnicity	White	58.0%	71.8%	60.8%	66.0%
	Black	21.2%	16.2%	17.3%	15.2%
	Latinx	8.2%	3.0%	6.3%	7.7%

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

*Based on examinees who reported their age as 20 or below

Table 9.2. WorkKeys NCRC Test Population Compared to Field Test Samples of Applied Math by Age, Gender, and Ethnicity

Group	Group Subcategory	WorkKeys NCRC Test Population	Field Test #1 Sample	Field Test #2 Sample	Field Test #3 Sample
Age	High school age*	40.6%	67.1%	60.5%	47.1%
	Adult	59.4%	32.9%	39.5%	52.9%
Gender	Female	46.0%	49.0%	52.6%	56.0%
	Male	54.0%	48.0%	47.4%	44.0%
Ethnicity	White	58.0%	71.8%	60.7%	63.4%
	Black	21.2%	16.4%	17.4%	16.4%
	Latinx	8.2%	3.7%	6.7%	7.9%

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

*Based on examinees who reported their age as 20 or below

Table 9.3. WorkKeys NCRC Test Population Compared to Field Test Samples of Graphic Literacy by Age, Gender, and Ethnicity

Group	Group Subcategory	WorkKeys NCRC Test Population	Field Test #1 Sample	Field Test #2 Sample	Field Test #3 Sample
Age	High school age*	40.6%	60.7%	60.9%	43.6%
	Adult	59.4%	39.3%	39.1%	56.4%
Gender	Female	46.0%	52.6%	53.0%	54.7%
	Male	54.0%	47.4%	47.0%	45.3%
Ethnicity	White	58.0%	60.7%	61.0%	63.4%
	Black	21.2%	17.4%	17.7%	16.9%
	Latinx	8.2%	6.7%	6.6%	8.1%

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

*Based on examinees who reported their age as 20 or below

9.2.1 Testing Time

ACT conducted two separate studies to assess the appropriate amount of time that examinees should be given to complete each WorkKeys NCRC assessment. In the first study, examinees were assigned either the online or the paper version of the assessment. They were also assigned either 55 or 60 minutes to test. Based on the study, ACT wanted to determine (a) whether the test modes (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 enough time to answer all items or tasks, so that the speed at 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 each WorkKeys NCRC assessment, the examinees' speed should not affect their scores, and 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 because 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. 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 test study.

From the first field test study, ACT found that examinees took approximately the same amount of time to complete the assessment regardless of mode (online vs. paper). ACT 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 less than one additional minute for testing than examinees in the 55-minute condition.

9.2.1.1 Workplace Documents

Ninety-five percent of the examinees in both time conditions (60-minute vs. 55-minute) completed the assessment in 46 minutes or less. The omit rate for the final test item in both conditions was less than 1%. For examinees in the 55-minute condition, 98% 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 the online and paper administrations, the allotted testing time should be 55 minutes. In the second field test study, ACT continued to evaluate testing time. The findings from the second study confirmed the conclusion of the first study: 55 minutes is a sufficient amount of testing time for examinees and thus speededness should not affect their Workplace Documents scores.

9.2.1.2 Applied Math

Ninety-five percent of the examinees in both time conditions (60-minute vs. 55-minute) completed the assessment in 51 minutes or less. The omit rate for the final test item in both conditions was less than 2%. For examinees in the 55-minute condition, 93% either strongly agreed or agreed with the statement that they had sufficient time to test. For examinees in the 60-minute condition, 97% either strongly agreed or agreed with the statement that they had sufficient time to test.

Based on these results, ACT concluded that for both the online and paper administrations, the allotted testing time should be 55 minutes. In the second field test study, ACT continued to evaluate testing time. The findings from the second study confirmed the conclusion of the first study: 55 minutes is a sufficient amount of testing time for examines and thus speededness should not affect their Applied Math scores.

9.2.1.3 Graphic Literacy

Ninety-five percent of the examinees in both time conditions (60-minute vs. 55-minute) completed the assessment in 47 minutes or less. The omit rate for the final test item in both conditions was less than 1%. 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 the online and paper administrations, the allotted testing time should be 55 minutes. In the second field test study, ACT continued to evaluate testing time. The findings from the second study confirmed the conclusion of the first study: 55 minutes is a sufficient amount of testing time for examinees and thus speededness should not affect their Graphic Literacy scores.

9.2.2 Scale Scores

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

9.2.3 Score Reliability and Generalizability

Score reliability and generalizability are 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 the assessment. Chapter 10 summarizes the analyses of the field test data that were conducted to provide estimates of score reliability and measurement error. According to the analyses, WorkKeys NCRC assessment scores are reliable and generalizable (i.e., measurement error is minimal) for use in estimating examinee skill levels.

9.2.4 Mode Comparability

ACT developed the WorkKeys NCRC assessments to be administered both on paper and online. The *Standards for Educational and Psychological Testing* state that evidence supporting score interpretations and use should be provided when a testing program maintains test forms “administered under different test administration conditions [that] are comparable for the intended purpose” (see Standard 5.17) (AERA et al., 2014, p. 106).

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

ACT conducted a field test study to evaluate the comparability of scores derived from paper and online administrations. In the field study, test centers randomly assigned each examinee to one of three testing conditions: current online form, current paper form, and previous online form. These three testing conditions are explained in detail in the subsections below.

ACT directed the centers to have each examinee take all three WorkKeys NCRC assessments on the same day or different days, with the test order varied across the sites. The examinees also completed a survey regarding their testing experience; surveys were administered either at the end of each online assessment or after the examinee had finished all three paper assessments.

9.2.4.1 Mode Comparability: Study Design and Sample

As was the case in the scaling study presented in Chapter 8, ACT recruited a sample of examinees representative of the WorkKeys NCRC examinee population.

Although ACT had instructed test centers to randomly assign examinees to the three testing conditions (see Sections 9.2.4.1.1, 9.2.4.1.2, and 9.2.4.1.3), 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 across the three conditions was extremely unbalanced (distribution was considered unbalanced when the difference between the number of examinees assigned to a specific testing condition and the number assigned to the other testing conditions was more than 10). Following the 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 (10 high schools and 27 adult testing centers across 20 states from different regions). Because the data may contain additional sampling errors, measurement precision may be affected. As a result, conclusions drawn from the results should be made with caution.

9.2.4.1.1 Workplace Documents

ACT directed the proctors to randomly assign each examinee to take one of the three test forms: a Workplace Documents online form (W2C_LM1), a Workplace Documents paper form (W2P_LM2), or a Reading for Information online form (W1C_LM3). Examinees who responded to the items on forms W2C_LM1 and W2P_LM2 were used to evaluate mode comparability, and examinees who responded to the items on forms W2C_LM1 and W1C_LM3 were used for the linking study.

Final examinee counts were 662 for online (form W2C_LM1) and 669 for paper (form W2P_LM2). Table 9.4 presents the demographic distribution information. In general, the recruited samples for the two modes acceptably represented the current WorkKeys NCRC test population, and the samples were quite similar except for White groups (63% for online and 57% for paper testing).

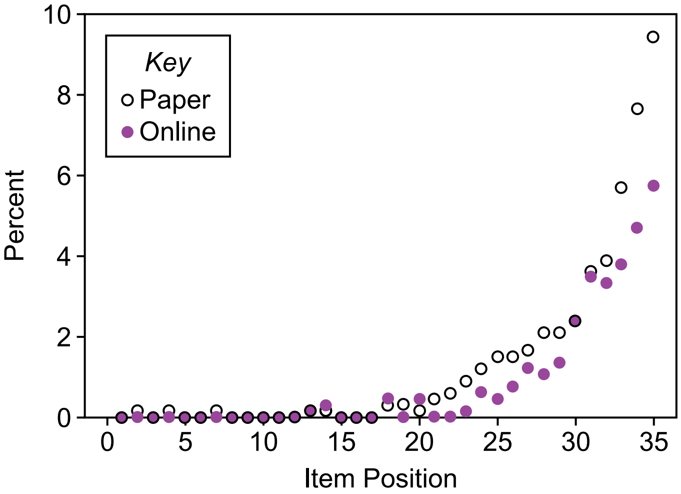
Table 9.4. Sample Demographic Information for the Two Delivery Modes of Workplace Documents

			Gender		Age		Ethnicity		
Mode	***N***	**M (SD)**	**F**	**M**	**HS**	**AD**	**W**	**B**	**L**
Online	662	19.37 (7.07)	55%	44%	45%	55%	63%	15%	10%
Paper	669	19.38 (6.87)	51%	45%	46%	54%	57%	15%	9%

Note. Non-respondent and multiracial examinees not included. F = female; M = male; HS = high school; AD = adult; W = White; B = Black; L = Latinx

For each item, the omit rates (no answer) for the online form and the paper form were compared. As shown in Figure 9.1, the omit rates were below 10% for both modes. The omit rates tended to be slightly higher for the paper form than for the online form.

Figure 9.1. Comparison of Item Omit Rates for the Two Delivery Modes of Workplace Documents



9.2.4.1.2 Applied Math

ACT directed the proctors to randomly assign each examinee to take one of the three test forms: an Applied Math online form (M2C_LM1), an Applied Math paper form (M2P_LM2), or an Applied Mathematics online form (M1C_LM3). Examinees who responded to the items on forms M2C_LM1 and M2P_LM2 were used to evaluate mode comparability, and examinees who responded to the items on forms M2C_LM1 and M1C_LM3 were used for the linking study.

Final examinee counts were 688 for online (form M2C_LM1) and 667 for paper (form M2P_LM2). Table 9.5 presents the demographic distribution information. In general, the recruited samples for the two modes acceptably represented the current WorkKeys NCRC test population, and the recruited samples were quite similar except for the White groups (64% for online and 57% for paper testing).

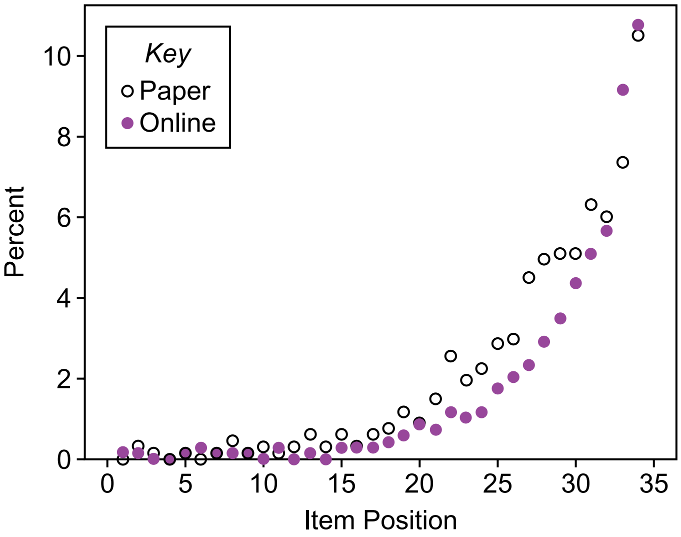
Table 9.5. Sample Demographic Information for the Two Delivery Modes of Applied Math

			Gender		Age		Ethnicity		
Mode	*N*	M (SD)	F	M	HS	AD	W	B	L
Online	688	19.37 (7.07)	54%	45%	46%	54%	64%	15%	9%
Paper	667	19.38 (6.87)	52%	44%	46%	54%	57%	15%	9%

Note. Non-respondent and multiracial examinees not included. F = female; M = male; HS = high school; AD = adult; W = White; B = Black; L = Latinx

For each item, the omit rates (no answer) for the online form and the paper form were compared. As shown in Figure 9.2, the omit rates were below 10% for both modes except for the last item. The omit rates tended to be slightly higher for the paper form than for the online form.

Figure 9.2. Comparison of Item Omit Rates for the Two Delivery Modes of Applied Math



9.2.4.1.3 Graphic Literacy

ACT directed the proctors to randomly assign each examinee 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 who responded to the items on forms G2C_LM1 and G2P_LM2 were used to evaluate mode comparability, and examinees who responded to the items on forms G2C_LM1 and G1C_LM3 were used for the linking study.

Final examinee counts were 701 for online (form G2C_LM1) and 668 for paper (form G2P_LM2). Table 9.6 presents the demographic distribution information. In general, the recruited samples for the two modes acceptably represented the current WorkKeys NCRC test population, and the recruited samples were quite similar except for the White groups (63% for online and 57% for paper testing).

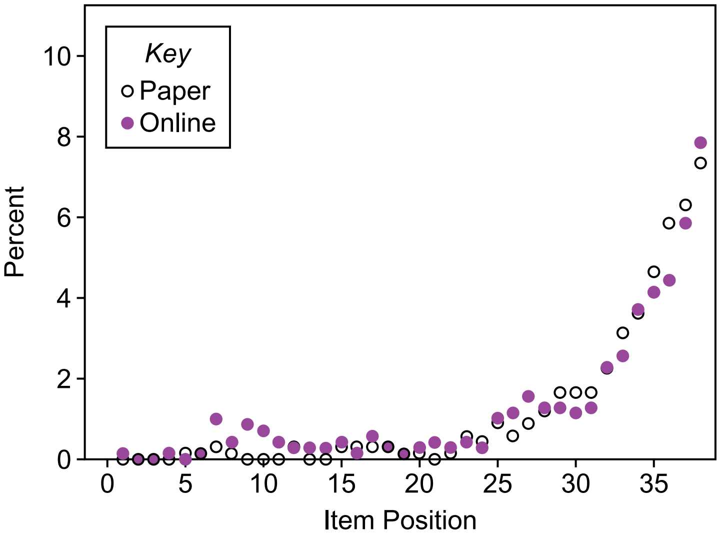
Table 9.6. Sample Demographic Information for the Two Delivery Modes of Graphic Literacy

Mode	*N*	M (SD)	Gender		Age		Ethnicity		
			F	M	HS	AD	W	B	L
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 and multiracial examinees not included. F = female; M = male; HS = high school; AD = adult; W = White; B = Black; L = Latinx

For each item, the omit rates (no answer) for the online form and the paper form were compared. As shown in Figure 9.3, the omit rates were below 10% for both modes. The omit rates for the two modes tended to be similar.

Figure 9.3. Comparison of Item Omit Rates for the Two Delivery Modes of Graphic Literacy



9.2.4.2 Mode Comparability: Comparisons of Items, Tests, and Score Conversions

9.2.4.2.1 Workplace Documents

Item-Level Comparison

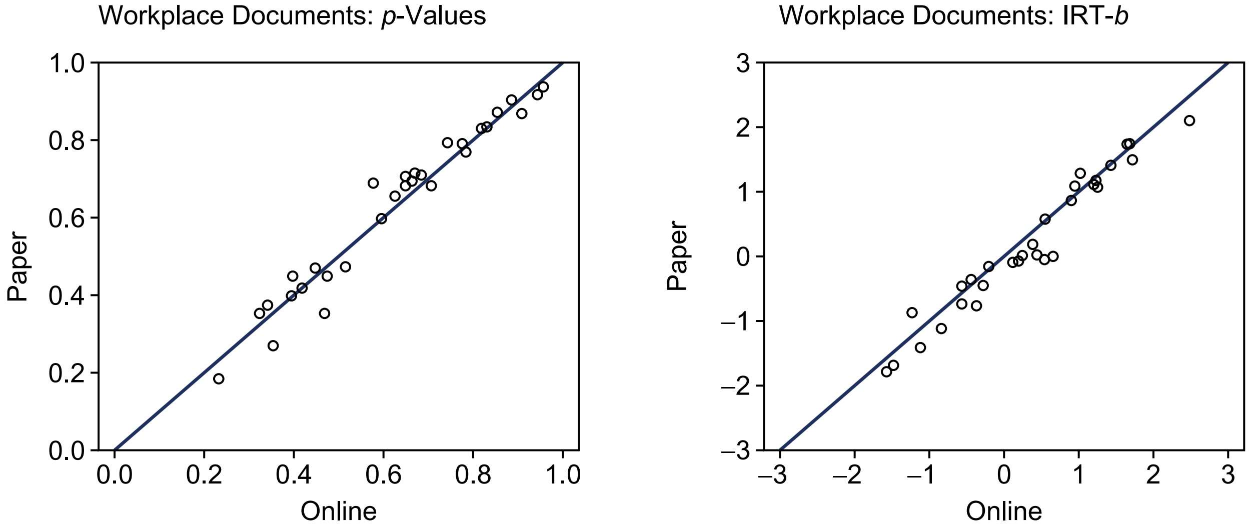
Separate calibrations were conducted for the online and paper forms, and the item parameter estimates were transformed to be on the same pool scale. Table 9.7 shows the summary statistics for the online and paper forms, and Figure 9.4 presents the scatterplots of item *p*-values and *b*-parameter estimates. These results indicate that the item statistics for the two modes are similar.

Table 9.7. Test Summary Statistics for Workplace Documents

Mode	*P*	PBIS	IRT A	IRT B	IRT C
Online	.625 (.204)	.495 (.095)	1.238 (.316)	.342 (1.040)	.155 (.045)
Paper	.630 (.213)	.484 (.081)	1.111 (.384)	.210 (1.066)	.131 (.041)

Note. P = *p*-value; PBIS = point-biserial correlation. Standard deviations are in parentheses.

Figure 9.4. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Workplace Documents

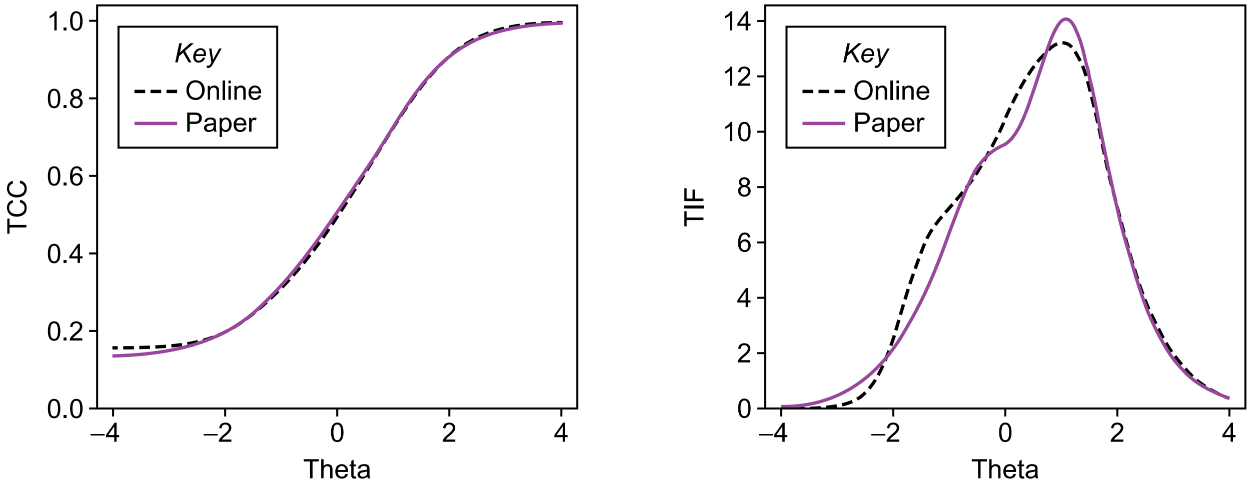


Differential item functioning (DIF) analysis was also conducted between the items on the paper and online forms. Three items were flagged as Category C (favoring one online-form item and two paper-form items) using the Mantel-Haenszel method.

Test Comparison

Figure 9.5 shows the comparisons of the test characteristic curves (TCCs) and test information functions (TIFs). The TCCs are almost identical, and the TIFs are very similar, which indicates that the average mode effect is negligible.

Figure 9.5. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Workplace Documents



Score Conversion Comparison

Figure 9.6 compares the raw-to-scale score conversions of the two testing modes. As is evident in the right side of the figure, there are only three raw score points that corresponded to different reported scale scores depending on the testing mode; however, the raw score cuts that corresponded to the level score cuts are identical.

Figure 9.6. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Workplace Documents

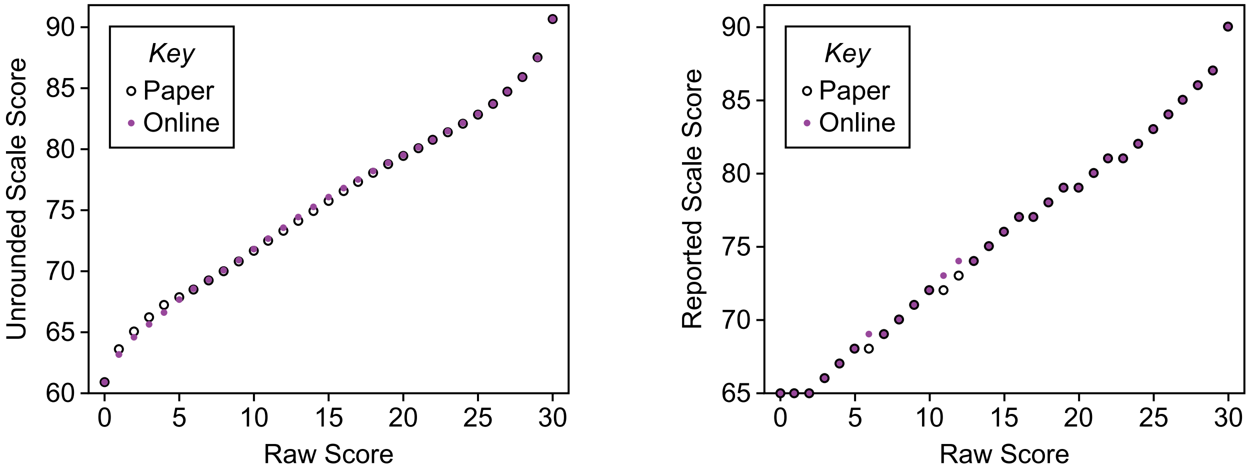
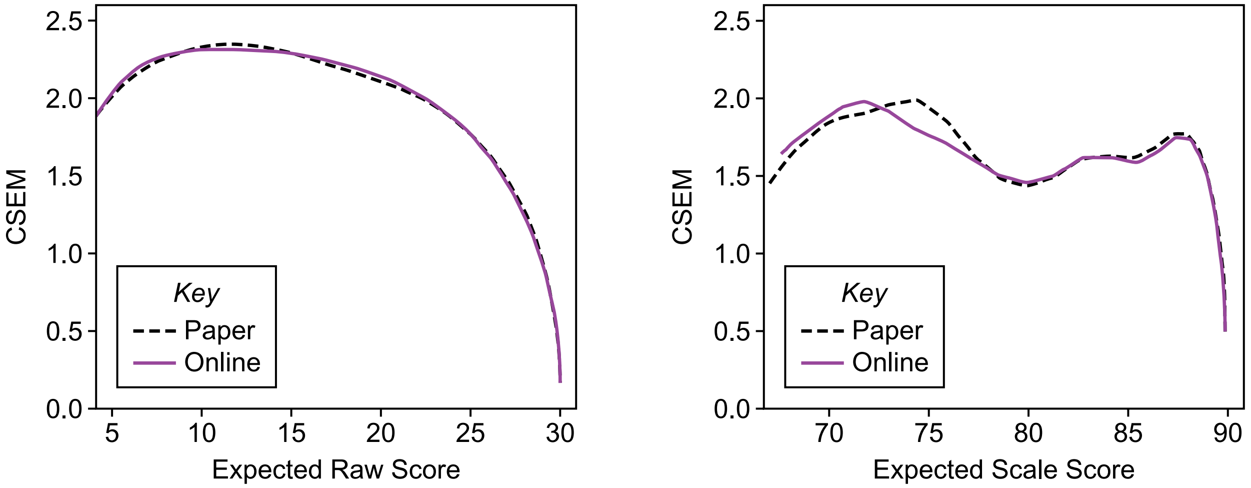


Figure 9.7 shows the conditional standard error of measurements (CSEMs). The raw score CSEMs tend to show a typical upside-down U shape, and the scale score CSEMs tend to be flat for most of the score points. For each score type, the CSEMs of the two modes are similar.

Figure 9.7. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Workplace Documents



9.2.4.2.2 Applied Math

Item-Level Comparison

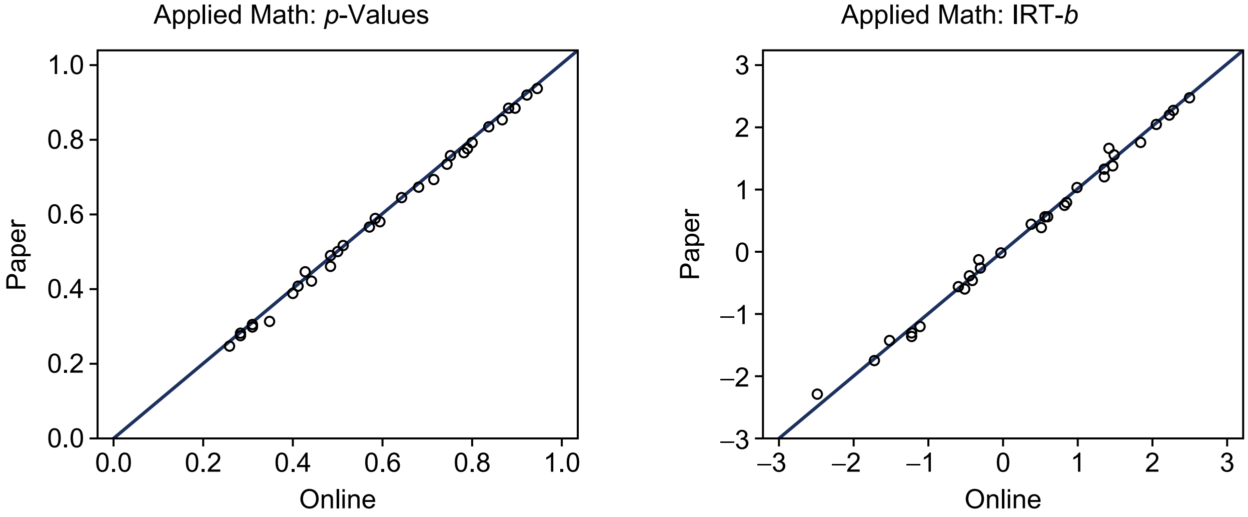
Separate calibrations were conducted for the online and paper forms, and the item parameter estimates were transformed to be on the same pool scale. Table 9.8 shows the summary statistics for the online and paper forms, and Figure 9.8 presents the scatterplots of item *p*-values and IRT *b*-parameter estimates. These results indicate that the item statistics for the two modes are similar.

Table 9.8. Test Summary Statistics for Applied Math

Mode	*P*	PBIS	IRT A	IRT B	IRT C
Online	.625 (.206)	.521 (.113)	1.152 (.309)	.428 (1.317)	.155 (.056)
Paper	.625 (.213)	.512 (.101)	1.102 (.259)	.448 (1.390)	.151 (.044)

Note. P = *p*-value; PBIS = point-biserial correlation. Standard deviations are in parentheses.

Figure 9.8. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Applied Math

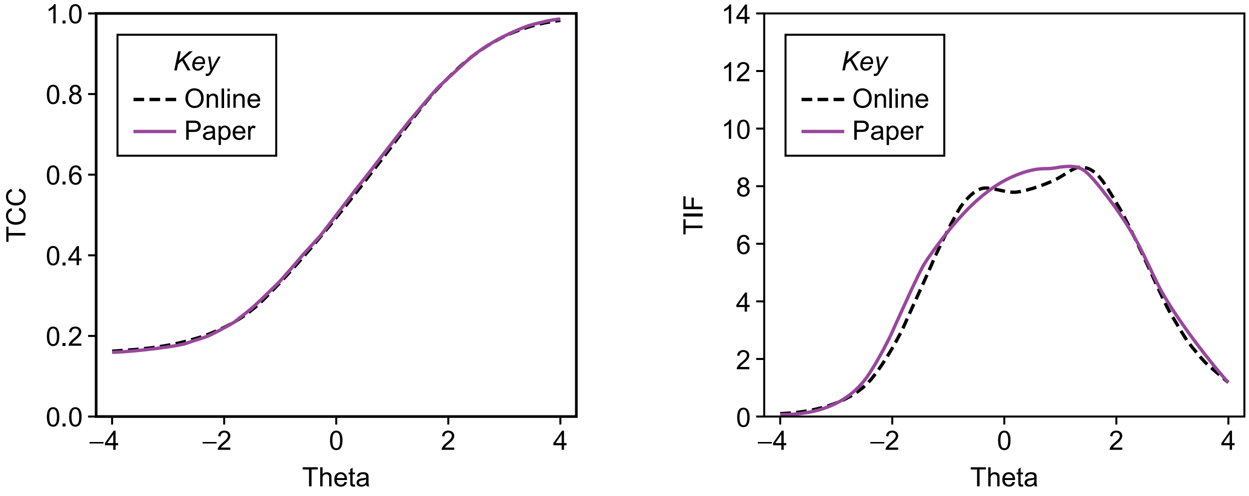


Differential item functioning (DIF) analysis was also conducted between the items on the paper and online forms. Only one item was flagged as Category C (favoring the paper testing) using the Mantel-Haenszel method.

Test Comparison

Figure 9.9 shows the comparisons of the test characteristic curves (TCCs) and test information functions (TIFs). The TCCs are almost identical, and the TIFs are very similar, which indicates that the average mode effect is negligible.

Figure 9.9. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Applied Math



Score Conversion Comparison

Figure 9.10 compares the raw-to-scale score conversions. The absolute differences between the unrounded scale scores of the two testing modes are smaller than 0.2 when the raw score points are lower than 21; the absolute differences are between 0.24 and 0.49 when the raw scores are larger than 22. As shown on the right side of Figure 9.10, only four raw score points correspond to different reported scale scores depending on the testing mode, which is mainly due to rounding errors. As for the raw score cuts that correspond to the level score cuts, only Level 5 had different raw score cuts depending on the testing mode; the raw score cut for Level 5 was either 21 (the corresponding unrounded scale score was 79.64) or 22 (the corresponding unrounded scale score was 80.24) depending on the testing mode, which was due to rounding error.

Figure 9.10. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Applied Math

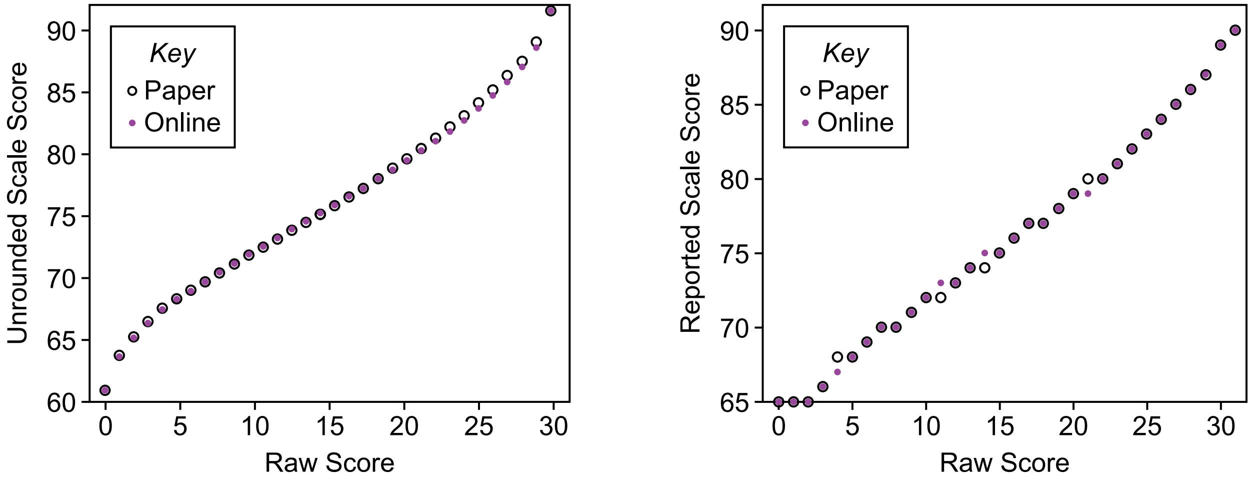
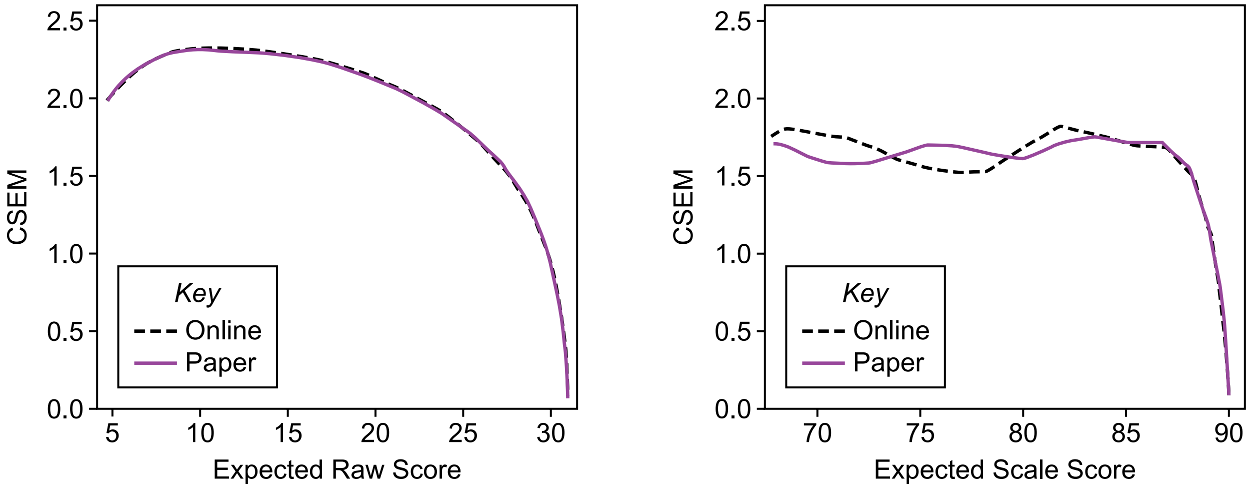


Figure 9.11 shows the conditional standard error of measurements (CSEMs). The raw score CSEMs tend to show a typical upside-down U shape, and the scale score CSEMs tend to be flat for most of the score points. For each score type, the CSEMs of the two modes are similar.

Figure 9.11. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Applied Math



9.2.4.2.3 Graphic Literacy

Item-Level Comparison

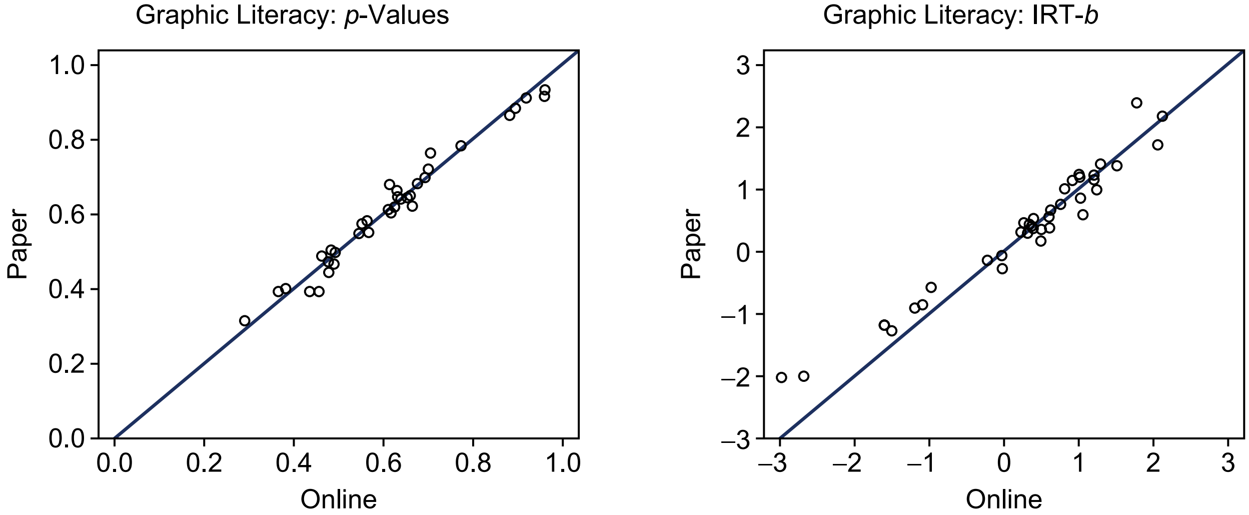
Separate calibrations were conducted for the online and paper forms, and the item parameter estimates were transformed to be on the same pool scale. Table 9.9 shows the summary statistics for the online and paper forms, and Figure 9.12 presents the scatterplots of item *p*-values and IRT *b*-parameter estimates. These results indicate that the item statistics for the two modes are similar.

Table 9.9. Test Summary Statistics for Graphic Literacy

Mode	*P*	PBIS	IRT A	IRT B	IRT C
Online	.616 (.173)	.474 (.100)	1.040 (.228)	.436 (1.096)	.185 (.052)
Paper	.618 (.171)	.465 (.084)	1.061 (.307)	.509 (1.002)	.176 (.043)

Note. P = *p*-value; PBIS = point-biserial correlation. Standard deviations are in parentheses.

Figure 9.12. Scatterplots of Item *p*-Values (Left) and IRT *b*-Parameter Estimates (Right) for the Two Delivery Modes of Graphic Literacy



Graphic Literacy: *p*-Values

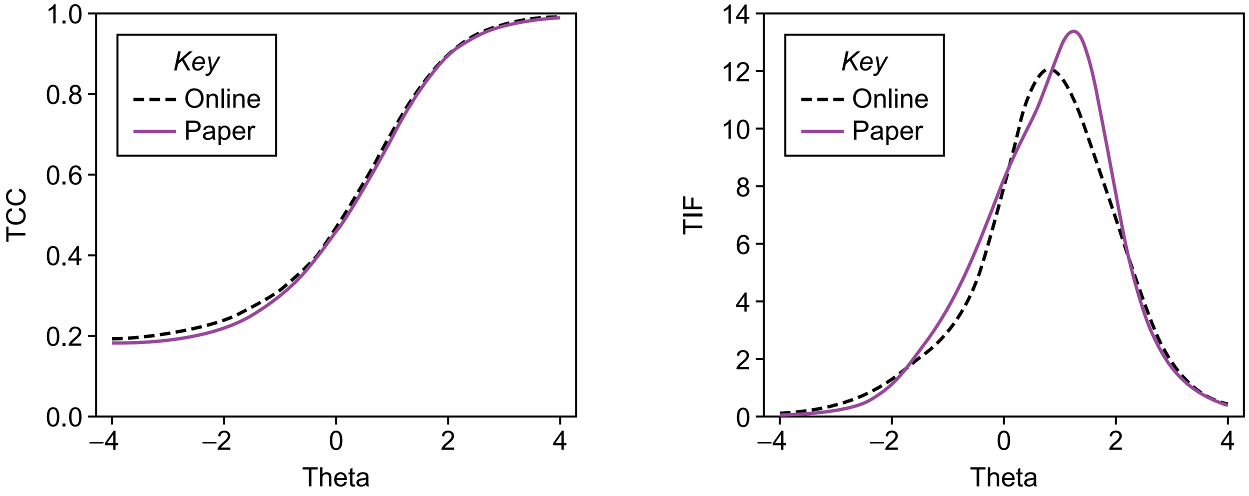
Graphic Literacy: IRT-*b*

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

Test Comparison

Figure 9.13 shows the comparisons of the test characteristic curves (TCCs) and test information functions (TIFs). The TCCs are almost identical, and the TIFs are very similar, which indicates that the average mode effect is negligible.

Figure 9.13. Comparisons of Test Characteristic Curves (Left) and Test Information Functions (Right) for the Two Delivery Modes of Graphic Literacy



Score Conversion Comparison

During the analysis, item 7 on the Graphic Literacy forms was found to be flawed, having no correct option. As a result, the following procedure was applied: (a) to obtain the raw-to-scale score conversion, IRT true-score equating was conducted, which equated the current form without the faulty item (31 items in total) to the base form (32 items in total); (b) to include all 32 raw score points, one score point was added to each raw score in the conversion that was obtained in step (a) (that is, the raw score 0 became the raw score 1, the raw score 1 became the raw score 2, and so on); and (c) to obtain the final conversion, a raw score of 0 was added to the conversion in step (b) and provided with the unrounded scale score of 60.00000 and the reported score of 65. Item 7 was scored as correct for all examinees who took the Graphic Literacy forms, and the scale score for each examinee was obtained by applying the conversion from step (c).

Figure 9.14 compares the raw-to-scale score conversions. Five raw score points corresponded to different scale scores depending on the testing mode; however, the raw score cuts that corresponded to the level score cuts were identical for the two modes.

Figure 9.14. Comparison of Unrounded (Left) and Reported (Right) Raw-to-Scale Score Conversions for the Two Delivery Modes of Graphic Literacy

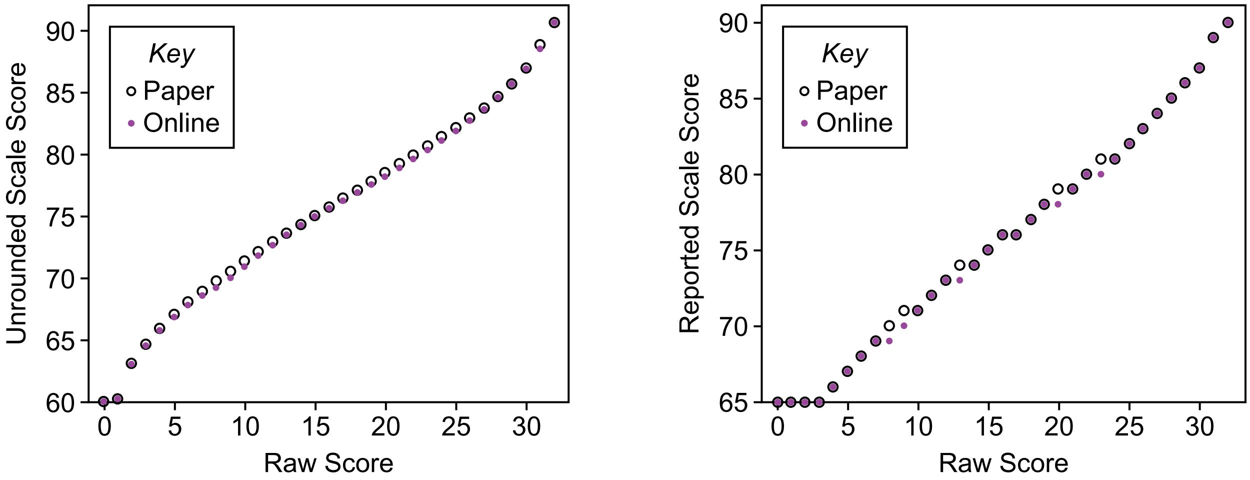
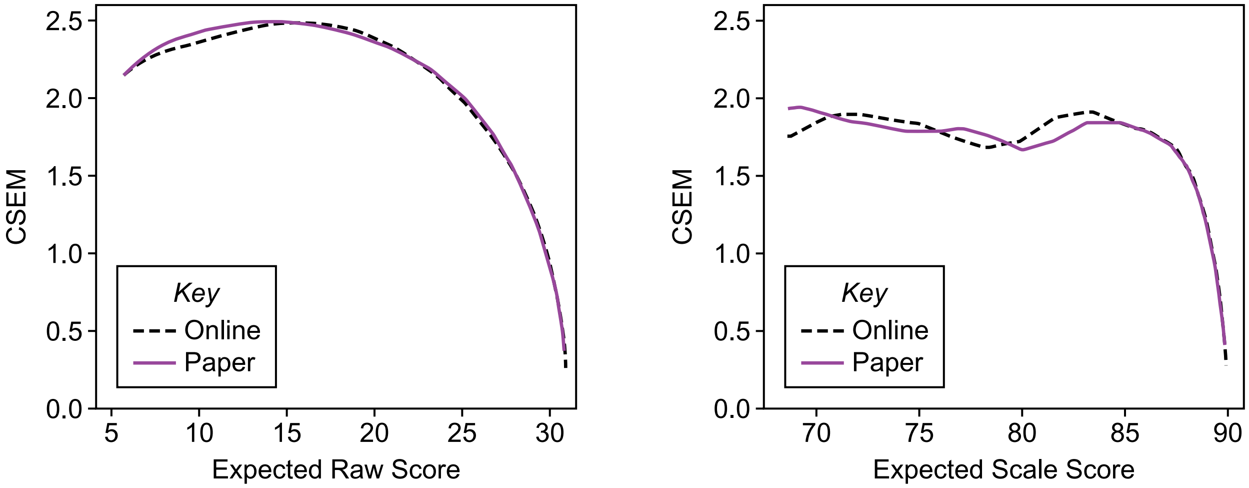


Figure 9.15 shows the conditional standard error of measurements (CSEMs). The raw score CSEMs tend to be a typical upside-down U shape, and the scale score CSEMs tend to be flat for most of the score points. For each score type, the CSEMs of the two modes are similar.

Figure 9.15. Comparison of CSEMs for Raw Scores (Left) and Scale Scores (Right) for the Two Delivery Modes of Graphic Literacy



9.2.4.3 Mode Comparability: Score Comparisons

9.2.4.3.1 Workplace Documents

Table 9.10 presents the summary statistics for the raw and scale scores by mode. Figure 9.16 presents the raw score distributions, and Figure 9.17 presents the scale score distributions. As is evident from the table and the figures, the results for the two modes are very similar. For both the raw and scale scores, the mean differences are below .17 and the effect sizes are below .025, which indicates that the two testing modes have almost identical score distributions.

Table 9.10. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Workplace Documents

Score	Mode	M	SD	P10	P25	P50	P75	P90	P95	M Diff.	ES	t-test prob
Raw Score	Online	18.74	6.53	10	14	19	24	27	28	.17	.025	.647
	Paper	18.91	6.31	10	14	19	24	27	28			
Scale Score	Online	78.57	5.09	72	75	79	82	85	86	.02	.003	.960
	Paper	78.59	4.97	72	75	79	82	85	86			

Note. M Diff. = mean difference; ES = effect size

Figure 9.16. Comparison of Raw Score Distributions for the Two Delivery Modes of Workplace Documents

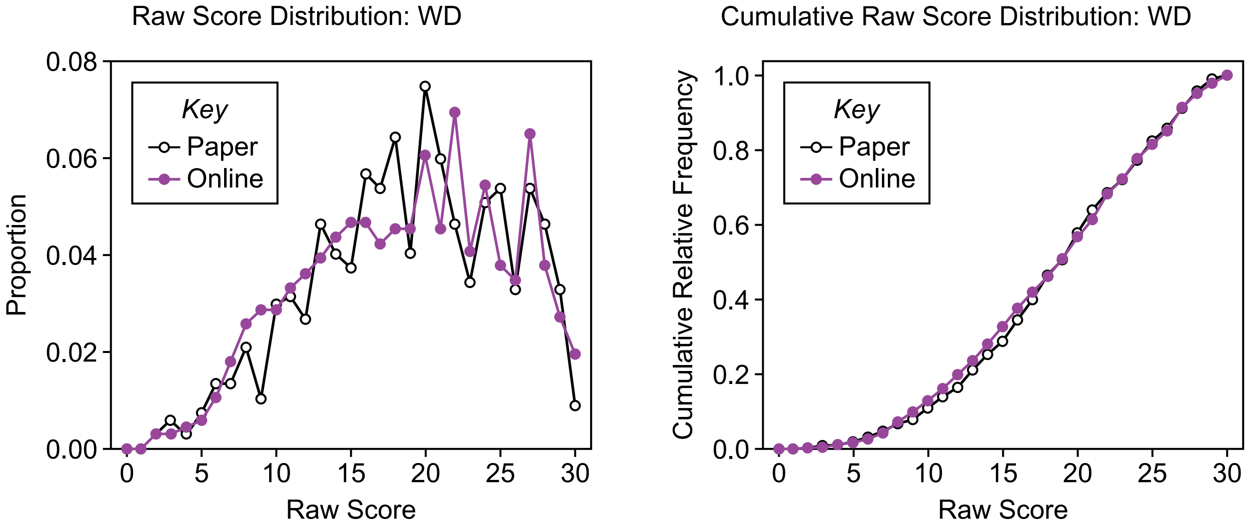
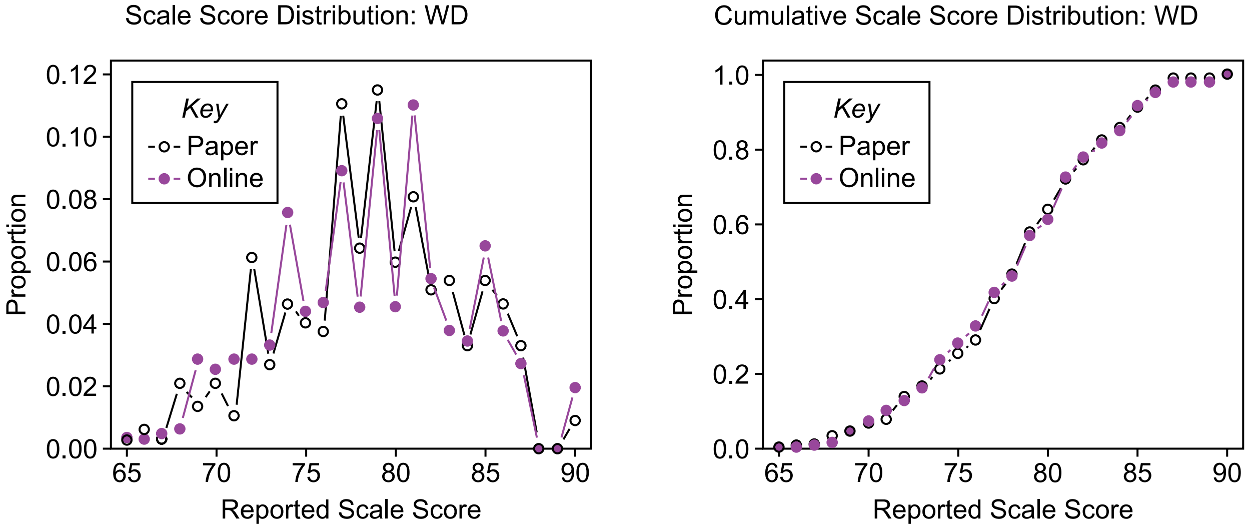


Figure 9.17. Comparison of Scale Score Distributions for the Two Delivery Modes of Workplace Documents



Based on the findings from the mode comparability analyses, ACT concluded that no significant mode effect existed in Workplace Documents. Due to the limitations of the field test data, ACT will continue to monitor the Workplace Documents assessment for potential mode effects to ensure that test scores derived from the paper and online administrations are comparable.

9.2.4.3.2 Applied Math

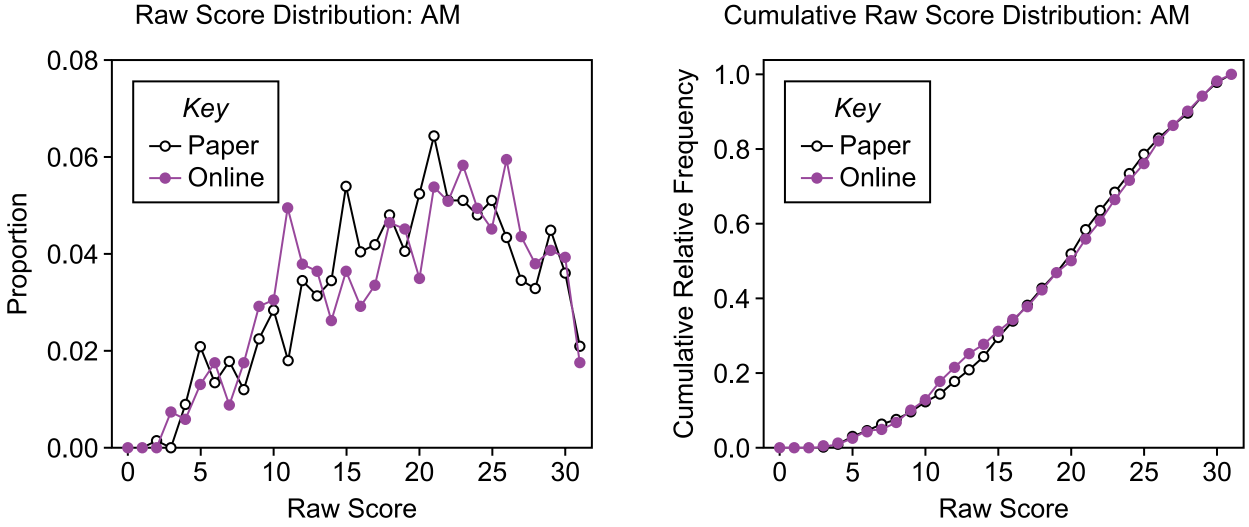
Table 9.11 presents the summary statistics for the raw and scale scores by mode. Figure 9.18 presents the raw score distributions, and Figure 9.19 presents the scale score distributions. The results for the two modes are very similar. For both the raw and scale scores, the mean differences are below .01 and the effect sizes are below .002, which indicates that the two modes have almost identical score distributions.

Table 9.11. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Applied Math

Score	Mode	M	SD	P10	P25	P50	P75	P90	P95	M Diff.	ES	t-test prob
Raw Score	Online	19.37	7.07	10	13	20	25	28	30	.01	.002	.969
	Paper	19.38	6.87	10	15	20	25	29	30			
Scale Score	Online	78.82	5.62	72	74	79	83	86	89	.00	.000	.993
	Paper	78.82	5.52	72	75	79	83	87	89			

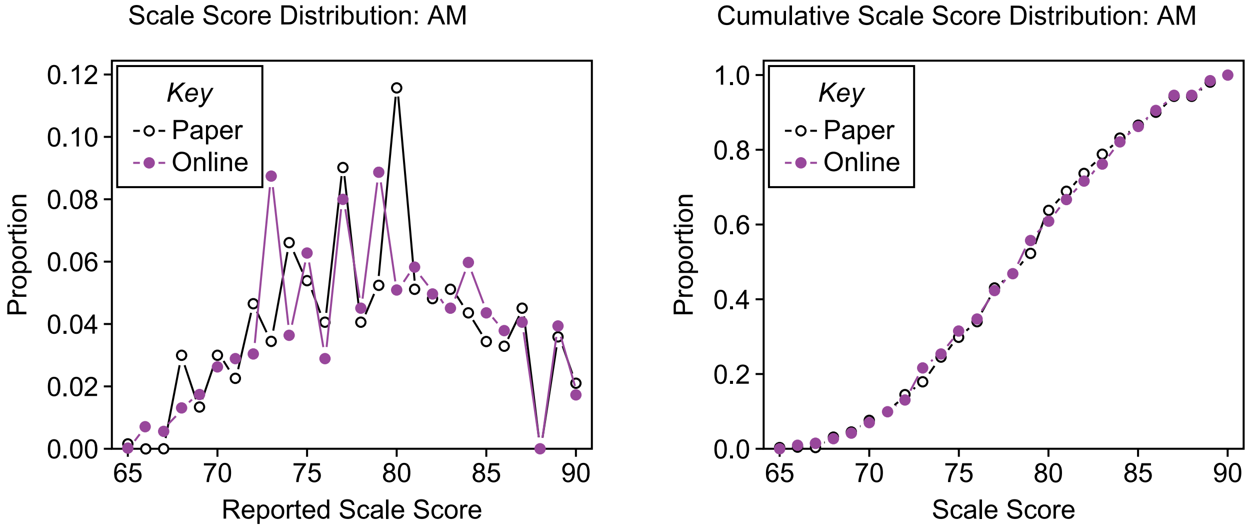
Note. M Diff. = mean difference; ES = effect size

Figure 9.18. Comparison of Raw Score Distributions for the Two Delivery Modes of Applied Math


Raw Score Distribution: AM

Cumulative Raw Score Distribution: AM

Figure 9.19. Comparison of Scale Score Distributions for the Two Delivery Modes of Applied Math


Scale Score Distribution: AM

Cumulative Scale Score Distribution: AM

Based on the findings from the mode comparability analyses, ACT concluded that no significant mode effect existed in Applied Math. Due to the limitations of the field test data, ACT will continue to monitor the Applied Math assessment for potential mode effects to ensure that test scores derived from the paper and online administrations are comparable.

9.2.4.3.3 Graphic Literacy

Table 9.12 presents the summary statistics for the raw and scale scores by mode. Figure 9.20 presents the raw score distributions, and Figure 9.21 presents the scale score distributions. The results for the two modes are very similar. For both the raw and scale scores, the mean differences are below .21 and the effect sizes are below .041, which indicates that the two modes have almost identical score distributions.

Table 9.12. Summary Statistics for Raw and Scale Scores for the Two Delivery Modes of Graphic Literacy

Score	Mode	M	SD	P10	P25	P50	P75	P90	P95	M Diff.	ES	t-test prob
Raw Score	Online	20.09	6.76	11	15	20	26	29	20	.08	.012	.829
	Paper	20.17	6.57	11	15	20	26	29	30			
Scale Score	Online	79.70	5.33	72	75	78	83	86	87	.21	.041	.453
	Paper	79.91	5.09	72	75	79	83	86	87			

Note. M Diff. = mean difference; ES = effect size

Figure 9.20. Comparison of Raw Score Distributions for the Two Delivery Modes of Graphic Literacy

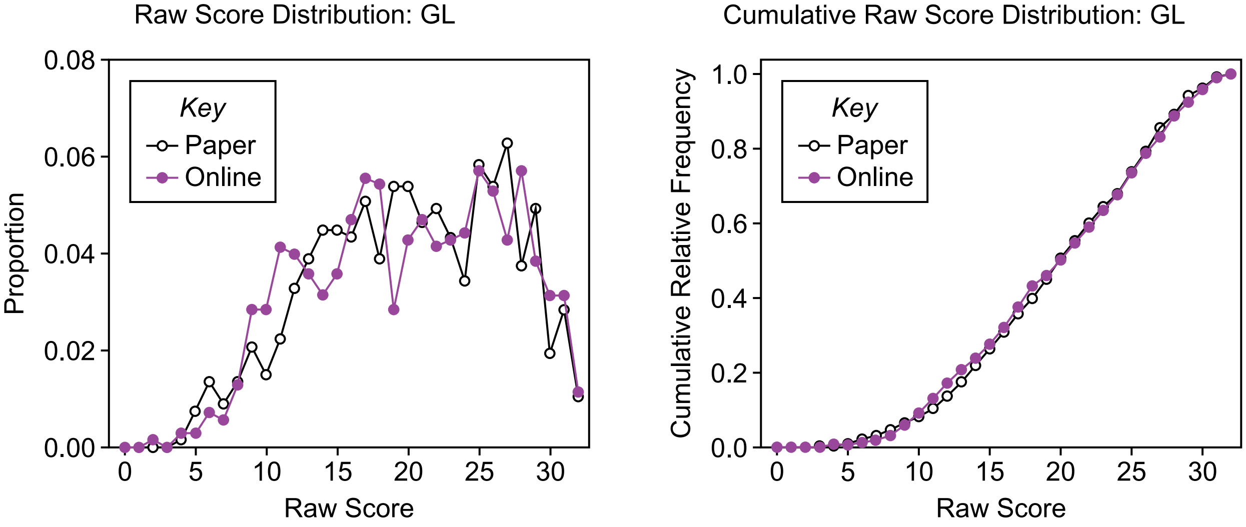
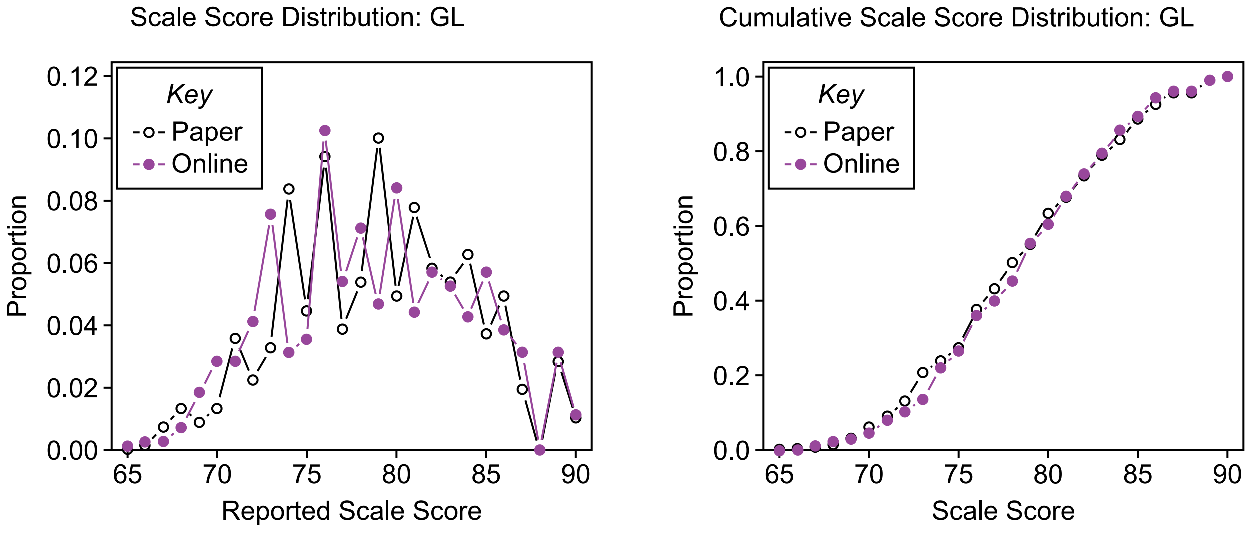


Figure 9.21. Comparison of Scale Score Distributions for the Two Delivery Modes of Graphic Literacy



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

9.2.5 Dimensionality

ACT analyzes WorkKeys NCRC assessment item data using a unidimensional item response theory (IRT) model (Hambleton & Swaminathan, 1985; Lord, 1980). WorkKeys NCRC 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 is attributable to a single underlying factor.

ACT applied exploratory factor analysis (EFA) to assess dimensionality for the WorkKeys NCRC assessments. EFA uses an inter-item correlation matrix to identify factors underlying observed item variance. In the analysis, ACT applied four criteria to assess dimensionality. The first criterion was a scree plot of eigenvalues, which 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. The second criterion was the percentage of total variance accounted for by a factor. Hatcher (1994) suggests that a factor should be retained if it accounts for at least 10% of the total variance. The third criterion was the percentage explained by the first factor. Reckase (1979) suggests that if the first factor explains 20% of the variance of a set of items, the item set should be considered unidimensional. Finally, ACT applied factor difference ratio index (FDRI). Hattie (1985) maintains that the first factor is relatively strong if the FDRI (Johnson et al., 2003) is greater than 3. The 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 factor and the third factor.

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

9.2.5.1 Workplace Documents

Figure 9.22 is the scree plot derived from the correlation matrix of item scores for the Workplace Documents assessment. This figure reveals that the “elbow” appears immediately after the first eigenvalue. Table 9.13 summarizes the eigenvalues and FDRI for both test forms. The table indicates that the percentage of variances accounted for by the first factor is approximately 40%, while the percentage accounted for by the second factor is less than 10%. Additionally, Table 9.13 indicates that the FDRI is 13.24, which is significantly greater than 3. These findings consistently indicate that a single factor underlies the item scores on the Workplace Documents assessment.

Figure 9.22. Workplace Documents Eigenvalue Scree Plot

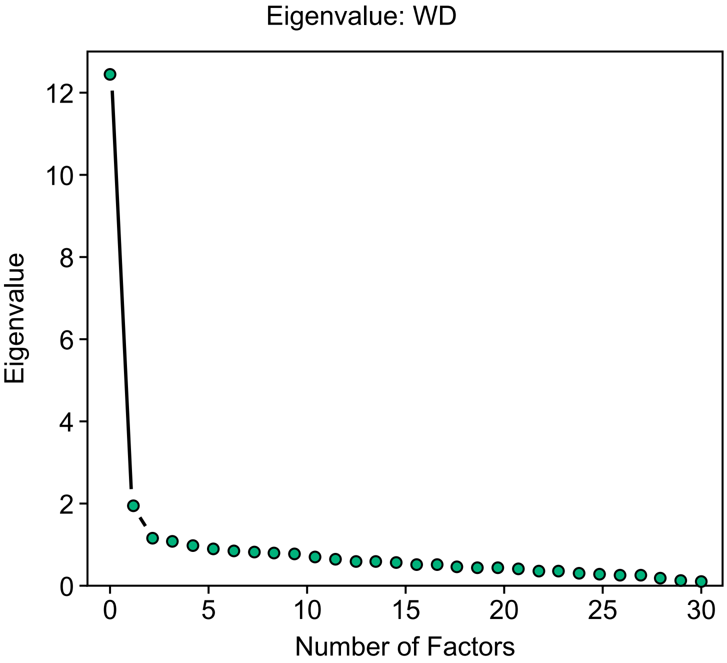


Table 9.13. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Workplace Documents

Factor	Eigenvalue	Difference Between Eigenvalues	FDRI
1	12.45 (41.5%)	—	—
2	1.96 (6.5%)	10.50	—
3	1.16 (3.9%)	.79	13.24

Note. Each number in parentheses is the percentage of total variance accounted for by that factor.

9.2.5.2 Applied Math

Figure 9.23 is the scree plot derived from the correlation matrix of item scores on the Applied Math assessment. The figure reveals that the “elbow” appears immediately after the first eigenvalue. Table 9.14 summarizes the eigenvalues and FDRI for both test forms. The table indicates that the percentage of variances accounted for by the first factor is nearly 40%, while the percentage accounted for by the second factor is less than 10%. Additionally, Table 9.14 indicates that the FDRI is 6.75, which is significantly greater than 3. These findings consistently indicate that a single factor underlies item scores on the Applied Math assessment.

Figure 9.23. Applied Math Eigenvalue Scree Plot

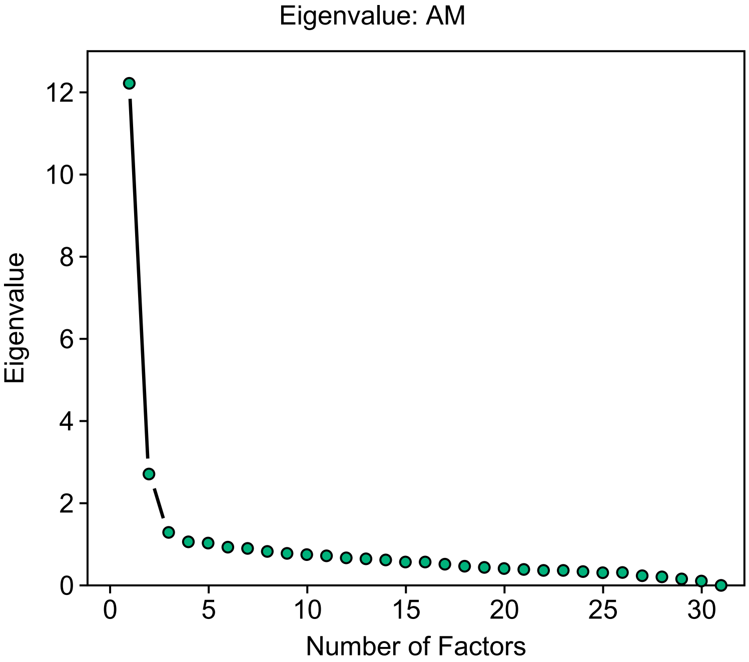


Table 9.14. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Applied Math

Factor	Eigenvalue	Difference Between Eigenvalues	FDRI
1	12.21 (39.4%)	—	—
2	2.70 (8.7%)	9.51	—
3	1.30 (4.2%)	1.41	6.75

Note. Each number in parentheses is the percentage of total variance accounted for by that factor.

9.2.5.3 Graphic Literacy

Figure 9.24 presents the scree plot derived from the correlation matrix of item scores on the Graphic Literacy assessment. The figure reveals that the "elbow" appears immediately after the first eigenvalue. Table 9.15 summarizes the eigenvalues and FDRI for both test forms. The table indicates that the percentage of variances accounted for by the first factor is approximately 30%, while the percentage accounted for by the second factor is less than 10%. Additionally, Table 9.15 indicates that the FDRI is 18.32, which is significantly greater than 3. These findings consistently indicate that a single factor underlies item scores on the Graphic Literacy assessment.

Figure 9.24. Graphic Literacy Eigenvalue Scree Plot

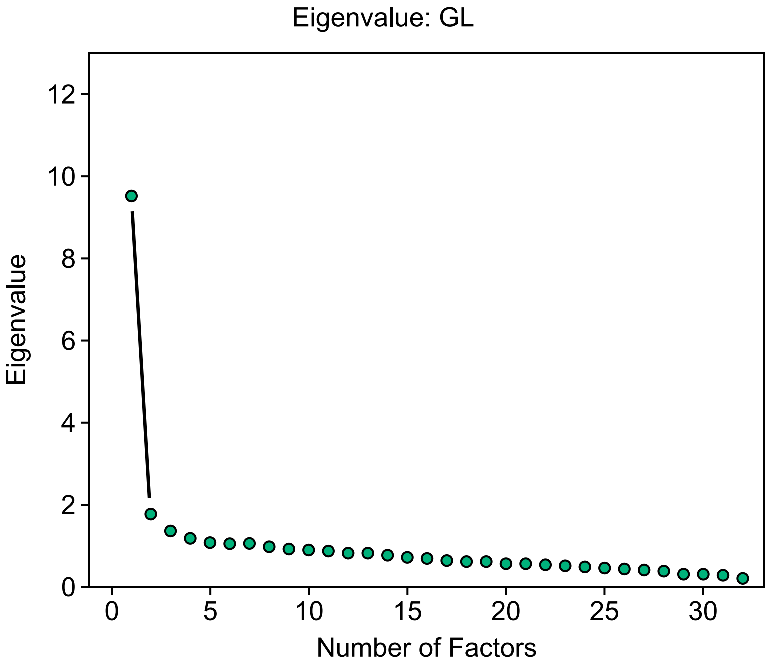


Table 9.15. Summary of Eigenvalues and Factor Difference Ratio Index (FDRI) for Graphic Literacy

Factor	Eigenvalue	Difference Between Eigenvalues	FDRI
1	9.53 (29.8%)	—	—
2	1.78 (5.6%)	7.75	—
3	1.36 (4.2%)	.42	18.32

Note. Each number in parentheses is the percentage of total variance accounted for by that factor.

9.2.6 IRT Modeling: Local Item Independence

The 3-PL IRT model assumes that items are locally independent, which means that an examinee's responses to different items on an assessment are statistically independent of each other after the examinee's ability has been considered. For the assumption to be met, an examinee's 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 the examinee's ability (Hambleton & Swaminathan, 1985).

9.2.6.1 Workplace Documents

The Workplace Documents assessment includes a series of reading passages. Each reading passage has two or three items associated with the passage. This raises the possibility of statistical dependence between items that share passages, so determining whether items are locally independent is critical for applying an IRT model to Workplace Documents.

ACT used Q_3 (Yen, 1984) to evaluate the local item dependence for the items within a Workplace Documents 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 interpreted as 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 within a set. The 95th percentile of the baseline was defined as the cut point. If the Q_3 for a pair of items within a set was larger than the cut point, the item pair was considered dependent.

ACT used test data from the second field test study to generate the Q_3 matrix to evaluate whether local item dependence was present. The Q_3 matrix for the Workplace Documents scaling form indicated that the items within each item set did not show higher correlations than the independent items outside of those item sets. Consequently, after reviewing all of the items on the form, ACT concluded that no compelling evidence of item dependence existed. Thus, the items on the Workplace Documents assessment met the assumption of local independence.

9.2.6.2 Applied Math

Items on the Applied Math assessment are discrete, meaning that each item is a single element on the assessment designed to assess a single skill in isolation. Discrete items do not share common stimulus materials (e.g., graphics or reading passages) with other items. After an Applied Math assessment form is assembled, content and research specialists review the form to ensure that the form is free of clueing (that is, that no item on the form provides information that may prompt or assist an examinee in answering a different item on the form). ACT form development quality assurance specifies that developers review each other's work to ensure that no item can be a clue for solving another item on the form. Consequently, items on the Applied Math assessment are locally independent because of the assessment's design properties and the assessment's quality assurance specifications.

9.2.6.3 Graphic Literacy

The Graphic Literacy assessment includes a series of graphics, where each graphic has two or three items associated with the graphic. This raises the possibility of statistical dependencies between items that share graphics, so determining whether items and item sets are independent 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 dependent.

ACT used test data from the second field test study to generate the Q_3 matrix to evaluate whether local item dependence was present. The Q_3 matrix for the Graphic Literacy scaling form indicated that the items within sets did not show higher correlations than the independent items. Consequently, after reviewing all of the items on the form, ACT concluded that no compelling evidence of item dependence existed. Thus, the items on the Graphic Literacy assessment met the assumption of local independence.

9.3 Linking

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 this understanding, test publishers often use a statistical procedure to link the scores from one test to those from the other. There are generally four types of linking, which are ordered in terms of the strength of the resulting relationship (strongest to weakest): equating, calibration, projection, and moderation (Linn, 1993; Mislevy, 1992). Concordance is a type of statistical moderation of matching distributions that uses percentile ranks to derive a table that links the scores from two tests. Dorans et al. (2007) pointed out that “concordances represent scalings 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 the same test, which produces comparable scores, the concordance of two tests produces scores that are *not* interchangeable.

9.3.1 Linking Reading for Information to the Workplace Documents Score Scale

The Workplace Documents assessment was developed based on modified test specifications from the Reading for Information assessment (see Chapter 3 for the test specifications). To facilitate a smooth transition from Reading for Information to Workplace Documents, ACT conducted a linking study in spring 2017. The focus of the linking study was to develop a concordance between the Reading for Information and Workplace Documents assessments. To find the scale score in Reading for Information that corresponded to a particular scale score in Workplace Documents, ACT identified the Reading for Information score that had the same percentile rank as the Workplace Documents score. For example, if the percentage of examinees who earned a scale score of 74 or below in Workplace Documents was 20%, and the percentage of examinees who earned a scale score of 76 or below in Reading for Information was 20%, then a Reading for Information scale score of 76 corresponds to (i.e., concords with) a Workplace Documents scale score of 74. This document summarizes the findings from the linking study to better illustrate the relationships between the two assessments and to assist users in appropriately interpreting the scores or score trends derived from the two assessments.

9.3.1.1 Study Design and Sample Representativeness

A total of 43 testing sites (10 high schools and 33 adult testing centers across 20 states) administered both forms W2C_LM1 (Workplace Documents online) and W1C_LM3 (Reading for Information online). A total of 1,613 examinees were given 55 minutes to complete one of the two linking forms; out of the 1,613 examinees, 800 examinees took W2C_LM1 and 813 examinees took W1C_LM3. In terms of demographic characteristics, the recruited sample was generally representative of the WorkKeys NCRC test population.

Because the Workplace Documents assessment was developed based on modified constructs and test specifications from the Reading for Information assessment, the resulting concorded scores are not interchangeable. When concorded forms have similar difficulty and measurement precision, the concordance results are likely to be stronger and more stable. A series of analyses was conducted to evaluate and compare the psychometric properties of the two assessments in terms of omit rates, testing times, scale score summary statistics, reliability, and standard error of measurement (SEM).

9.3.1.2 Comparison of Omit Rates and Testing Times Between Reading for Information and Workplace Documents

Figure 9.25 presents the omit rates for each item on both the Workplace Documents and the Reading for Information forms administered in the linking study. Figure 9.25 indicates that the omit rates were less than 10% for all items. In addition, as summarized in Table 9.16, on average, the examinees spent more time on form W1C_LM3 than on form W2C_LM1. It should be noted that there were two more pretest items on the Workplace Documents assessment than on the Reading for Information assessment.

Figure 9.25. Comparison of Item Omit Rates Between Reading for Information and Workplace Documents

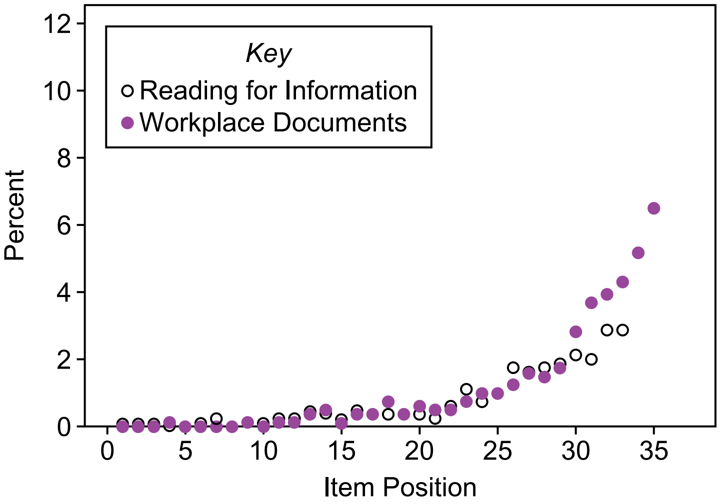


Table 9.16. Summary Statistics for Total Testing Times (in Minutes) for Reading for Information and Workplace Documents

Form	*N*	Mean (SD)	Min	P5	P10	P25	P50	P75	P90	P95
W1C_LM3	813	33.83 (13.12)	5	9	12	19	29	40	49	53
W2C_LM1	800	29.66 (13.20)	6	12	15	24	34	45	52	54

9.3.1.3 Scale Score Distributions for Reading for Information and Workplace Documents

Because no significant mode effect was observed in the mode study, the item parameter estimates were then recalibrated using the combined data from both the paper and online administrations to derive the conversion for the Workplace Documents form (W2C_LM1). Tables 9.17 and 9.18 provide the summary statistics for the raw scores and the scale scores for the linking study. Based on the average IRT-*b* statistics, the Workplace Documents form, W2C_LM1, appears to be slightly easier than the Reading for Information form, W1C_LM3.

Table 9.17. Test Summary Statistics for Reading for Information and Workplace Documents

Form	*P*	PBIS	IRT A	IRT B	IRT C
W1C_LM3	.613 (.219)	.457 (.094)	1.079 (.319)	.377 (1.150)	.155 (.054)
W2C_LM1	.626 (.208)	.491 (.093)	1.232 (.335)	.293 (1.048)	.147 (.045)

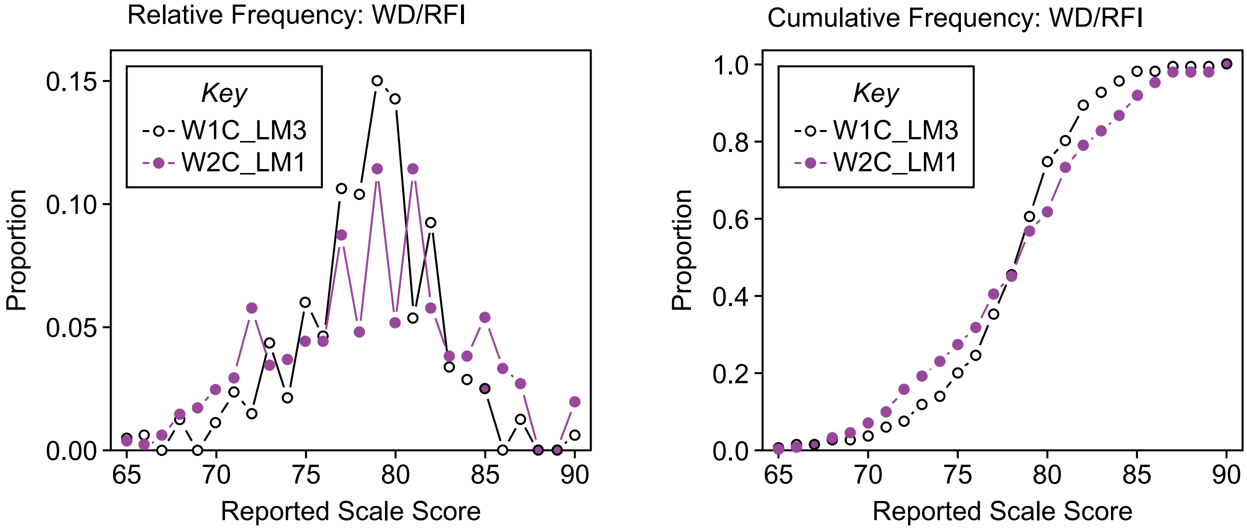
Note. P = *p*-value; PBIS = point-biserial correlation. Standard deviations are in parentheses.

Table 9.18. Scale Score Summary Statistics for Reading for Information and Workplace Documents

Form	*N*	Mean (SD)	P5	P10	P25	P50	P75	P90	P95
W1C_LM3	800	78.38 (3.91)	71	73	77	79	81	83	84
W2C_LM1	813	78.51 (5.10)	70	72	75	79	82	85	86

Figure 9.26 presents the relative frequency distributions and cumulative frequency distributions for the Reading for Information and Workplace Documents forms. These plots suggest that the Reading for Information and Workplace Documents scale score distributions are different because significant modifications were made to the Workplace Documents assessment.

Figure 9.26. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Reading for Information and Workplace Documents



9.3.1.4 Concordance From Reading for Information to Workplace Documents

Given the changes in test specifications and the need to link the Reading for Information and Workplace Documents assessments, statistical moderations using an equating method were performed to link scores from the Reading for Information assessment (RFI 1.0) to those from the Workplace Documents assessment (WD 2.0). Concordance from Reading for Information to Workplace Documents was conducted using the equipercentile method with a smoothing (S) value of .05.

9.3.1.5 Evaluation of Reading for Information Forms After Linking

Table 9.19 provides the summary statistics for the scale scores for the Reading for Information form before the form was transformed to be on the Workplace Documents scale (W1C_LM3) and after the form was transformed to be on the Workplace Documents scale (W1C_LM3*). The table also provides the summary statistics for the scale scores for the Workplace Documents form (W2C_LM1). It can be observed that the means, standard deviations, and quantiles of the transformed Reading for Information form (W1C_LM3*) are very similar to those of the Workplace Documents form (W2C_LM1).

Table 9.19. Summary Statistics for Scale Scores for Reading for Information and Workplace Documents

Scale	Form	*N*	Mean (SD)	P10	P25	P50	P75	P90	P95
RFI 1.0	W1C_LM3	800	78.38 (3.91)	73	77	79	81	83	84
WD 2.0	W1C_LM3*	800	78.46 (5.05)	71	76	79	82	85	86
WD 2.0	W2C_LM1	813	78.51 (5.10)	72	75	79	82	85	86

Note. W1C_LM3* was derived by applying the concordance table to transform the scores in W1C_LM3 to be on the WD 2.0 scale.

Table 9.20 provides the summary statistics for the level scores for the Reading for Information form before the form was transformed to be on the Workplace Documents scale (W1C_LM3) and after the form was transformed to be on the Workplace Documents scale (W1C_LM3*). The table also provides the summary statistics for the level scores for the Workplace Documents form (W2C_LM1). The means and standard deviations of W1C_LM3* and W2C_LM1 are very similar except for the P10 quantile. The level cuts for the Workplace Documents assessment were developed based on a standard-setting study using the Mapmark method. (For greater detail on the standard-setting process, please see Chapter 8.)

Table 9.20. Summary Statistics for Level Scores of Workplace Documents

Scale	Form	*N*	Mean (SD)	P10	P25	P50	P75	P90	P95
RFI 1.0	W1C_LM3	800	4.43 (1.56)	3	4	5	5	6	7
WD 2.0	W1C_LM3*	800	3.96 (1.86)	<3	3	4	5	6	7
WD 2.0	W2C_LM1	813	4.06 (1.79)	3	3	4	5	6	7

The results suggest that to compare the scores from the Reading for Information and Workplace Documents assessments and to understand the score relationships between the two assessments, the scale scores on the Reading for Information assessment need to be transformed to be on the Workplace Documents scale based on the concordance table. Test users need to be aware that the concordant scale scores do not represent the test scores that examinees would achieve if they took the Workplace Documents assessment. Similarly, test users should be cautious when using only concordance scores to compare group performance averages or to analyze year-to-year performance trends. As noted at the beginning of Section 9.3, the concorded scores are not comparable or interchangeable across different forms, unlike the equated scores.

9.3.2 Linking Applied Mathematics to the Applied Math Score Scale

The Applied Math assessment was developed based on modified test specifications from the Applied Mathematics assessment (see Chapter 3 for the test specifications). To facilitate a smooth transition from Applied Mathematics to Applied Math, ACT conducted a linking study in spring 2017. The focus of the linking study was to develop a concordance between the Applied Mathematics and Applied Math assessments. To find the scale score in Applied Mathematics that corresponded to a particular scale score in Applied Math, ACT identified the Applied Mathematics score that had the same percentile rank as the Applied Math score. For example, if the percentage of examinees who earned a scale score of 74 or below in Applied Math was 20%, and the percentage of examinees who earned a scale score of 76 or below in Applied Mathematics was 20%, then an Applied Mathematics scale score of 76 corresponds to (i.e., concords with) an Applied Math scale score of 74. This document summarizes the findings from the linking study to better illustrate the relationships between the two assessments and to assist users in appropriately interpreting the scores or score trends derived from the two assessments.

9.3.2.1 Study Design and Sample Representativeness

A total of 43 testing sites (10 high schools and 33 adult testing centers across 20 states) administered both forms M2C_LM1 (Applied Math online) and M1C_LM3 (Applied Mathematics online). A total of 1,656 examinees were given 55 minutes to complete one of the two linking forms; out of the 1,656 examinees, 835 examinees took M2C_LM1 and 821 examinees took M1C_LM3. In terms of demographic characteristics, the recruited sample was generally representative of the WorkKeys NCRC test population.

Because the Applied Math assessment was developed based on modified constructs and specifications from the Applied Mathematics assessment, the resulting concordance scores are not interchangeable. When concorded forms have similar difficulty and measurement precision, the concordance results are likely to be stronger and more stable.

A series of analyses was conducted to evaluate and compare the psychometric properties of the two assessments in terms of omit rates, testing times, scale score summary statistics, reliability, and standard error of measurement (SEM).

9.3.2.2 Comparison of Omit Rates and Testing Times Between Applied Mathematics and Applied Math

Figure 9.27 presents the omit rates for each item on both the Applied Math and the Applied Mathematics forms administered in the linking study. The figure indicates that the omit rates were less than 10% for all items except the last item on the Applied Math form, M2C_LM1. In addition, as summarized in Table 9.21, on average, examinees spent slightly more time on form M2C_LM1 than on form M1C_LM3. It should be noted that there was one more operational item on the Applied Math assessment than on the Applied Mathematics assessment.

Figure 9.27. Comparison of Item Omit Rates Between Applied Mathematics and Applied Math

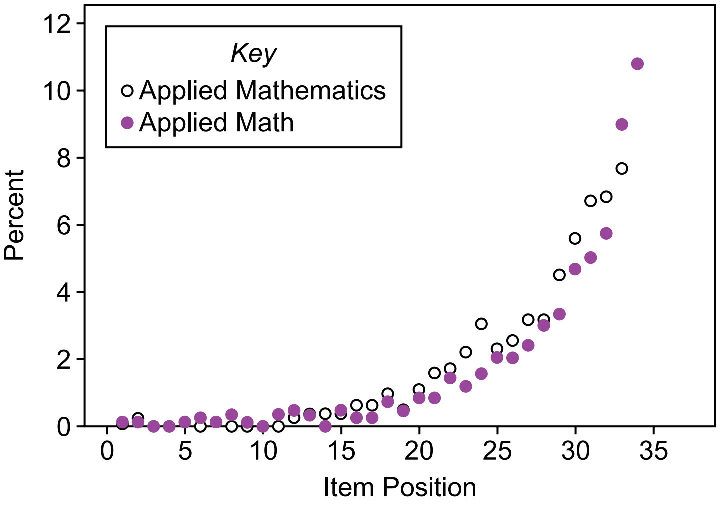


Table 9.21. Summary Statistics for Total Testing Times (in Minutes) for Applied Mathematics and Applied Math

Form	*N*	Mean (SD)	Min	P5	P10	P25	P50	P75	P90	P95
M1C_LM3	821	37.28 (14.00)	5	11	17	27	39	50	54	55
M2C_LM1	835	37.85 (13.73)	6	11	16	28	40	51	54	54

9.3.2.3 Scale Score Distributions for Applied Mathematics and Applied Math

Because no significant mode effect was observed in the mode study, the item parameter estimates were then recalibrated using the combined data from both the paper and online administrations to derive the conversion for the Applied Math form (M2C_LM1). Tables 9.22 and 9.23 provide the summary statistics for the raw and the scale scores for the linking study. Based on the average IRT-*b* statistics, the Applied Math form, M2C_LM1, appears to be slightly easier than the Applied Mathematics form, M1C_LM3.

Table 9.22. Test Summary Statistics for Applied Mathematics and Applied Math

Form	*P*	PBIS	IRT A	IRT B	IRT C
M1C_LM3	.633 (.205)	.492 (.106)	1.111 (.315)	.467 (1.312)	.171 (.055)
M2C_LM1	.620 (.214)	.510 (.110)	1.120 (.278)	.431 (1.350)	.152 (.055)

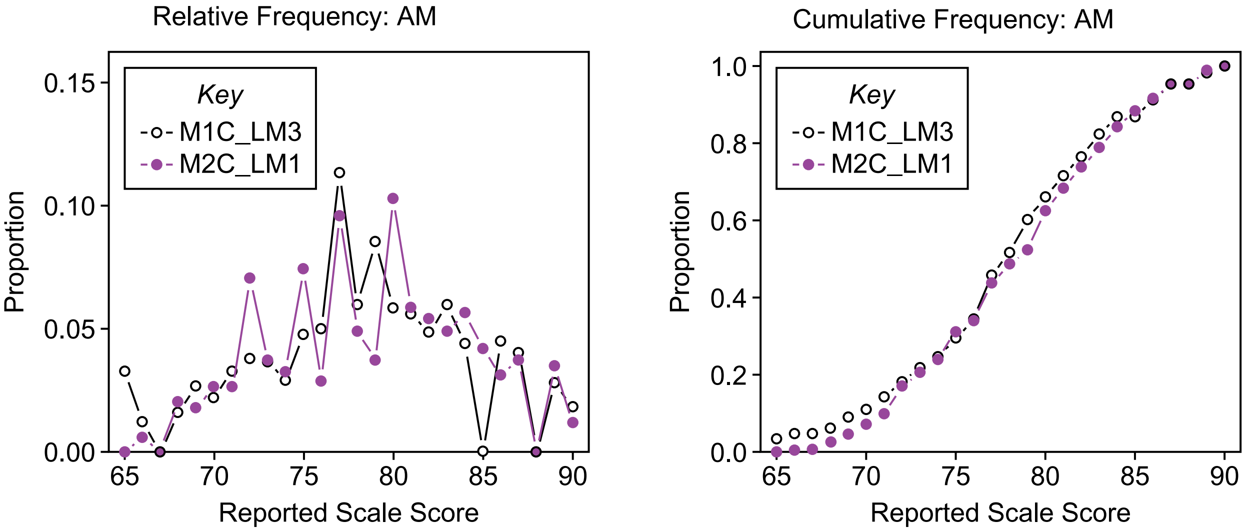
Note. P = *p*-value; PBIS = point-biserial correlation. Standard deviations are in parentheses.

Table 9.23. Scale Score Summary Statistics for Applied Mathematics and Applied Math

Form	*N*	Mean (SD)	P5	P10	P25	P50	P75	P90	P95
M1C_LM3	821	78.12 (5.87)	68	70	75	78	82	86	87
M2C_LM1	835	78.67 (5.46)	70	72	75	79	83	86	87

Figure 9.28 presents the relative frequency distributions and cumulative frequency distributions for the Applied Mathematics and Applied Math forms. These plots suggest that the scale score distributions are similar for the two assessments.

Figure 9.28. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Applied Mathematics and Applied Math



9.3.2.4 Concordance From Applied Mathematics to Applied Math

Given the changes in the test specifications and the need to link the Applied Mathematics and Applied Math assessments, statistical moderations using an equating method were performed to link scores from the Applied Mathematics assessment (AM 1.0) to those from the Applied Math assessment (AM 2.0). Concordance from Applied Mathematics to Applied Math was conducted using the equipercentile method with a smoothing (S) value of .05.

9.3.2.5 Evaluation of Applied Mathematics Forms After Linking

Table 9.24 provides the summary statistics for the scale scores for the Applied Mathematics form before the form was transformed to be on the Applied Math scale (M1C_LM3) and after the form was transformed to be on the Applied Math scale (M1C_LM3*). The table also provides the summary statistics for the scale scores for the Applied Math form (M2C_LM1). As indicated in Table 9.24, the means, standard deviations, and quantiles of the transformed Applied Mathematics form (M1C_LM3*) are very similar to those of the Applied Math form (M2_LM).

Table 9.24. Summary Statistics for Scale Scores for Applied Mathematics and Applied Math

Scale	Form	*N*	Mean (SD)	P10	P25	P50	P75	P90	P95
AM 1.0	M1C_LM3	821	78.12 (5.87)	70	75	78	82	86	87
AM 2.0	M1C_LM3*	821	78.60 (5.56)	71	76	79	82	86	87
AM 2.0	M2C_LM1	835	78.67 (5.46)	72	75	79	83	86	87

Note. M1C_LM3* was derived by applying the concordance table to transform the score in M1C_LM3 to be on the AM 2.0 scale.

Table 9.25 provides the summary statistics for the level scores for the Applied Mathematics form before the form was transformed to be on the Applied Math scale (M1C_LM3) and after the form was transformed to be on the Applied Math scale (M1C_LM3*). The table also provides the summary statistics for the level scores for the Applied Math form (M2C_LM1). The means and standard deviations of M1C_LM3* and M2C_LM1 are very similar except for some of the quantiles. The level cuts for the Applied Math assessment were developed based on a standard-setting study using the Mapmark method. (For greater detail on the standard-setting process, please see Chapter 8.)

Table 9.25. Summary Statistics for Level Scores of Applied Mathematics and Applied Math

Scale	Form	*N*	Mean (SD)	P10	P25	P50	P75	P90	P95
AM 1.0	M1C_LM3	821	4.34 (1.90)	<3	4	5	6	6	7
AM 2.0	M1C_LM3*	821	4.19 (1.90)	<3	4	4	5	7	7
AM 2.0	M2C_LM1	835	4.28 (1.88)	3	3	4	6	7	7

The results suggest that to compare the scores from the Applied Mathematics and Applied Math assessments and to understand the score relationships between the two assessments, the scale scores on the Applied Mathematics assessment need to be transformed to be on the Applied Math scale based on the concordance table. Test users need to be aware that the concordant scale scores do not always represent the test scores that examinees would achieve if they took the Applied Math assessment. Similarly, test users should be cautious when using only concordance scores to compare group performance averages or to analyze year-to-year performance trends. As noted at the beginning of Section 9.3, the concorded scores are not comparable or interchangeable across different forms, unlike the equated scores.

9.3.3 Linking Locating Information to the Graphic Literacy Score Scale

The Graphic Literacy assessment was developed based on redesigned test specifications from the Locating Information assessment (see Chapter 3 for the test specifications). To facilitate a smooth transition from Locating Information to Graphic Literacy, ACT conducted a linking study in spring 2017. The focus of the linking study was to develop a concordance between the Locating Information and the Graphic Literacy assessments. To find the scale score in Locating Information that corresponded to a particular scale score in Graphic Literacy, ACT identified the Locating Information score that had the same percentile rank as the Graphic Literacy score. For example, if the percentage of examinees who earned a scale score of 74 or below in Graphic Literacy was 20%, and the percentage of examinees who earned a scale score of 76 or below in Locating Information was 20%, then a Locating Information scale score of 76 corresponds to (i.e., concords with) a Graphic Literacy scale score of 74. This document summarizes the findings from the linking study to better illustrate the relationships between the two assessments and to assist users in appropriately interpreting the scores or score trends derived from the two assessments.

9.3.3.1 Study Design and Sample Representativeness

A total of 43 testing sites (10 high schools and 33 adult testing centers across 20 states) administered both forms G2C_LM1 (Graphic Literacy online) and G1C_LM3 (Locating Information online). A total of 1,684 examinees were given 55 minutes to complete one of the two linking forms; out of the 1,684 examinees, 854 examinees took G2C_LM1 and 830 examinees took G1C_LM3. In terms of demographic characteristics, the recruited sample was generally representative of the WorkKeys NCRC test population.

Because the Graphic Literacy assessment is a new assessment with new constructs and test specifications that differ from those of the Locating Information assessment, the resulting concorded scores are not interchangeable. When concorded forms have similar difficulty and measurement precision, the concordance results are likely to be stronger and more stable. A series of analyses was conducted to evaluate and compare the psychometric properties of the two assessments in terms of omit rates, testing times, scale score summary statistics, reliability, and standard error of measurement (SEM).

9.3.3.2 Comparison of Omit Rates and Testing Times Between Locating Information and Graphic Literacy

Figure 9.29 presents the omit rates for each item on both the Graphic Literacy and the Locating Information forms administered in the linking study. Figure 9.29 indicates that the omit rates were less than 10% for all items except the last item in the Locating Information form, G1C_LM3. In addition, as summarized in Table 9.26, the examinees spent an average of four minutes less on form G2C_LM1 than on form G1C_LM3. The shorter testing time 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 9.29. Comparison of Item Omit Rates Between Locating Information and Graphic Literacy

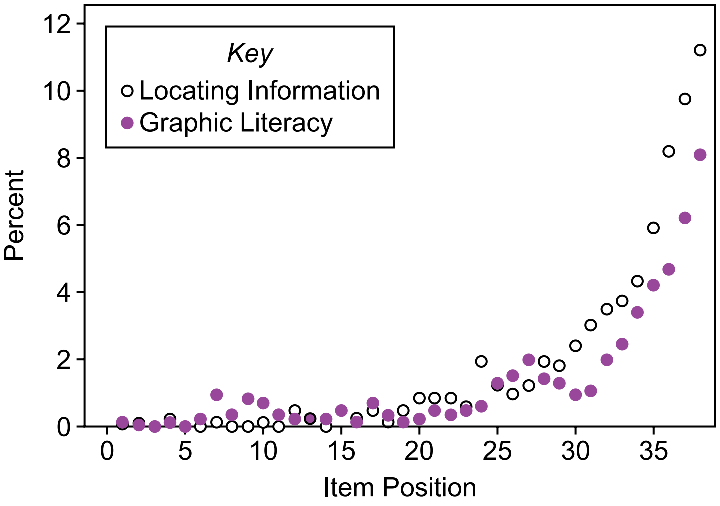


Table 9.26. Summary Statistics for Total Testing Times (in Minutes) for Locating Information and Graphic Literacy

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

9.3.3.3 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 recalibrated using the combined data from both the paper and online administrations to derive the conversion for the Graphic Literacy form (G2C_LM1). 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 9.27 and 9.28 provide the summary statistics for the raw scores and the scale scores for the linking study. Based on the average IRT-*b* statistics, the Graphic Literacy form, G2C_LM1, appears to be slightly more difficult than the Locating Information form, G1C_LM3.

Table 9.27. Test Summary Statistics for Locating Information and Graphic Literacy

Form	*P*	PBIS	IRT A	IRT B	IRT C
G1C_LM3	.604 (.256)	.381 (.121)	.973 (.282)	.297 (1.769)	.172 (.052)
G2C_LM1	.614 (.177)	.462 (.100)	1.022 (.261)	.435 (1.124)	.178 (.054)

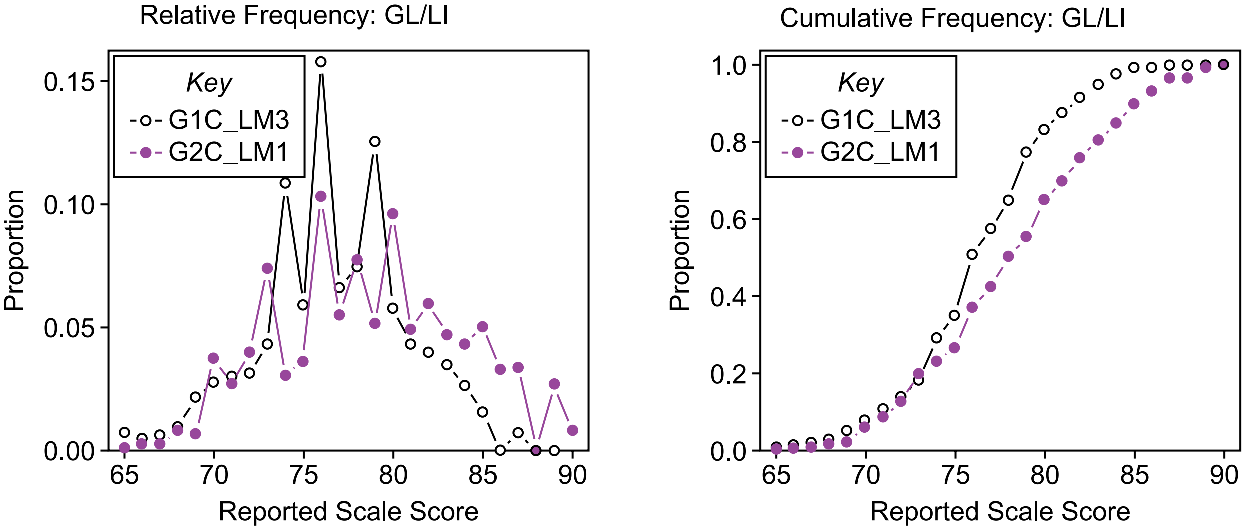
Note. P = *p*-value; PBIS = point-biss correlation. Standard deviations are in parentheses.

Table 9.28. Scale Score 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
G2C_LM1	854	78.64 (5.14)	70	72	75	78	82	86	87

Figure 9.30 presents the relative frequency distributions and cumulative frequency distributions for the Locating Information and Graphic Literacy forms. These plots suggest that the scale score distributions are different for the two assessments because significant modifications were made to the Graphic Literacy assessment.

Figure 9.30. Comparison of Relative Frequency Distributions (Left) and Cumulative Frequency Distributions (Right) Between Locating Information and Graphic Literacy



9.3.3.4 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 the Locating Information assessment (LI 1.0) to those from the Graphic Literacy assessment (GL 2.0). Concordance from Locating Information to Graphic Literacy was conducted using the equipercentile method with a smoothing (S) value of .05.

9.3.3.5 Evaluation of Locating Information Forms After Linking

Table 9.29 provides the summary statistics for the scale scores for the Locating Information form before the form was transformed to be on the Graphic Literacy scale (G1C_LM3) and after the form was transformed to be on the Graphic Literacy scale (G1C_LM3*). The table also provides the summary statistics for the scale scores for the Graphic Literacy form (G2C_LM1). It can be observed that the means, standard deviations, and quantiles of the transformed Locating Information form (G1C_LM3*) are very similar to those of the Graphic Literacy form (G2C_LM1).

Table 9.29. Summary Statistics for Scale Scores for Locating Information and Graphic Literacy

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	G2C_LM1	854	78.64 (5.14)	72	75	78	82	86	87

Note. G1C_LM3* was derived by applying the concordance table to transform the score in G1C_LM3 to be on the GL 2.0 scale.

Table 9.30 provides the summary statistics for the level scores for the Locating Information form before the form was transformed to be on the Graphic Literacy scale (G1C_LM3) and after the form was transformed to be on the Graphic Literacy scale (G1C_LM3*). The table also provides the summary statistics for the level scores for the Graphic Literacy form (G2C_LM1). The means and standard deviations of G1C_LM3* and G2C_LM1 are very similar except for the P90 quantile. The level cuts for the Graphic Literacy assessment were developed based on a standard-setting study using the Mapmark method. (For greater detail on the standard-setting process, please see Chapter 8.)

Table 9.30. Summary Statistics for Level Scores of Locating Information and Graphic Literacy

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	G2C_LM1	854	4.46 (1.83)	3	3	5	6	7	7

The results suggest that 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 be transformed to be on the Graphic Literacy scale based on the concordance table. Test users need to be aware that the concordant scale scores do not always represent the test scores that examinees would achieve if they took the Graphic Literacy assessment. Similarly, test users should be cautious when using only concordance scores to compare group performance averages or to analyze year-to-year performance trends. As noted at the beginning of Section 9.3, the concorded scores are not comparable or interchangeable across different forms, unlike the equated scores.

Chapter 10: Reliability and Measurement Error

10.1 Overview

This chapter reports the reliability evidence of the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessments. Reliability and measurement error are fundamental for evaluating the psychometric qualities of an assessment to substantiate the assessment claims defined in Chapter 1. As the *Standards for Educational and Psychological Testing* (hereafter, *Standards*) state, "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 examinees are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual examinee; that is, reliability is 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 NCRC foundational skill 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 the examinees are classified based on 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 the raw and scale scores based on classical test theory, (b) reliability coefficients of the level scores based on generalizability theory, and (c) classification consistency of the level scores.

10.2 Reliability Coefficients and Standard Error of Measurement (SEM)

Reliability coefficients quantify the consistency level of test scores. The reliability values typically range from zero to one, with the values near one indicating high consistency and the values near zero indicating little or no consistency. In a single test administration, the internal consistency reliability, 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 the 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 scores.

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 estimated scale score SEM and s_t^2 is the sample variance of the observed scale scores. The scale score SEMs were estimated using a four-parameter beta compound binomial model (Kolen et al., 1992). If the distribution of measurement error is approximated by a normal distribution, the true scale scores of about two-thirds of the examinee group are within plus or minus one SEM of the reported scale scores.

10.2.1 Workplace Documents

Table 10.1 presents the coefficient alpha (i.e., Cronbach's alpha) and the SEM for the Workplace Documents assessment for both the raw scores and scale scores. The reliability and SEM estimates are based on the sample used for the scaling study described in Chapter 8. The sample included 1,136 examinees after data cleaning. For reliable test score interpretations, a minimum reliability value of .80 is required. As shown in Table 10.1, the reliability estimates for both the raw and scale scores exceed this .80 threshold. (The corresponding plots of the conditional standard error of measurement [CSEM] of the raw scores and scale scores are presented in Chapter 8.)

Table 10.1. Reliability Estimates and Standard Error of Measurement for Form W2C_S1

		Raw Score		Scale Score	
Form	*N*	Cronbach's Alpha	SEM	Cronbach's Alpha	SEM
W2C_S1	1,136	.89	2.21	.90	1.70

10.2.2 Applied Math

Table 10.2 presents the coefficient alpha (i.e., Cronbach's alpha) and the SEM for the Applied Math assessment for both the raw scores and scale scores. The reliability and SEM estimates are based on the sample used for the scaling study described in Chapter 8. The sample included 1,185 examinees after data cleaning. For reliable test score interpretations, a minimum value of .80 is required. The reliability estimates for both the raw and scale scores exceed this .80 threshold. (The corresponding plots of the CSEM of the raw scores and scale scores are presented in Chapter 8.)

Table 10.2. Reliability Estimates and Standard Error of Measurement for Applied Math Form M2C_S1

		Raw Score		Scale Score	
Form	*N*	Cronbach's Alpha	SEM	Cronbach's Alpha	SEM
M2C_S1	1,185	.88	2.16	.89	1.61

10.2.3 Graphic Literacy

Table 10.3 presents the coefficient alpha (i.e., Cronbach's alpha) and the SEM for the Graphic Literacy assessment for both the raw scores and scale scores. The reliability and SEM estimates are based on the sample used for the scaling study described in Chapter 8. The sample included 1,170 examinees after data cleaning. For reliable test score interpretations, a minimum value of .80 is required. The reliability estimates for both the raw and scale scores exceed the .80 threshold. (The corresponding plots of the CSEM of the raw scores and scale scores are presented in Chapter 8.)

Table 10.3. Reliability Estimates and Standard Error of Measurement for Graphic Literacy Form G2C_S1

		Raw Score		Scale Score	
Form	*N*	Cronbach's Alpha	SEM	Cronbach's Alpha	SEM
G2C_S1	1,170	.85	2.34	.85	1.71

10.3 Generalizability Theory

Reliability based on generalizability theory was also investigated. The generalizability theory provides a broad conceptual and statistical framework for evaluating measurement precision (Cronbach et al., 1972). The 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 from a stratified test using a table of specifications. In the WorkKeys NCRC forms, items are nested (stratified) within specific levels of difficulty: Levels 3 to 7. A mixed model of *persons x (items:strata)* or *p x (i:h)* from a multivariate perspective was used. The results for Workplace Documents, Applied Math, and Graphic Literacy are presented in Tables 10.4, 10.5, and 10.6, respectively.

In the tables, the following results can be observed for all three WorkKeys NCRC Assessments:

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

- The variability of item difficulty, $\hat{\sigma}^2(i)$, is small, suggesting that difficulty is similar across items within each level.

- The interactions of person-by-item, $\hat{\sigma}^2(pi)$, are greater for the items at Levels 5, 6, and 7 than the items at Levels 3 and 4, indicating that examinees' performance is less consistent across the items at Levels 5, 6, and 7 than at Levels 3 and 4.

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

- The estimated effective weights which indicate the relative contributions of the items in each level to the total variance are higher for Levels 4, 5, and 6 than for Levels 3 and 7. The results, which are related to the numbers of items in each level, suggest that moderately difficult items are more heavily weighted in forming the total scores than the other items on the test.

The following lists the results that are different depending on the assessment:

- Reliability-like coefficients for norm-reference decisions, $E\hat{\rho}^2$, and for criterion-reference decisions $\hat{\phi}$:
  - ◊ **Workplace Documents:** .42 or higher at each level with Level 7 having the lowest value
  - ◊ **Applied Math:** .52 or higher at each level with Level 7 having the lowest value
  - ◊ **Graphic Literacy:** .43 or higher at each level with Level 3 having the lowest value
- Total scores:
  - ◊ **Workplace Documents:** The reliability-like coefficients for both the rank ordering of examinees and the judging performance levels of examinees are both equal to .90.
  - ◊ **Applied Math:** The reliability-like coefficients for both the rank ordering of examinees and the judging performance levels of examinees are .89 and .88, respectively.
  - ◊ **Graphic Literacy:** The reliability-like coefficients for both the rank ordering of examinees and the judging performance levels of examinees are both equal to .83.

Table 10.4. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Workplace Documents Form W2C_S1

Level	I	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(\delta)$	$\hat{\sigma}^2(\Delta)$	$E\hat{\rho}^2$	$\hat{\phi}$	Effective Weight
3	4	.045	.002	.104	.026	.027	.63	.63	.11
4	8	.078	.005	.117	.015	.015	.84	.83	.33
5	7	.059	.002	.186	.027	.027	.69	.69	.26
6	6	.056	.018	.178	.030	.033	.65	.63	.20
7	5	.027	.000	.185	.037	.037	.43	.42	.10

Note. I = the number of items.

Table 10.5. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Applied Math Form M2C_S1

Level	I	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(\delta)$	$\hat{\sigma}^2(\Delta)$	$E\hat{\rho}^2$	$\hat{\phi}$	Effective Weight
3	6	.027	.000	.010	.010	.010	.73	.73	.12
4	6	.061	.001	.022	.022	.023	.73	.73	.24
5	7	.059	.002	.025	.025	.027	.70	.68	.28
6	6	.058	.001	.031	.031	.032	.65	.64	.23
7	6	.028	.001	.025	.025	.026	.52	.52	.13

Note. I = the number of items.

Table 10.6. Estimated Variance Components, Error Variances, and Generalizability Coefficients at Each Level for Graphic Literacy Form G2C_S1

Level	I	$\hat{\sigma}^2(p)$	$\hat{\sigma}^2(i)$	$\hat{\sigma}^2(pi)$	$\hat{\sigma}^2(\delta)$	$\hat{\sigma}^2(\Delta)$	$E\hat{\rho}^2$	$\hat{\phi}$	Effective Weight
3	4	.020	.008	.098	.024	.026	.45	.43	.08
4	7	.028	.011	.158	.023	.024	.55	.54	.20
5	9	.041	.019	.169	.019	.021	.69	.66	.32
6	7	.040	.003	.207	.030	.030	.57	.57	.24
7	5	.040	.010	.175	.035	.037	.53	.52	.16

Note. I = the number of items.

10.4 Classification Consistency of Level Scores

The *Standards* (AERA et al., 2014, p. 46 as Standard 2.16) recommend that test publishers provide information about the percentage of examinees who would be classified in the same way on classification tests if the examinees took a test twice using alternate forms. Classification consistency ranges from 0% to 100%, with values near 100 indicating greater 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.

10.4.1 Workplace Documents

Table 10.7 displays the classification consistency indices for the Workplace Documents form W2C_S1 data, which were computed using the IRT methodology (Schulz et al., 1997, 1999). The second row of the table that is labeled “Exact” shows the percentages of the examinees who would receive the same level score on two parallel forms. For example, if an examinee took two parallel forms of a test and scored at Level 3 on both forms, this would be a case of exact agreement. For Workplace Documents form W2C_S1, the estimated exact agreement is 56%. The remaining rows show the consistency of aggregated classifications at each level. To compute the aggregated classification consistency at each level, examinees were classified into two groups: the group that scored below that specific level (analogous to the “fail” group) and the group that scored at or higher than that specific level (analogous to the “pass” group). Using this two-group classification, the classification consistency at each level was computed. For example, to be able to achieve a consistent classification of examinees who scored at Level 4 on one testing occasion, examinees needed to score at Level 4 and above on another testing occasion. In this study, the aggregated classification consistency of the level scores (agreement index *p*) is estimated to be 87% 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 the skill levels in an examinee sample. For example, in the W2C_S1 data, the mean of the examinee sample is slightly above the Level 4 theta cutoff, suggesting that the true skill of a relatively large proportion of these examinees was close to the Level 4 theta cutoff. Generally, examinees are more likely to be misclassified because of measurement error when their true skill is closer to the cutoff.

Table 10.7. Estimated Classification Consistency Indices for Level Scores for Form W2C_S1

Level	p	κ
Exact	56%	45%
3	92%	71%
4	88%	75%
5	87%	69%
6	90%	63%
7	96%	52%

In summary, the reliability and classification consistency findings in W2C_S1 are deemed acceptable based on the available field study data presented in Chapter 8. The updated test score reliability and classification consistency based on the recent data are available in Chapter 13.

10.4.2 Applied Math

Table 10.8 provides the classification consistency indices for the Applied Math form M2C_S1 data, which were computed using the IRT methodology (Schulz et al., 1997, 1999). The second row of the table that is labeled "Exact" shows the percentages of examinees who would receive the same level score on two parallel forms. For example, if an examinee took two parallel forms of a test and scored at Level 3 on both forms, this would be a case of exact agreement. For Applied Math form M2C_S1, the estimated exact agreement is 57%. The remaining rows show the consistency of aggregated classifications at each level. To compute the aggregated classification consistency at each level, examinees were classified into two groups: the group that scored below that specific level (analogous to the "fail" group) and the group that scored at or higher than that specific level (analogous to the "pass" group). Using this two-group classification, the classification consistency at each level was computed. For example, to be able to achieve a consistent classification of examinees who scored at Level 4 on one testing occasion, the examinees needed to score at Level 4 and above on another testing occasion. In this study, the aggregated classification consistency of the level scores (agreement index *p*) is estimated to be 87% 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 the skill levels in an examinee sample. For example, in the M2C_S1 data, the mean of the examinee sample is between the Level 4 and Level 5 theta cutoffs, suggesting that the true skill of a relatively large proportion of these examinees was close to the two θ cutoffs. Generally, examinees are more likely to be misclassified because of measurement error when their true skill is closer to the cutoff.

Table 10.8. Estimated Classification Consistency Indices for Level Scores for Form M2C_S1

Level	*p*	κ
Exact	57%	45%
3	94%	53%
4	87%	69%
5	87%	72%
6	91%	68%
7	96%	65%

In summary, the reliability and classification consistency findings in M2C_S1 are deemed acceptable based on the available field study data presented in Chapter 8. The updated test score reliability and classification consistency based on the recent data are available in Chapter 13.

10.4.3 Graphic Literacy

Table 10.9 provides the classification consistency indices for Graphic Literacy form G2C_S1 data, which were computed using the IRT methodology (Schulz et al., 1997, 1999). The second row of the table that is labeled "Exact" shows the percentages of examinees who would receive the same level score on two parallel forms. For example, if an examinee took two parallel forms of a test and scored 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%. The remaining rows show the consistency of aggregated classifications at each level. To compute the aggregated classification consistency at each level, examinees were classified into two groups: the group that scored below that specific level (analogous to the "fail" group) and the group that scored at or higher than that specific level (analogous to the "pass" group). Using this two-group classification, the classification consistency at each level was computed. For example, to be able to achieve a consistent classification of examinees who scored at Level 4 on one testing occasion, examinees needed to score at Level 4 and above on another testing occasion. In this study, the aggregated classification consistency of the level scores is estimated to be 84% 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 the skill levels in an examinee sample. For example, in the G2C_S1 data, the mean of the examinee sample is at the Level 5 theta cutoff, suggesting that the true skill of a relatively large proportion of these examinees was close to the Level 5 theta cutoff. Generally, examinees are more likely to be misclassified because of measurement error when their true skill is closer to the cutoff.

Table 10.9. 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 in G2C_S1 are deemed acceptable based on the available field study data presented in Chapter 8. The updated test score reliability and classification consistency based on the recent data are available in Chapter 13.

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® National Career Readiness Certificate® (NCRC®) assessments teams incorporated an approach of gathering evidence 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) maintains 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, assessment 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 be altogether inappropriate and problematic for another 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 the WorkKeys NCRC assessment test score interpretations and usage, the WorkKeys NCRC assessments teams 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, the WorkKeys NCRC assessments teams also used 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.

The WorkKeys NCRC teams implemented a process that began with articulating the purpose of the assessment and its associated claims, as outlined in Chapter 1, and culminated with the collection of data from various sources to evaluate the validity use argument, presented in this chapter. The purpose of this validity chapter is to present each WorkKeys NCRC assessment claim and then provide evidence to evaluate the appropriateness of the proposed interpretations and uses.

11.2 WorkKeys Assessments Claims

Drawing on its understanding of the skills gap and skills-based hiring practices, the design team developed three primary claims for the WorkKeys NCRC assessments.

> **Claim #1:** U.S. examinees of high school or workforce age whose scores reach at least a given level on the WorkKeys NCRC assessments are more likely to perform successfully in more U.S. jobs (in the ACT job taxonomy classified in the ACT® WorkKeys® JobPro® database) and at higher levels than examinees whose scores do not reach that level.

Claim #1 Assumptions:

1. Each of the three WorkKeys NCRC assessments is a component of foundational workplace skills; these skills are required for success in many jobs (based on ACT's job profile database).

2. ACT has developed a professionally valid and appropriate definition of the WorkKeys NCRC assessments construct.

3. The WorkKeys NCRC assessments elicit observable evidence of the construct and provide reliable and interpretable scores that reflect the construct.

4. ACT has defined workplace-appropriate performance level descriptors (PLDs), and ACT has established standards (e.g., cut points [cut scores]) aligned to the PLDs.

5. Cut points (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 WorkKeys NCRC assessments are positively related to measures of employee performance, including productivity and turnover rates.

8. Examinees who score well on the WorkKeys NCRC assessments are more likely to receive higher performance ratings and are more likely to have greater job success (defined as job retention and performance evaluations) than lower-scoring examinees.

Claim #2: U.S. companies that hire U.S. examinees of high school or workforce age whose scores reach at least a given level on the WorkKeys NCRC assessments 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. These include the first seven Claim #1 assumptions.

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 more effective and productive business.

Claim #3: U.S. companies that hire U.S. examinees of high school or workforce age whose WorkKeys NCRC assessments scores 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. These include the first seven Claim #1 assumptions.

2. Employees with higher foundational skill levels are less likely to be terminated in the first 6 months of employment than employees with lower foundational skill levels.

3. Employees with higher foundational skill levels are less likely to quit in the first 6 months of employment than employees with lower foundational skill levels.

4. Businesses that use scores from the WorkKeys NCRC assessments as part of their hiring process will tend to experience less turnover than businesses that do not use the WorkKeys NCRC assessment as part of their hiring process.

These three WorkKeys NCRC assessments claims address questions around examinee job success, improving worker productivity, reducing employee turnover rates, and improving business productivity. 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 and career advancement.

The *Standards for Educational and Psychological Testing* (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.

In Chapter 1, we noted that all three primary claims are dependent on the validity of five initial assumptions and then presented data and analysis related to the five assumptions. The following table provides the initial assumptions and the chapter number(s) where evidence supporting the assumption can be found.

Table 11.1. Initial Assumptions and Chapter Numbers

Evidence Location	Initial Assumptions
Chapter 1	1. The skills required in reading workplace documents, in applied mathematics, and in graphic literacy are foundational workplace skills and are required for success in many jobs.
Chapters 2 and 3	2. ACT has developed a valid and appropriate construct definition of workplace documents, applied mathematics, and graphic literacy.
Chapter 9	3. ACT's WorkKeys NCRC assessments provide reliable and interpretable scores measuring the construct.
Chapters 2 and 3	4. ACT has defined appropriate WorkKeys NCRC assessment PLDs, and ACT has established standards aligned to the PLDs.
Chapters 9 and 10	5. The cut scores used to delineate each performance level have sufficient classification accuracy.

11.3 Primary Claims and Relevant Findings

The purpose of the WorkKeys Assessment program is to help build a high-performance workforce by connecting job skills, curriculum, and testing in a manner that benefits both employers and employees. The WorkKeys NCRC assists educators in identifying skill gaps between student skills and employment needs identified using ACT WorkKeys Job Profiling so that educators may better address the gaps and thereby improve students' employment prospects.

The three primary claims articulate how scores from each WorkKeys NCRC assessment may provide actionable information to examinees, employers, educators, and workforce development officials to make these connections. The claims differ regarding 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 each WorkKeys NCRC assessment are related to workplace success. In other words, an examinee who scores at a prescribed level (as defined through data from a job profile) will have a greater probability of achieving success in a corresponding job (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 each WorkKeys NCRC 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 each WorkKeys NCRC assessment may help an individual in the labor market.

The focus of the second and third claims is the employer or business. Scores on each WorkKeys NCRC assessment are related to workplace success in ways that will result in improved business productivity and efficiency. Claim #2 states that if a business first determines the WorkKeys NCRC assessment scores that are required for specific jobs through a job analysis or job profile, and the business subsequently hires people who achieved those scores, the productivity gains provided by the new employees will 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 will 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 content-related and construct evidence provided in Section 11.4 and Chapter 1 Section 1.5. Additionally, they can be supported through the analysis of outcome data. Claim #3 requires the analysis of employee turnover rates to be plausible.

11.3.1 Evidence Based on Test Content

One source of evidence used to establish the validity of test score interpretations and uses is evidence based on test content (AERA et al., 2014). Content evidence is often the first type of evidence used to support employment selection practices. The *Uniform Guidelines on Employee Selection Procedures (1978)* (Equal Employment Opportunity Commission [EEOC] et al., 1978), the *Standards for Educational and Psychological Testing* (AERA et al., 2014), and the *Principles for the Validation and Use of Personnel Selection Procedures* (SIOP, 2018) all describe the need to demonstrate that knowledge and skills in employment measures should be demonstrably linked to work behaviors and job tasks.

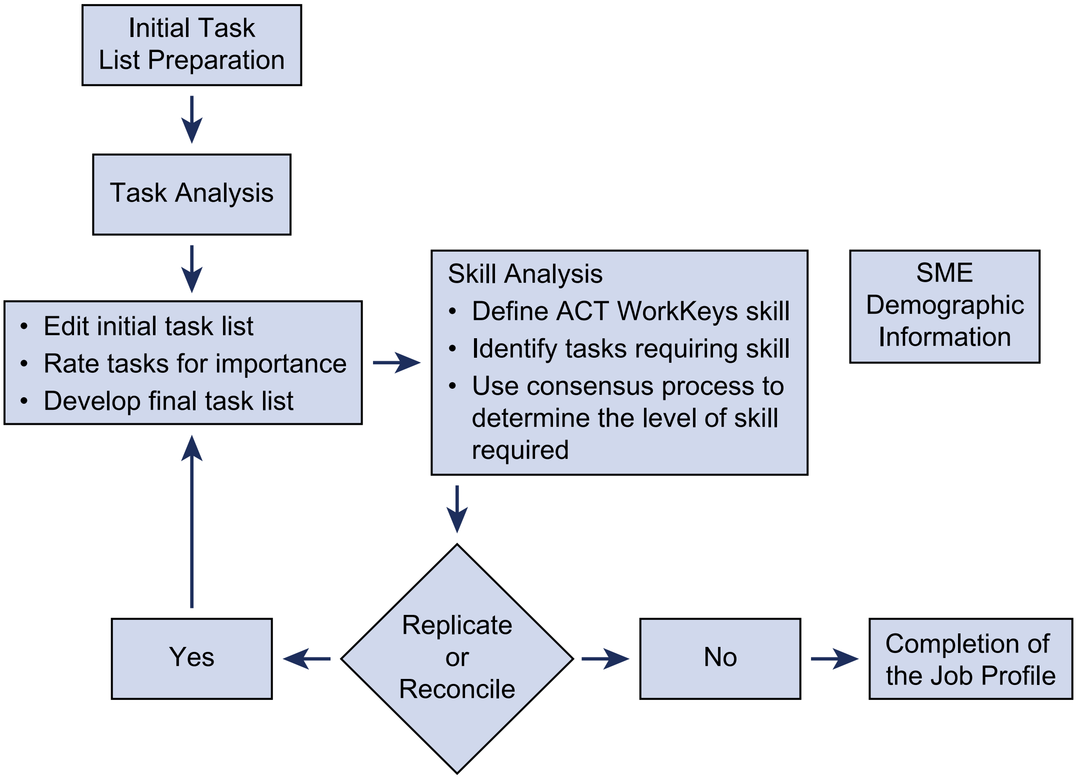
Both the *Standards for Educational and Psychological Testing* (AERA et al., 2014) and the *Principles for the Validation and Use of Personnel Selection Procedures* (SIOP, 2018) suggest that expert judgment can be used to determine the importance and criticality of job tasks and to relate such tasks to the content domain of a measure. This evaluation is commonly conducted through a job analysis that identifies the tasks required for performance on a job; it is subsequently used to develop the content blueprint and items to ensure content validity (Cascio, 1982; Dunnette & Hough, 1990). The WorkKeys NCRC assessments—Workplace Documents, Applied Math, and Graphic Literacy—were designed to assess foundational skills and proficiency levels associated with many jobs. As such, the content-related validity evidence for the assessments was originally established across numerous jobs by the subject matter experts (SMEs) who aligned the skills and PLDs in the WorkKeys NCRC assessments 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 suite of assessments. Creating a job profile 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.

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

The process begins with a task analysis where the group of SMEs edit a task list previously generated by the profiler. The goal of the task analysis is to develop a task list that accurately represents the job at an organization. The SMEs then rate each task in terms of its importance. Figure 11.1 details the steps in the job profiling procedure where tasks and skills are identified leading to the completion of the job profile.

Figure 11.1. Job Profile Process Designed to Align Job Tasks to Skill Levels


Initial Task List Preparation
Task Analysis
• Edit initial task list
• Rate tasks for importance
• Develop final task list
Skill Analysis
• Define ACT WorkKeys skill
• Identify tasks requiring skill
• Use consensus process to determine the level of skill required
SME Demographic Information
Yes
Replicate or Reconcile
No
Completion of the Job Profile

Equally important is the skill analysis where the SMEs review each skill measured by the WorkKeys Assessments. Once the SMEs understand the definition of each skill and have determined its relevancy to the job, the SMEs independently identify the important tasks on the final task list that require the skill. The SMEs 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 profiler presents detailed descriptions of the WorkKeys skill levels to the SMEs, which includes examples of problems or situations employees deal with at each level. SMEs use successive approximations to determine the skill level required for job entry and for effective performance. 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 tasks. 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 SMEs determine that the skills are about the same as the level they are reviewing, the SMEs are still shown the next higher level before confirming their agreement between a skill and its designated level.

No decision is reached until the SMEs have considered a range of skill levels; that is, 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 22,000 job profiles.

11.3.2 Evidence Based on Relationships to Work-Related Variables

The correlations between ACT WorkKeys Assessments and job performance ratings provide criterion-related evidence for the validity of using ACT WorkKeys Assessments in relation to a specific job. Several studies have been conducted across a range of organizations which examine the relationship between the ACT WorkKeys cognitive test scores and employee job performance ratings. Sample sizes and correlations vary across studies of a wide spectrum of occupations across the assessments. Early ACT WorkKeys criterion validity studies relied on measures of job performance based on job- and company-specific task lists developed during the job profiling process. Studies conducted since 2006 have utilized the ACT Supervisor Survey or ACT WorkKeys Appraise, both of which rely on more generalized categories of job performance based on literature about common dimensions of job performance (ACT, 2018).

As of January 2015, there have been numerous criterion validation studies conducted on the ACT WorkKeys Assessments since 1993. A breakout of the number of unique studies by assessment, including the ranges of sample sizes and correlations, is provided in Table 11.2 (ACT, 2018). The table also presents the relationship of composite scores from Locating Information (LI), Applied Mathematics (AM), and Reading for Information (RFI) with different outcome measures (LeFebvre, 2016). The results of the studies show a modest relationship with supervisor ratings of overall job performance; a positive relationship with education outcomes such as grade point average (GPA), course grades, and postsecondary persistence; and a positive relationship with reduction in safety incidents and customer complaints, absenteeism, and turnover. Research on the ACT NCRC shows a modest relationship with increase in earnings, employment attainment, and employment retention rates.

Table 11.2. WorkKeys NCRC Validation Studies

ACT WorkKeys Assessment	No. of Studies	*N* Size		Range of Validity Coefficients			Outcomes
		Min	Max	Min	Med	Max	
Applied Mathematics (AM)	1	2,162	2,162	0.21	0.21	0.21	Career Tech Course Grades
	1	1,246	1,246	0.28	0.28	0.28	Postsecondary GPA
	13	13	165	–0.23	0.12	0.41	Overall Job Performance—Supervisor Ratings
Locating Information (LI)	1	1,216	1,216	0.21	0.21	0.21	Career Tech Course Grades
	1	96	96	–0.33	–0.33	–0.33	HRIS Data—Turnover
	1	96	96	–0.22	–0.22	–0.22	HRIS Data—Absenteeism
	1	96	96	–0.11	–0.11	–0.11	HRIS Data—Safety Incidents
	1	96	96	–0.11	–0.11	–0.11	HRIS Data—Customer Complaints
	14	13	314	–0.51	0.16	0.32	Overall Job Performance—Supervisor Ratings
Reading for Information (RI)	1	2,223	2,223	0.22	0.22	0.22	Career Tech Course Grades
	1	1,251	1,251	0.25	0.25	0.25	Postsecondary GPA
	1	96	96	0.12	0.12	0.12	HRIS Data—Turnover
	1	96	96	–0.13	–0.13	–0.13	HRIS Data—Absenteeism
	1	96	96	–0.15	–0.15	–0.15	HRIS Data—Safety Incidents
	1	96	96	–0.24	–0.24	–0.24	HRIS Data—Customer Complaints
	16	10	314	–0.32	0.2	0.86	Overall Job Performance—Supervisor Ratings
Composite RI and AM	1	10,744	10,744	0.3	0.3	0.3	Postsecondary GPA
	1	277,631	277,631	0.23	0.23	0.23	College Persistence
Composite RI, LI, and AM	3	68	951	0.29	0.29	0.29	Overall Job Performance—Supervisor Ratings
	1	951	951	0.25	0.25	0.25	Career Tech Course Grades

Note. Of the many dimensions of job performance studied, only the overall job performance correlations are reported in this table for summary purposes. HRIS = Human Resource Information System data collected by employer.

Two additional third-party studies specifically analyzed scores on the WorkKeys Assessments or levels achieved on the WorkKeys NCRC and compared them to outcome measures, including job performance ratings and grades in career and technical education programs.

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 NCRC 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 employees requiring less training time and less close supervision.

Greene (2008) analyzed the use of the WorkKeys cognitive assessments in business and industry in North Carolina. Greene surveyed employers of small and large companies, focusing primarily on the use of the WorkKeys NCRC, and 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 rework 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 higher quality work overall.

There has been and will continue to be a need to conduct studies to support the use of the ACT WorkKeys program for pre-employment selection and other high-stakes purposes. The results of more recent studies are provided below.

Through the collaboration between the Missouri Department of Economic Development—Missouri Economic Research Information Center (MERIC) and ACT Research, ACT gathered WorkKeys performance data for Missouri examinees from 2012–2014, and MERIC merged those data with demographic information, unemployment insurance records, postsecondary degree attainment, and cumulative GPA. Quarterly wage records from the unemployment insurance database were provided 9 months before and 15 months after taking WorkKeys. The complete data set included 27,475 records. Overall, results from this study were consistent with the NCRC as a signal of the skills needed for success in the labor market and postsecondary education in Missouri. Individuals with higher NCRC levels were more likely to have wage records in the database, which indicates that people earning higher NCRC levels were more likely to be employed. The study also detected a significant positive relationship between WorkKeys performance and postsecondary cumulative GPA. Each increase in NCRC level was associated with an approximate 0.22 increase in cumulative GPA (Steedle & Lefebvre, 2018).

As part of a U.S. Department of Labor grant, Cincinnati State Technical and Community College identified evidence-based best practices in workforce development. Overall, results from the study support the use of WorkKeys products and services as tools for predicting and improving educational and labor market outcomes such as grades, program completion rates, employment rates, and post-enrollment wages (Steedle et al., 2017). More specifically, the results showed:

- 90% of students who completed the WorkKeys curriculum obtained an NCRC credential.
- Students who received support services, often including ACT training services, were more likely to complete their program than students who did not receive services (65% versus 35%).
- Of the participants in the sample who were unemployed before enrollment, 69% of those who received ACT training services became employed, compared with 57% of those who did not have ACT training services.
- In the same group of students, 70% of those who earned an NCRC became employed, compared with 57% of those who did not earn an NCRC.
- Participants who received ACT training services or obtained an NCRC experienced greater wage increases than those who did not participate or obtain an NCRC.

In 2018, ACT published a study showing that 2010 WorkKeys examinees who earned higher NCRC levels typically had higher incomes and increased their incomes over time at a faster rate. ACT recently conducted a study of 2011 WorkKeys examinees to replicate the prior study and to investigate income trend differences between high school and adult test takers. For this study, ACT delivered a representative sample of 50,000 WorkKeys examinees to Equifax—one of the nation's largest consumer credit reporting agencies—and Equifax matched the WorkKeys testers to income records from 2011 through 2016. Workers who earned higher NCRC levels in 2011 tended to earn higher incomes and increase their incomes more in the five years after testing. Specifically, the groups with Silver, Gold, or Platinum NCRC increased their median incomes within two years of taking WorkKeys, and this was true for both high school and adult examinees. Moreover, income increases for NCRC earners outpaced national trends by a wide margin. In all, these findings are consistent with the claim that performing well on WorkKeys and earning higher NCRC levels can help people secure higher incomes in the short and long term (Steedle & LeFebvre, 2018).

11.3.3 Evidence from Meta-Analysis Studies

The studies discussed in the previous section specifically analyzed scores on the WorkKeys Assessments or levels achieved on the WorkKeys NCRC and compared them to outcome measures, including job performance ratings, income, and grades. Other researchers 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 the measurement problems associated with outcome variables, researchers believed that validity coefficients varied a great deal from one job to the next. For the first 70 years of the 20th century, researchers evaluated employment selection methods by correlating scores on selection tests to measures of job performance. The researchers found that using the same tests for nearly identical jobs often resulted in quite different validity coefficients, and they concluded that the differences in validity coefficients stemmed from subtle differences in job requirements resulting in situation-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 Schmidt and Hunter's 1998 paper, Schmidt, Oh, and Shaffer (2016) evaluated 31 different methods of personnel selection, including cognitive ability testing, job interview ratings, and handwriting analysis. Schmidt, Oh, and Shaffer concluded that general cognitive ability was the "gold standard" of selection methods and thus 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). Schmidt and Sharf defined the Reading for Information assessment as a measure of reading skills that are highly relevant to job performance and learning and defined the Applied Mathematics assessment as a measure of quantitative reasoning skills that are highly relevant to job performance and learning. Schmidt and Sharf defined the Locating Information assessment as a measure of technical/problem-solving skills that are highly relevant to job performance and learning.

Combining Schmidt and Sharf's (2010) results with LeFebvre's summary reveals a median correlation of 0.29, which appears similar to correlations of the SAT and ACT to first-year college grades. Taking into account selection effects, range restriction, and the low reliability of outcome measures, the correlation of 0.29 is a conservative estimate. The disattenuated correlation is likely much greater (Sackett, Borneman, & Connelly, 2008).[6]

[6] Sackett, Borneman, and Connelly (2008), applying meta-analytic methods to address range restriction and low reliability of outcome measures, estimate that the disattenuated correlation of general cognitive ability with job performance is 0.47.

11.3.4 WorkKeys Assessments and Return on Investment

Hunter, Schmidt, and Judiesch (1990) published a groundbreaking 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 used 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 required 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 are 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 the outcomes compared pre- and post-WorkKeys NCRC implementation. Mayo 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, Mayo concluded that using the WorkKeys NCRC as part of the hiring process resulted in employers making a minimal investment for a very large return.

In 2011, E. & J. Gallo Winery partnered with ACT to determine the utility of the ACT WorkKeys Assessments based on organizational data (e.g., cost of system, number of employees selected, salary, employee tenure). The utility results indicated that ACT WorkKeys resulted in a 23.2% increase in employee productivity in task performance, a 22.1% increase in output due to increased employee safety, an 18.9% reduction in hiring needs due to increased performance, and a 19.3% reduction in hiring needs due to increased employee safety.

11.3.5 Educational Outcomes

LeFebvre (2016) reviewed studies that related WorkKeys NCRC assessment scores to post-secondary educational outcomes (see Table 11.2). In career and technical education programs, individuals who achieved higher scores on the assessments tended to have higher completion rates and earn higher grades in career and technical education programs and tended to have higher grade point averages in their post-secondary studies.

Schultz and Stern (2013) studied changes in examinee perceptions of career readiness following the administration of WorkKeys 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 As and Bs, and they regarded their skills as strong based on their WorkKeys NCRC scores, slightly more than 50% of the students did not meet the college or career readiness standards.

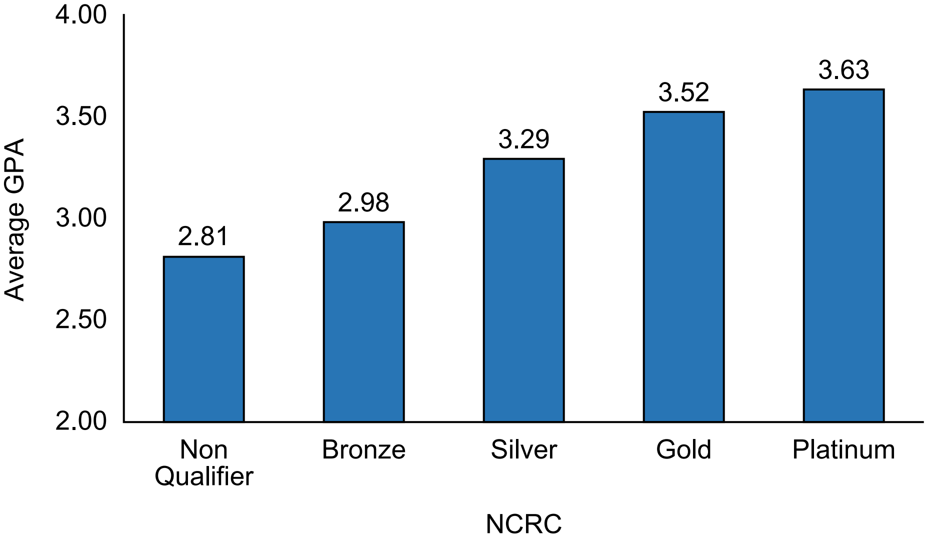
A study with GPS Education Partners (GPSEd), a community-based nonprofit organization and work-based learning intermediary offering a statewide manufacturing youth apprenticeship in Wisconsin, provided evidence that the WorkKeys NCRC is an indicator of skills needed for success in work-based learning. Analyses revealed that WorkKeys scores correlated with academic outcomes like reading ability level and math course-taking patterns. WorkKeys scores and WorkKeys NCRC Levels were also related to Manufacturing Skill Standards Council (MSSC) Certified Production Technician (CPT) certificate attainment and attendance in GPSEd classes and apprenticeships. In all, the results suggest that WorkKeys provides useful signals of the foundational workplace skills required for success in a work-based learning program. Moreover, average WorkKeys scores increased significantly from junior to senior year, which is consistent with the notion that students can improve their foundational workplace skills in a work-based learning program (Steedle & Hepburn, 2020).

A recent study examined the relationship between the WorkKeys NCRC assessments and first-year college GPA (Conway, 2022). The correlations between each WorkKeys NCRC assessment and the GPA are listed in Table 11.3. The sample size for the study was 964, and all correlations in Table 11.3 were statistically significant ($p < .01$). To aid interpretation, Figure 11.2 shows the average GPA at each NCRC level. The figure shows that the average GPA increases as a function of the NCRC level.

Table 11.3. Correlations Between WorkKeys Assessments and GPA

WorkKeys Assessment	Validity Coefficient
Applied Math	.40
Graphic Literacy	.40
Workplace Documents	.40
NCRC	.41

Figure 11.2. Average GPA by NCRC Distinction



11.3.6 WorkKeys Assessments at the State and Regional Levels

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, LeFebvre found that individuals who achieved higher levels experienced faster time to hire, earned higher wages, and stayed in their jobs longer.

11.4 Evaluation of Claims

The cited studies analyzed data from each WorkKeys Assessment, the WorkKeys NCRC, and general measures of cognitive ability. As mentioned earlier, the Workplace Documents, Applied Math, and Graphic Literacy assessments constitute the WorkKeys NCRC. Each WorkKeys NCRC assessment was designed to build on the information that ACT has collected over the past 25 years from the Reading for Information assessment, the Applied Mathematics assessment, and the original Locating Information assessment. To better reflect the current uses of written materials and of applied mathematics in the workforce, Workplace Documents and Applied Math contents were updated from the original Reading for Information and Applied Mathematics assessments, respectively. 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 Workplace Documents, Applied Math, and Graphic Literacy assessments met or exceeded the psychometric standards that were used to develop the earlier assessments of Reading for Information, Applied Mathematics, and Locating Information. For these reasons, data collected on Reading for Information, Applied Mathematics, and Locating Information can be used tentatively to evaluate the three claims. Additional types of evidence are always sought to bolster a validity argument, and there are opportunities to augment the evidence for the ACT NCRC and specific WorkKeys NCRC assessments. Validation is an ongoing process and often the joint responsibility of the test developer and organizations using assessments. ACT remains committed to providing multiple sources of evidence to support the interpretative arguments and intended uses for the WorkKeys NCRC tests and the ACT NCRC.

From the examinee perspective, based on the findings when score information from the WorkKeys Assessments and the WorkKeys NCRC are used as part of employment selection or for educational evaluation, it appears that individuals who achieved sufficient scores on the WorkKeys NCRC assessments tended to experience the following:

- reduction in time to hire (LeFebvre, 2016; Mayo, 2012)
- higher wages (Steedle & LeFebvre, 2018; 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 (Conway, 2022; Steedle & LeFebvre, 2018; Steedle et al., 2017; LeFebvre, 2016)
- information that provides insight useful in evaluating career readiness and career planning (Steedle & Hepburn, 2020; Schultz & Stern, 2013)

The findings from these studies provide evidence supporting Claim #1 that examinees who score at a given level of the WorkKeys NCRC 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 above, when score information from the WorkKeys Assessments and the WorkKeys NCRC were used as part of the employment selection process, it appears that businesses tended to have the following outcomes:

- higher levels of productivity (LeFebvre, 2016; Greene, 2008; Hunter et al., 1990)
- lower rates of re-work (Greene, 2008)
- lower turnover rates and 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 provide evidence supporting Claims #2 and #3 that businesses that use the WorkKeys NCRC assessment as part of the hiring process will experience increases in business productivity and reduced worker turnover rates.

11.5 Ongoing Validation

ACT continually collects and analyzes data related to the validation of its products. As outcome data continues to be collected and analyzed, ACT will publish the findings through research reports to supplement the WorkKeys Technical Manual. In collecting and analyzing the data, ACT is cognizant of the two main populations served by each WorkKeys NCRC 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 populations. Although specific details of the analyses are dependent on the available outcome data, ACT will analyze the relationships of scores on each WorkKeys NCRC 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.

Chapter 12: Test Fairness

This chapter contains evidence to address assessment fairness related to the ACT® WorkKeys® National Career Readiness Certificate® (NCRC®). This 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 of which requires evaluation: (a) fairness in treatment during the testing process, (b) fairness in access to the construct or constructs measured, (c) fairness as lack of measurement bias, and (d) fairness as validity of individual test score interpretations for the intended uses.

Whenever tests are used as part of a 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 WorkKeys NCRC assessments is not limited to this chapter and is drawn from other chapters in this technical manual. Furthermore, ACT continually collects and analyzes assessment data. As additional data are collected and analyzed, ACT will continually issue reports related to the fairness of each assessment's 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 examinees' ability 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 WorkKeys NCRC assessments incorporate principles of Universal Design (Center for Applied Special Technologies, 2011) and Evidence-Centered Design (Mislevy et al., 2004) to assist in ensuring fairness to all examinees. ACT developed and documented standardized procedures for training test center staff for test administration. ACT articulated room and equipment standards to support standardized and fair conditions for all examinees. ACT further defined protocols for handling 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 to allow ACT to decide whether the event compromised validity. WorkKeys has implemented these procedures to attain fairness for all examinees in the administration of the WorkKeys NCRC assessments. (See Chapter 4 for a comprehensive review of test administration procedures.)

The WorkKeys NCRC assessments are administered to examinees in both paper and online formats. To provide evidence of the fairness of scores across both administration formats, ACT conducted a mode comparability study. ACT evaluated the mode effects at the item and score levels. Through the analysis, ACT concluded that mode effects on examinee responses and scores were negligible. (For greater detail regarding the mode analysis, see Chapter 9.)

Although ACT recognizes that standardizing procedures for test administrations is critically important to ensure that all examinees have an equal opportunity to demonstrate their abilities in the construct being measured, ACT also recognizes that flexibility is required to achieve true fairness. When the standardized administration procedures hinder examinees from demonstrating their abilities in the construct being measured, and when the examinees provide proper documentation, modifications to the standardized procedures are considered fair and appropriate.

12.3 Fairness in Access to the Measured Construct

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, examinees with a visual impairment may not be able to appropriately answer questions on a WorkKeys NCRC assessment because they cannot clearly see the test materials. In such cases, not being able to read the test materials creates construct-irrelevant variance. A second example involves an examinee who was diagnosed with mild autism spectrum disorder. This examinee may require additional time to complete the test and a special testing location that is free from distractions. ACT provides a variety of accessibility options for examinees that are designed to provide access to the intended test construct that is measured without violating or interfering with the construct being measured or giving the examinee an unfair advantage. (See Chapter 5 for a comprehensive review of test accessibility features.)

The supports provided on the WorkKeys NCRC assessments 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: (a) embedded tools, (b) open access tools, and (c) accommodations. Embedded tools are commonly used by many people, are available to all examinees, and do not need to be requested in advance. Open access tools are used by fewer people and are 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 or equipment that can be used effectively and securely only with expertise, special training, and/or extensive monitoring.

All accessibility supports permitted for the WorkKeys NCRC assessments 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 complete list of the available accommodation features available in paper and online administrations.)

12.4 Fairness as a 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 et al., 1985, p. 79). Measurement fairness requires that examinees with equal standing (i.e., equal performance levels) on the construct being measured obtain, on average, the same scores on the assessment regardless of group membership (Sackett et al., 2008). Consequently, measurement bias occurs when score interpretations are differentially valid across groups (i.e., when they are valid for some examinee groups but invalid for other examinee groups). To investigate the potential for measurement bias, ACT evaluates the internal structure of the WorkKeys NCRC assessments by evaluating the invariance of the items and the overall assessment depending on examinee groups.

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 in one demographic group respond to differently from individuals in 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). DIF is statistical evidence of item bias; however, statistical evidence alone is not sufficient evidence of measurement bias. ACT WorkKeys has established a process for conducting DIF analyses that is followed by external reviews of flagged items to determine whether the flagged items include measurement bias.

In conducting DIF analyses, ACT compares item responses in two groups of examinees: a focal group and a reference group. The focal group is the group of primary interest and includes protected classes under federal employment antidiscrimination laws. The reference group serves as the basis for group comparison.

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

Table 12.1. WorkKeys DIF Evaluations—Group Comparisons

Focal group	Reference group
Female	Male
Black	White non-Latinx
Latinx	White non-Latinx

An item is flagged as showing DIF when one group has a higher probability of answering an item correctly than the other group. Because groups may differ in ability level, the DIF analysis matches examinees based on ability. For the WorkKeys DIF studies, ACT matches examinees by their total test scores.

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

Table 12.2. WorkKeys NCRC DIF Classification Rules

Group	Rules
Group A	MHD not significantly different from 0 (based on chi-square test, alpha = .05) or $\lvert MHD \rvert < 1.0$
Group B	MHD significantly different from 0 (based on chi-square test, alpha = .05) and $\lvert MHD \rvert \geq 1.0$ and < 1.5; or MHD not significantly different from 0 and $\lvert MHD \rvert \geq 1.0$
Group C	MHD significantly different from 0 (based on chi-square test, alpha = .05) and $\lvert MHD \rvert \geq 1.5$

Note. Classification rules adopted from National Assessment of Educational Progress guidelines (Allen et al., 1999)

After ACT has analyzed the DIF statistics and classified items into Groups A, B, or C, the content specialists evaluate all flagged items (Groups B and C) for possible bias. Item bias occurs when an aspect of the item content places a group at a disadvantage. To determine if a flagged item contains bias, the content team reviews the item's content internally. In addition to being reviewed by the content team, any flagged item is potentially also reviewed by external evaluators who have training and expertise in cultural anthropology or multicultural education.

The review includes evaluating (a) the item's vocabulary or use of numbers and symbols, (b) the knowledge needed to correctly answer the item, (c) the examinees' accessibility to knowledge, (d) the cognitive processes required, and (e) possible examinee misinterpretations that might occur because of differences in life experiences or learning opportunities. To assist in this review, ACT has identified five questions for reviewers to use:

1. **Status:** Are members of a particular group depicted in situations that do not involve authority or leadership?
2. **Stereotype:** Are members of a particular group portrayed as uniformly having certain aptitudes, interests, occupations, or personality characteristics?
3. **Familiarity:** Is there a 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?
4. **Offensive Choice of Words:** Has a demeaning or gender-biased label been applied where a neutral term could be substituted?
5. **Other:** Are there any other indications of bias?

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

1. Maintain and continue to use the item as it is currently constructed.
2. Send the item back to the content team for revision; the evaluator identifies what aspect of the item should be revised.
3. Remove the item from the item pool.

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

12.4.1.1 Workplace Documents

During the second step in the field-testing process, ACT administered the two forms of the Workplace Documents assessment to 2,317 field test participants. Forty testing sites in 22 states participated (13 high schools and 27 adult testing centers). Approximately 59% of the examinees were high school students and 41% were adults. Prior to the test administration, ACT instructed 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 Workplace Documents assessment.

Table 12.3. Workplace Documents—Number and Percent of Field Test Participants by Demographic Group

Demographic characteristic	Form WS1		Form WS2		Total	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Total participants	1,136	49.0	1,181	51.0	2,317	100
Male	504	44.4	511	43.3	1,015	43.8
Female	596	52.5	639	54.1	1,235	53.3
Black	197	17.3	218	18.5	415	17.9
American Indian/Alaska Native	24	2.1	21	1.8	45	1.9
Asian	9	.8	6	.5	15	.6
Latinx	71	6.3	81	6.9	152	6.6
Native Hawaiian/Pacific Islander	2	.2	1	.1	3	.1
Two or more races/ethnicities	93	8.2	199	16.9	292	12.6
White non-Latinx	691	60.8	720	61.0	1,411	60.9
Prefer not to respond	49	4.3	34	2.9	83	3.6

DIF analyses were generated to compare male and female and also White and Black. (The number of Latinx examinees in the field test sample was too low to conduct a DIF analysis.) For the two forms, consisting of a combined total of 70 items, two items were flagged for C-level DIF. The summary of the DIF analyses for the two forms is provided in Table 12.4.

Table 12.4. Identifications of C-Level DIF Items on the Two Workforce Documents Forms

Form	Number of flagged items	Favored group
WS1	1	White
WS2	1	White

The DIF analysis from the field test study needs to be interpreted with caution because the sample size for Black for each form was small (*n* = 197 and *n* = 218). The results from this small sample size might not be generalizable to the general testing population. Because DIF methods require large sample sizes, ACT could not conduct DIF analyses for the other demographic groups. ACT continues to conduct DIF analyses whenever test forms are administered and thoroughly reviews items that are flagged by the DIF analyses.

12.4.1.2 Applied Math

During the second step in the field-testing process, ACT administered the two forms of the Applied Math assessment to 2,381 field test participants. Forty testing sites in 22 states participated (13 high schools and 27 adult testing centers). Approximately 61% of the examinees were high school students and 39% were adults. Prior to the test administration, ACT instructed 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 12.5 presents the demographic characteristics by test form for the Applied Math assessment.

Table 12.5. Applied Math—Number and Percent of Field Test Participants by Demographic Group

Demographic characteristic	Form MS1		Form MS2		Total	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Total participants	1,185	49.8	1,196	50.2	2,381	100
Male	549	46.3	513	42.9	1,062	44.6
Female	609	51.4	648	54.2	1,257	52.8
Black	206	17.4	233	19.5	439	18.4
American Indian/Alaska Native	27	2.3	19	1.6	46	1.9
Asian	9	.8	6	.5	15	.6
Latinx	79	6.7	82	6.9	161	6.8
Native Hawaiian/Pacific Islander	1	.1	1	.1	2	.1
Two or more races/ethnicities	102	8.6	101	8.4	202	8.5
White non-Latinx	719	60.7	714	59.7	1,433	60.9
Prefer not to respond	42	3.5	40	3.3	82	3.4

DIF analyses were generated to compare male and female and also Black and White. (The number of Latinx examinees in the field test sample was too low to conduct a DIF analysis.) For the two forms, consisting of a combined total of 68 items, six items were flagged for C-level DIF. The summary of the DIF analyses for the two forms is presented in Table 12.6.

Table 12.6. Identifications of C-Level DIF Items on the Two Applied Math Forms

Form	Number of flagged items	Favored group
MS1	3	Black; White; White
MS2	3	Male; Female; Female

The DIF analysis from the field test study needs to be interpreted with caution because the sample size for Black for each form was small (*n* = 206 and *n* = 233). The results from this small sample size might not be generalizable to the general testing population. Because DIF methods require large sample sizes, ACT could not conduct DIF analyses for the other demographic group comparisons. ACT continues to conduct DIF analyses whenever test forms are administered and thoroughly reviews items that are flagged by the DIF analyses.

12.4.1.3 Graphic Literacy

During the second step in the field-testing process, ACT administered the two forms of the Graphic Literacy assessment to 2,265 field test participants. Forty testing sites in 22 states participated (13 high schools and 27 adult testing centers). Approximately 61% of the examinees were high school students and 39% were adults. Prior to the administration, ACT instructed 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 12.7 presents the demographic characteristics by test form for the Graphic Literacy assessment.

Table 12.7. Graphic Literacy—Number and Percent of Field Test Participants by Demographic Group

Demographic characteristic	Form GS1		Form GS2		Total	
	Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Total participants	1,170	51.6	1,096	48.4	2,266	100
Male	534	45.6	471	43.0	1,005	44.4
Female	601	51.4	586	53.5	1,187	52.4
Black	207	17.7	225	20.5	432	19.1
American Indian/Alaska Native	24	2.1	18	1.6	42	1.9
Asian	7	.6	4	.4	11	.5
Latinx	77	6.6	51	4.7	128	5.6
Native Hawaiian/Pacific Islander	1	.1	1	.1	2	.1
Two or more races/ethnicities	90	7.7	87	7.9	177	7.8
White non-Latinx	714	61.0	667	60.9	1,381	60.9
Prefer not to respond	50	4.3	43	3.9	93	4.1

DIF analyses were generated to compare male and female and also Black and White. (The number of Latinx examinees in the field test sample was too low to conduct a DIF analysis.) For the two forms, consisting of a combined total of 76 items (18 items were common for the two forms), four items were flagged for C-level DIF. The summary of the DIF analyses for the two forms is presented in Table 12.8.

Table 12.8. Identifications of C-Level DIF Items on the Two Graphic Literacy Forms

Form	Number of flagged Items	Favored group
GS1	3	Female; Black; White
GS2	1	Male

The DIF analysis from the field test study needs to be interpreted with caution because the sample size for Black for each form was small (*n* = 207 and *n* = 224). The results from this small sample size might not be generalizable to the general testing population. Because DIF methods require large sample sizes, ACT could not conduct DIF analyses for the other demographic groups. ACT continues to conduct DIF analyses whenever test forms are administered and thoroughly reviews items that are flagged by the DIF analyses.

12.5 Fairness as the Validity of Individual Score Interpretations

Fairness of individual score interpretations becomes an important point when an assessment score is used as part of a process for making high-stakes decisions. When employment decisions are made based in part on WorkKeys NCRC scores, ACT considers this high-stakes test use. As a result, federal rules and procedures should be followed so that score interpretations can be valid, fair, and legal.

The federal agencies responsible for enforcing civil rights legislation collectively published the *Uniform Guidelines on Employee Selection Procedures (1978)* (EEOC et al., 1978), which regulates how an assessment process may be used to assist in employee selection. If a selection procedure adversely affects a protected group, the procedure should not be followed unless the employer is able to demonstrate that the assessment measures skills that are job-related.

Adverse impact occurs when a seemingly neutral employee-selection practice has a disproportionately negative effect on members of a protected group (Society for Human Resource Management, 2022). Under the applicable federal law, adverse impact does not require any intended discrimination on the part of the employer. 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 rate at which the reference group is selected. The second method is referred to as the statistical significance test. This method attempts to determine whether the difference in selection rates is greater than what would be expected if selections happened by chance. The statistical significance test uses Fisher’s exact test and interprets a difference of two standard deviations as indicative of adverse impact.

When a selection process that uses assessment scores shows an 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 are not evidence of test bias. There are many reasons why such differences may exist in a cognitive ability test. Ultimately, a differential prediction study may be conducted to examine test bias and to examine whether there are differences across demographic groups in the slopes and intercepts of the regression equations that are used to predict outcomes (e.g., job performance, or turnover, etc.). Differential prediction studies can be conducted with applicants if the applicants are later employed, or they can be conducted 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. Furthermore, organizations that use WorkKeys should conduct a job analysis if they intend to use any WorkKeys NCRC assessment test scores as part of their employment decisions.

When any WorkKeys NCRC 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, the cutoff scores should be established at appropriate levels, and the process for identifying the levels should be clearly documented (AERA et al., 2014; Society for Industrial and Organizational Psychology, 2018).

The *Uniform Guidelines* and the *Standards* recognize the use of job analysis coupled with 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 12.9 describes the validation requirements of the *Uniform Guidelines* and how ACT's job-profiling process meets the requirements. (Chapter 11 provides a detailed description of job profiling and content validation.)

Table 12.9. 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	Subject matter experts (SMEs) participating in the job-profiling procedure establish a list that describes behaviors and tasks using the tasks from O*NET API in SkillPro software; the SMEs then customize the list using information gained from company materials, interviews, and job shadowing. Then the SMEs rate each task for importance, and the SkillPro software averages the SMEs' ratings to yield a list of tasks in the 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 (i.e., task with 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. The SMEs link specific tasks to a skill level and explain how skills at the specific level are 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 the skill 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 effectively 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 the SMEs' own skill levels.
Scores are interpreted as pass/fail only; they must not be interpreted as a rank ordering of examinees.	WorkKeys scores show that examinees either have or do not have the skills at the required levels. It is not appropriate to rank the order of examinees on the basis of 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.

Any time an employer wants to use a WorkKeys Assessment as part of the selection process, ACT recommends that the employer use the job-profiling process to assist in determining both the requisite skills and the skill levels for the job. By using job profiling, the employer can make more efficient use of the WorkKeys Assessment suite. Furthermore, the employer is also providing job applicants with a fair method of selection consistent with the *Uniform Guidelines*.

12.5.1 Level Scores Analysis of WorkKeys NCRC

Tables 12.10, 12.11, and 12.12 provide the average WorkKeys level scores and the standard deviations for different groups who took Applied Math, Graphic Literacy, and Workplace Documents assessments, respectively, from June 2020 to July 2022. Cohen's *d* was computed to examine how meaningful the mean difference between the groups was. Cohen's *d* is a statistic useful for interpreting mean differences by providing results in standard deviation units. Cohen's *d* was computed using the following formula:

$$d = \frac{\overline{x_1} - \overline{x_2}}{s}$$

where $\overline{x_1}$ is the mean of Group 1, $\overline{x_2}$ is the mean of Group 2, and *s* is the pooled standard deviation. The pooled standard deviation was computed as follows:

$$s = \sqrt{\frac{\left(SD_1^2 + SD_2^2\right)}{2}}$$

where SD_1 is the standard deviation for Group 1 and SD_2 is the standard deviation for Group 2. When Cohen's *d* is smaller than 0.2, differences can be viewed as negligible, whereas when the value is larger than 0.8, differences can be viewed as large.

Mean differences for ethnic groups on cognitive assessments are a common finding (Gottfredson, 1988; Sackett & Wilk, 1994; Roth et al., 2001). When Cohen's *d* was computed between the gender groups and between the ethnic groups (e.g., White vs. others and Asian vs. others), the major findings were as follows: First, the gender differences across all three assessments were all negligible (Cohen's $d < 0.20$). The largest Cohen's *d* value across the three assessments was a 0.15 value favoring female examinees on Workplace Documents. Second, the ethnic group differences were larger than the gender group differences; nevertheless, most ethnic groups (except in Applied Math between White and Black, Asian and Black, and Asian and Native Hawaiian/Pacific Islander) still showed only small to moderate mean score differences (i.e., no meaningful differences) when each non-White ethnic group was compared to the White group and when each non-Asian ethnic group was compared to the Asian group.

Table 12.10. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Applied Math

Category	Demographic	*n*	Mean	SD
Gender	Female	249,680	4.22	1.68
	Male	283,883	4.41	1.81
Race/ ethnicity	American Indian/Alaska Native	5,975	3.73	1.74
	Asian	10,373	5.08	1.74
	Black	110,142	3.37	1.67
	Latinx	57,845	4.03	1.69
	Native Hawaiian/Pacific Islander	1,411	3.66	1.79
	Two or more races/ethnicities	20,182	4.38	1.66
	White	272,575	4.74	1.60

Table 12.11. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Graphic Literacy

Category	Demographic	*n*	Mean	SD
Gender	Female	249,680	4.55	1.46
	Male	283,883	4.50	1.66
Race/ ethnicity	American Indian/Alaska Native	5,975	4.03	1.67
	Asian	10,373	5.06	1.51
	Black	110,142	3.89	1.55
	Latinx	57,845	4.27	1.57
	Native Hawaiian/Pacific Islander	1,411	4.06	1.62
	Two or more races/ethnicities	20,182	4.60	1.50
	White	272,575	4.83	1.46

Table 12.12. Average Level Scores, Standard Deviations (SD), and Numbers of Examinees (*n*) for Gender and Ethnicity of Workplace Documents

Category	Demographic	*n*	Mean	SD
Gender	Female	249,680	4.52	1.36
	Male	283,883	4.30	1.60
Race/ ethnicity	American Indian/Alaska Native	5,975	3.95	1.54
	Asian	10,373	4.77	1.52
	Black	110,142	3.88	1.44
	Latinx	57,845	4.11	1.51
	Native Hawaiian/Pacific Islander	1,411	3.95	1.54
	Two or more races/ethnicities	20,182	4.46	1.44
	White	272,575	4.67	1.42

Chapter 13: Operational Validation

13.1 Overview

Since the launch of the updated ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) assessments, it has been important to continuously monitor and review the psychometric properties of operational testing forms. This chapter summarizes the analyses and findings from the WorkKeys NCRC assessments administered from May 2021 to April 2022. The chapter includes demographic statistics of the assessment population and psychometric analyses from four operational form administrations to provide evidence of test quality. The findings from these analyses and demographic statistics should be interpreted as an extension of the psychometric analyses (presented in earlier chapters) that were based on the field study. The following results are reported in this chapter to further support the analyses summarized in earlier chapters:

- gender and racial/ethnic group summary
- summary statistics for four operational forms (three computer-based testing [CBT] forms and one paper form)
- reliability results that include classification consistency results for the forms
- an example of dimensionality evaluation from one of the three CBT forms

13.2 Workplace Documents

13.2.1 Examinees

This section summarizes the results of analyses of different gender and racial/ethnic groups from the examinees who took at least one ACT® WorkKeys® Workplace Documents assessment between May 1, 2021, and April 30, 2022.

A total of 474,196 examinees who took one of the Workplace Documents forms between these dates and had valid scores were included in the analyses. Based on the gender and racial/ethnic group distributions, the assessment samples shown in Table 13.1 are consistent with samples from previous WorkKeys NCRC test administrations, as shown in Chapter 9, Table 9.1. Current demographic trends consistent with demographic trends seen in past administrations are (a) more male examinees than female examinees and (b) White, Black, and Latinx examinees at approximate proportions of 40%, 20%, and 10%, respectively. As evident in Table 13.1, the average scale score earned by male examinees (78.6) is over a half score point lower than that earned by female examinees (79.4). Among the four largest racial/ethnic groups, the average scale score, ordered from highest to lowest, was earned by Asian, White, Latinx, and Black examinees. The average scale score for this large sample of 474,196 examinees is 78.9 with a standard deviation (SD) of 4.5.

Table 13.1. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Workplace Documents (5/1/2021–4/30/2022)

				Scale score		Percentage distributions for level scores					
Group		*N*	%	Mean	SD	Below 3	3	4	5	6	7
Full group		474,196	—	78.9	4.5	7	18	35	20	15	5
Gender	Female	188,172	39.7	79.4	4.2	5	16	36	22	16	5
	Male	215,944	45.5	78.6	4.7	9	19	33	19	14	5
	Missing	70,080	14.8	78.6	4.5	8	20	35	18	14	5
Race/ ethnicity	White	196,136	41.4	79.9	4.3	5	13	33	23	19	7
	Black	91,564	19.3	77.3	4.2	11	26	39	15	7	1
	Latinx	48,443	10.2	78.1	4.4	10	21	37	18	11	3
	Asian	8,227	1.7	80.2	4.7	6	14	28	20	21	11
	American Indian/ Alaska Native	4,577	1.0	77.6	4.5	11	24	37	16	10	2
	Native Hawaiian/ Pacific Islander	1,070	0.2	77.5	4.4	12	23	37	17	10	1
	Two or more races/ ethnicities	16,030	3.4	79.3	4.3	6	16	35	22	16	6
	Missing	108,149	22.8	78.8	4.6	8	19	34	19	15	5

Note. Results are based on test records with valid scale scores. Missing groups include the response category “prefer not to respond” for gender and racial/ethnic groups. Percentages of CBT and paper test administrations are 67% and 33%, respectively.

13.2.2 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 13.2.1. The results in Table 13.2 include the sample sizes, gender and racial/ethnic group distributions, test completion rates, and scale score means and standard deviations. There are three CBT forms (CBT 1, CBT 2, and CBT 3) and one paper form. The CBT forms were administered from September 2021 to December 2021, and the paper form was administered from May 2021 to April 2022.

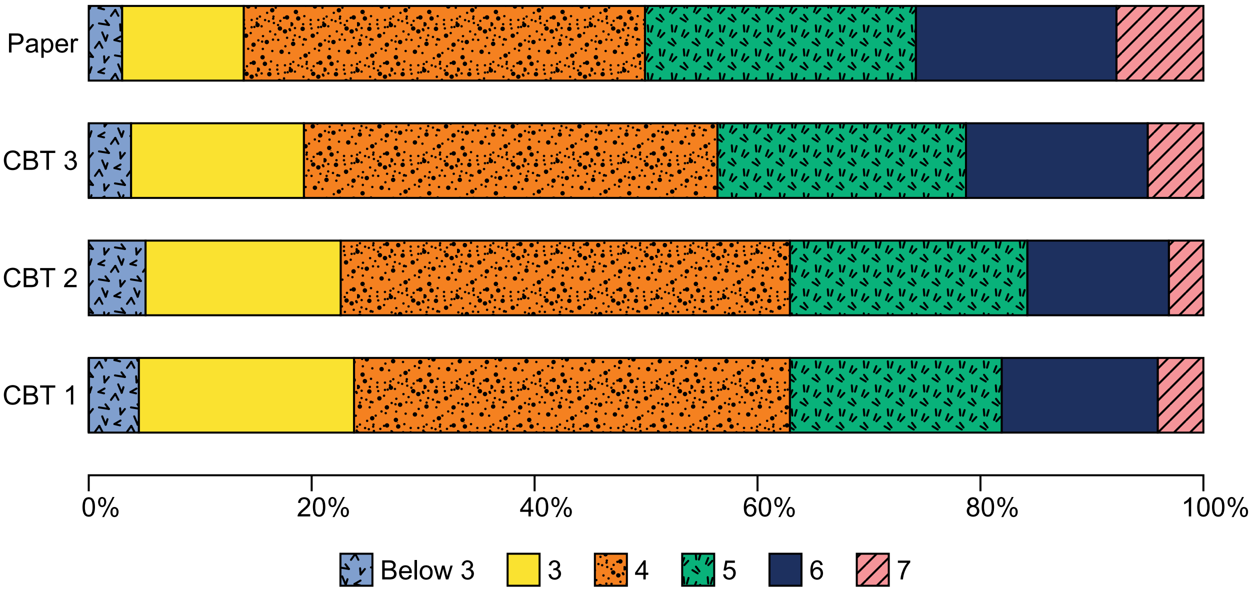
As shown in Table 13.2, the demographic proportions for each of the four examinee groups are similar to the demographic proportions given in Table 13.1 for the total sample of 474,196 examinees. One prominent exception is a higher percentage of male examinees who took the paper form (60.9%). As shown in Table 13.2, the CBT forms have larger sample sizes than the paper form (approximately 34,000 and 5,000, respectively). The demographic percentages of the three largest racial/ethnic groups are similar to those reported in Table 13.1.

Table 13.2. Summary Statistics for Four Workplace Documents Forms

Form	*N*	Female (%)	Male (%)	White (%)	Black (%)	Latinx (%)	Test completion (%)	Scale score mean	Scale score SD
CBT 1	33,823	38.1	40.7	39.5	19.6	11.2	96.1	78.95	3.94
CBT 2	33,872	38.3	40.4	39.4	19.9	11.2	95.9	78.90	3.84
CBT 3	34,169	38.0	41.3	39.9	19.8	11.2	97.0	79.52	3.90
Paper	5,213	27.7	60.9	50.2	16.6	7.3	94.7	80.17	3.88

Test completion rates are over 90% for the four forms. The average scale scores range from 78.90 to 80.17, which are higher scale scores than the targeted mean scale score of 77.3 (see Chapter 8, Section 8.4, Procedures for Establishing the Score Scale). The standard error of measurement (SEM) is consistent with the targeted SEM of 1.7 (see Table 13.4 for the SEM for each form). Figure 13.1 presents the level score distributions for these forms (note that for all four forms, the percentages for both the Below Level 3 and the Level 7 groups are 7.7% or lower).

Figure 13.1. Level Score Distributions for Workplace Documents



The test characteristic curves (TCCs) and test information functions (TIFs) for the four forms are presented in Figure 13.2 and Figure 13.3, respectively. For comparison, the scaling form is included as the base form (identical to that in Chapter 8, Figure 8.3). Note that these forms were built to meet the same test blueprint as presented in Chapter 3, Section 3.3. The TCCs are placed tightly across the forms, as shown in Figure 13.2.

Figure 13.2. Test Characteristic Curves (TCCs) for Workplace Documents—Base Form and Four Operational Forms

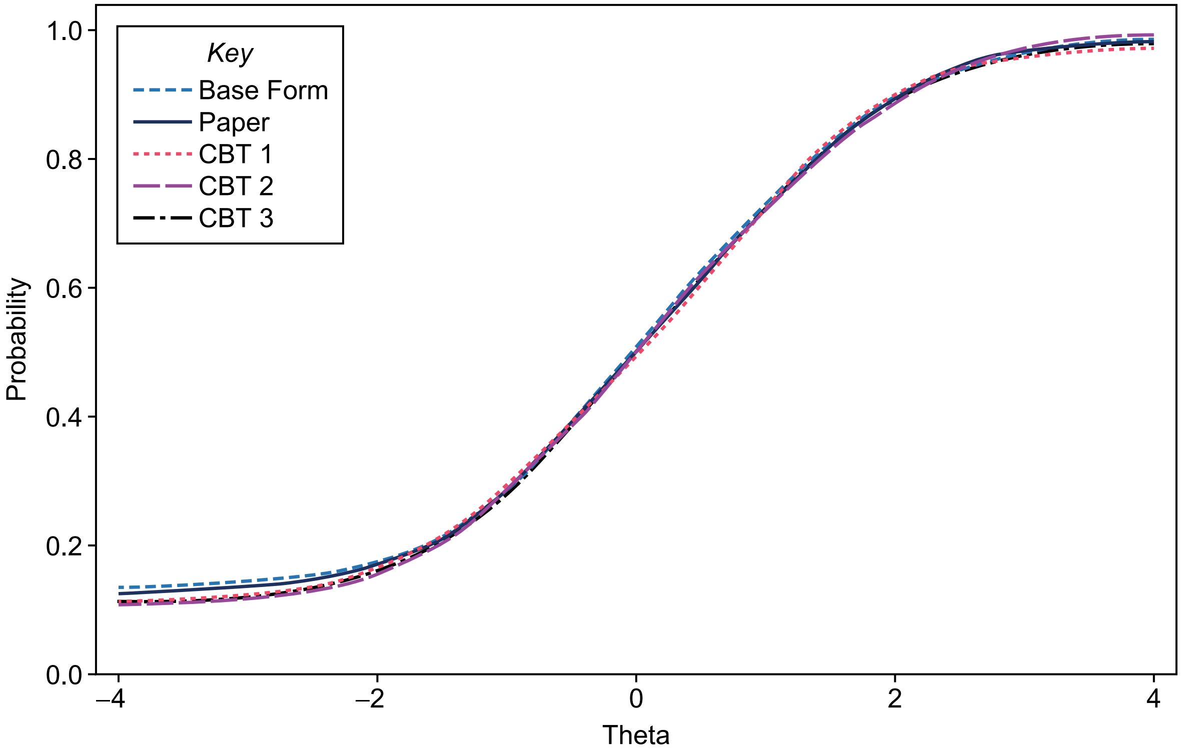
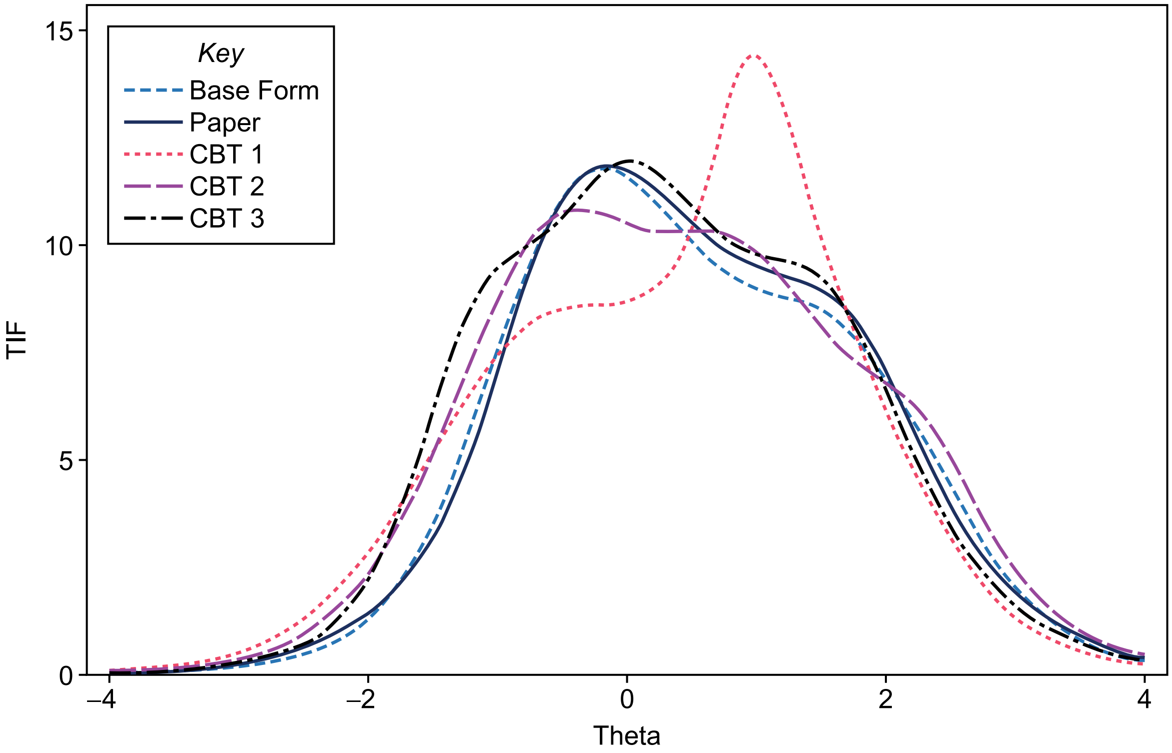


Figure 13.3. Test Information Function (TIF) Curves for Workplace Documents—Base Form and Four Operational Forms



ACT researchers also monitor differential item functioning (DIF) for both pretest and operational items using the method presented in Chapter 12, Section 12.4. On the four forms, four items were flagged as Group C DIF based on comparisons between Black and White or Latinx and White, as presented in Table 13.3. ACT further evaluated the four items, and there were no fairness concerns for them; thus, these four items continue to be used.

Table 13.3. Summary of C-Level DIF Items on Four Workplace Documents Forms

Form	Number of flagged items	Favored group
CBT 1	None	NA
CBT 2	2	Black (White vs. Black); White (White vs. Black)
CBT 3	None	NA
Paper	2	Black (White vs. Black); Black[a] (White vs. Black); Latinx[a] (White vs. Latinx)

[a] One item was flagged as favoring two races/ethnicities.

13.2.3 Reliability Analyses

The reliability analyses were divided into two parts. The first was based on computing familiar estimates of reliability, including Cronbach’s alpha, scale score reliability, and SEMs for the scale scores from the four forms. Cronbach’s alpha estimates were .83 or .84 (see Table 13.4), which are slightly lower than .89 for the scaling form (reported in Chapter 10, Section 10.2). For the four forms, the reliability estimates for the scale scores were .89 or .90, and the scale score SEM values ranged from 1.67 to 1.71, most of which are slightly lower than 1.70 for the scaling form.

Table 13.4. Reliability and SEM Results for Four Workplace Documents Forms

Form	Cronbach’s alpha	Scale score reliability	Scale score SEM
CBT 1	.84	.90	1.67
CBT 2	.83	.89	1.71
CBT 3	.84	.89	1.69
Paper	.84	.90	1.68

The second part of the reliability analyses was based on computing the classification consistency of each WorkKeys level for the four forms. Classification consistency analysis (described in Chapter 10, Section 10.4) was computed for these forms using item parameter estimates that were used in pre-equating. By comparing Table 13.5 to Chapter 10, Table 10.7, one can observe that the classification consistency results are relatively stable.

Table 13.5. Estimated Classification Consistency Indices for Level Scores for Four Workplace Documents Forms

	CBT 1		CBT 2		CBT 3		Paper	
Level	*P* (%)	κ (%)	*P* (%)	κ (%)	*P* (%)	κ (%)	*P* (%)	κ (%)
Exact	57	46	57	46	57	46	57	46
3	91	70	91	69	91	69	91	70
4	88	74	88	74	88	75	88	76
5	88	72	88	72	88	72	88	72
6	91	66	91	66	91	66	91	66
7	95	54	95	54	95	53	95	53

13.2.4 Dimensionality Evaluation

This section provides evidence that the Workplace Documents assessment is unidimensional. This evidence was derived using the method described in Chapter 9, Section 9.2.5 (i.e., comparing the eigenvalues of the first three factors from the exploratory factor analysis [EFA]). Table 13.6 presents the EFA results for the CBT 1 form. As shown in Table 13.6, the factor difference ratio index (FDRI) value is significantly greater than 3, and the first factor explains 18% of the total variance for the full set of operational items. These findings indicate an underlying single-factor structure of the Workplace Documents assessment.

Table 13.6. Eigenvalues and Factor Difference Ratio Index (FDRI) for Workplace Documents—CBT 1 Form

Factor	Eigenvalue	Difference	FDRI
1	5.43 (17.51%)	—	—
2	1.68 (5.41%)	3.75	—
3	1.03 (3.33%)	0.65	5.77

Note. The percentage in each set of parentheses is the percentage of the total variance accounted for by that factor.

In summary, the results in this chapter that were obtained using recent operational data consistently support the findings from the field study and also provide strong evidence of the sound psychometric quality of the Workplace Documents forms. As additional Workplace Documents forms are developed according to the WorkKeys Assessment blueprint and statistical guidelines, ACT researchers continue to conduct similar analyses to review and monitor test form and item quality.

13.3 Applied Math

13.3.1 Examinees

This section summarizes the results of analyses of different gender and racial/ethnic groups from the examinees who took at least one ACT® WorkKeys® Applied Math assessment between May 1, 2021, and April 30, 2022.

A total of 479,629 examinees who took one of the Applied Math forms between these dates and had valid scores were included in the analyses. Based on the gender and racial/ethnic group distributions, the assessment samples shown in Table 13.7 are consistent with samples from previous test administrations, as shown in Chapter 9, Table 9.2. Current demographic trends consistent with demographic trends seen in past administrations are (a) more male examinees than female examinees and (b) White, Black, and Latinx examinees at approximate proportions of 40%, 20%, and 10%, respectively. As evident in Table 13.7, the average scale score earned by male examinees (78.6) is over a half score point higher than that earned by female examinees (77.9). Among the four largest racial/ethnic groups, the average scale score, ordered from highest to lowest, was earned by Asian, White, Latinx, and Black examinees. The average scale score for this large sample of 479,629 examinees is 78.1 with a standard deviation of 5.3.

Table 13.7. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Applied Math Assessment (5/1/2021–4/30/2022)

				Scale score		Percentage distributions for level scores					
	Group	*N*	%	Mean	SD	Below 3	3	4	5	6	7
Full group		479,629	—	78.1	5.3	11	22	27	18	12	9
Gender	Female	193,904	40.4	77.9	5.0	10	24	29	19	11	8
	Male	216,099	45.1	78.6	5.5	11	19	25	19	14	12
	Missing	69,626	14.5	77.3	5.3	14	26	27	14	11	7
Race/ ethnicity	White	197,564	41.2	79.8	5.1	6	15	26	23	17	13
	Black	95,428	19.9	75.3	4.3	20	35	28	11	4	1
	Latinx	48,716	10.2	77.4	4.9	12	25	29	18	10	6
	Asian	8,003	1.7	81.1	5.7	6	13	20	19	17	26
	American Indian/ Alaska Native	4,466	0.9	76.6	4.8	16	26	31	16	8	4
	Native Hawaiian/ Pacific Islander	1,099	0.2	76.2	4.9	18	28	28	16	6	4
	Two or more races/ ethnicities	16,187	3.4	78.5	5.1	8	21	29	20	12	9
	Missing	108,166	22.6	77.8	5.5	13	24	26	16	12	10

Note. Results are based on test records with valid scale scores. Missing groups include the response category “prefer not to respond” for gender and racial/ethnic groups. Percentages of CBT and paper test administrations are 67% and 33%, respectively.

13.3.2 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 13.3.1. The results in Table 13.8 include the sample sizes, gender and racial/ethnic group distributions, test completion rates, and scale score means and standard deviations. There are three CBT forms (CBT 1, CBT 2, and CBT 3) and one paper form. The CBT forms were administered from September 2021 to December 2021, and the paper form was administered from May 2021 to March 2022.

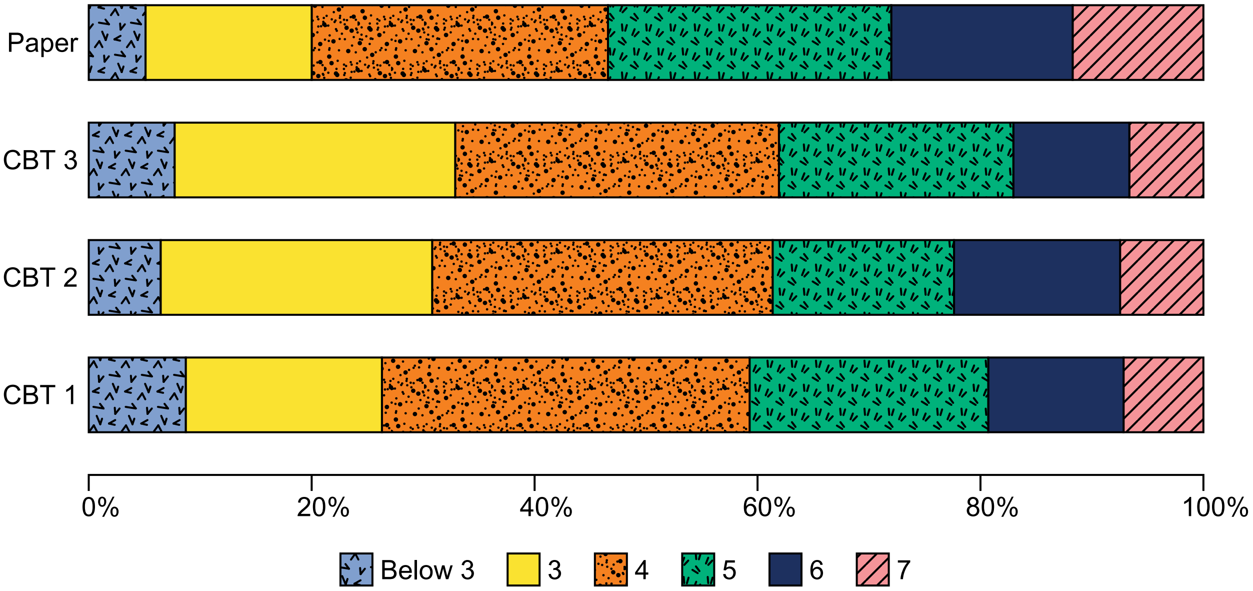
As shown in Table 13.8, the demographic proportions for each of the four examinee groups are similar to the demographic proportions given in Table 13.7 for the total sample of 479,629 examinees. One prominent exception is a higher percentage of male examinees who took the paper form (61.5%). As shown in Table 13.8, the CBT forms have larger sample sizes than the paper form (approximately 34,000 and 5,000, respectively). The demographic percentages of the three largest racial/ethnic groups are similar to those reported in Table 13.7.

Table 13.8. Summary Statistics for Four Applied Math Forms

Form	*N*	Female (%)	Male (%)	White (%)	Black (%)	Latinx (%)	Test completion (%)	Scale score mean	Scale score SD
CBT 1	34,082	38.6	40.6	40.1	19.5	11.3	96.4	78.33	4.91
CBT 2	34,015	38.2	40.8	39.5	19.9	11.6	96.7	78.43	4.84
CBT 3	33,964	39.0	40.5	39.4	20.4	11.0	96.7	78.13	4.83
Paper	5,454	26.7	61.5	50.5	15.9	7.2	94.7	79.86	4.97

Test completion rates are over 90% for the four forms. The average scale scores are about 78, which is consistent with the targeted mean scale score of 77.9 (see Chapter 8, Section 8.4, Procedures for Establishing the Score Scale). The SEM of 1.7 is consistent with the targeted SEM of 1.6 (see Table 13.10 for the SEM for each form). Figure 13.4 presents the level score distributions for these forms (note that for all four forms, the percentages for both the Below Level 3 and the Level 7 groups are below 12%).

Figure 13.4. Level Score Distributions for Applied Math



TCCs and TIFs for the four forms are presented in Figure 13.5 and Figure 13.6, respectively. For comparison, the scaling form is included as the base form (identical to that in Chapter 8, Figure 8.8). Note that these forms were built to meet the same test blueprint as presented in Chapter 3, Section 3.3. The TCCs are placed tightly across the forms, as shown in Figure 13.5.

Figure 13.5. Test Characteristic Curves (TCCs) for Applied Math—Base Form and Four Operational Forms

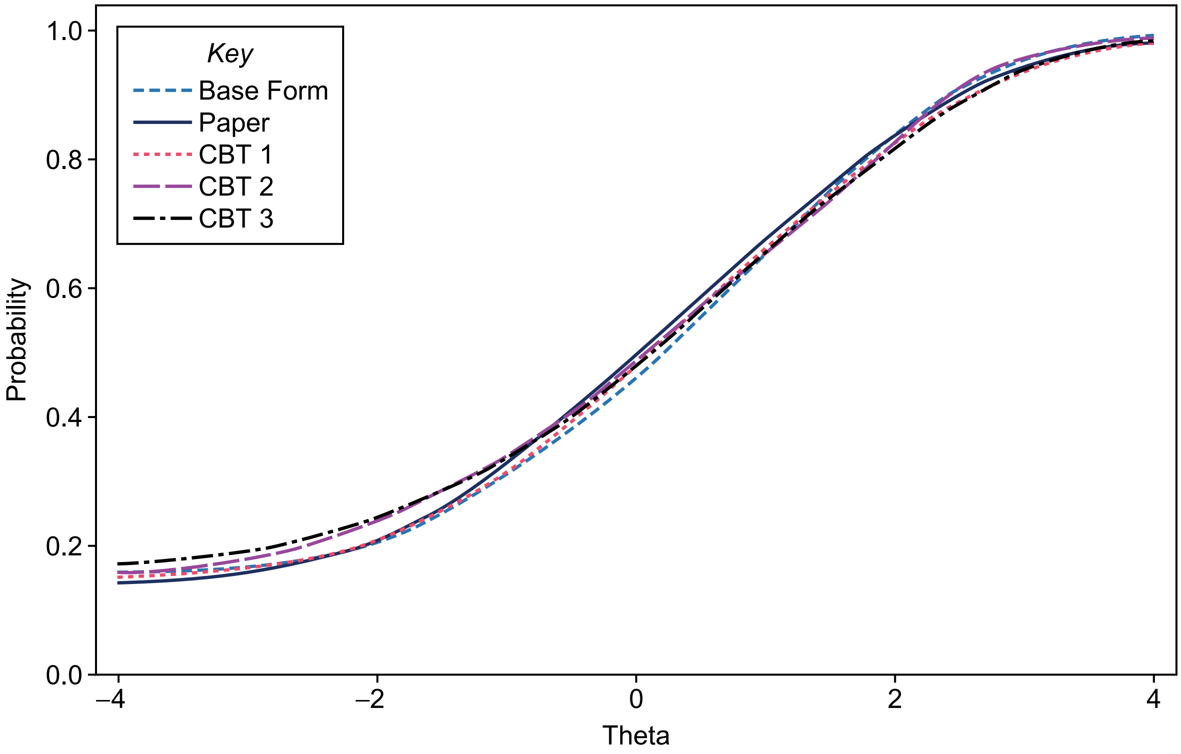
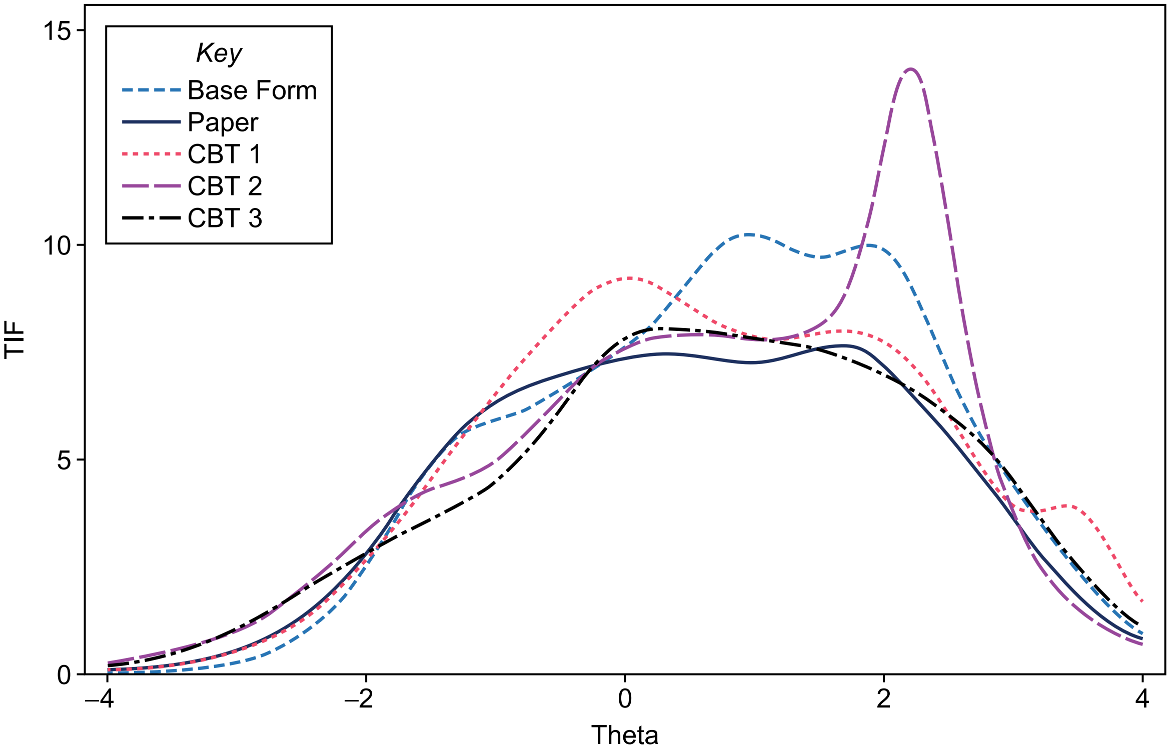


Figure 13.6. Test Information Function (TIF) Curves for Applied Math—Base Form and Four Operational Forms



ACT researchers also monitor DIF for both pretest and operational items using the method presented in Chapter 12, Section 12.4. For the four forms, three items were flagged as Group C DIF based on the comparisons between female and male and Black and White, as presented in Table 13.9. ACT further evaluated the three items, and there were no fairness concerns for them; thus, these three items continue to be used.

Table 13.9. Summary of C-Level DIF Items on Four Applied Math Forms

Form	Number of flagged items	Favored group
CBT 1	1	Male (female vs. male)
CBT 2	None	NA
CBT 3	1	White (White vs. Black)
Paper	1	White (White vs. Black)

13.3.3 Reliability Analyses

The reliability analyses were divided into two parts. The first was based on computing familiar estimates of reliability, including Cronbach's alpha, scale score reliability, and SEMs for the scale scores from the four forms. Cronbach's alpha estimates were .87 or .88 (see Table 13.10), which are similar to .88 for the scaling form (reported in Chapter 10, Section 10.2). For the four forms, the reliability estimates for the scale scores were .87 or .88, and the scale score SEM values ranged from 1.66 to 1.74, which are slightly higher than 1.61 for the scaling form.

Table 13.10. Reliability and SEM Results for Four Applied Math Forms

Form	Cronbach's alpha	Scale score reliability	Scale score SEM
CBT 1	.88	.87	1.68
CBT 2	.87	.88	1.66
CBT 3	.87	.87	1.70
Paper	.88	.87	1.74

The second part of the reliability analyses was based on computing the classification consistency of each WorkKeys level for the four forms. Classification consistency analysis (described in Chapter 10, Section 10.4) was computed for these forms using item parameter estimates that were used in pre-equating. By comparing Table 13.11 to Chapter 10, Table 10.8, one can observe that the classification consistency results are quite stable.

Table 13.11. Estimated Classification Consistency Indices for Level Scores for Four Applied Math Forms

	CBT 1		CBT 2		CBT 3		Paper	
Level	P (%)	κ (%)	P (%)	κ (%)	P (%)	κ (%)	P (%)	κ (%)
Exact	55	42	55	43	54	42	54	41
3	93	53	93	52	93	51	93	55
4	87	68	86	68	86	69	86	68
5	86	70	86	69	85	69	85	69
6	91	66	90	68	91	65	91	65
7	96	62	96	66	96	62	96	60

13.3.4 Dimensionality Evaluation

This section provides evidence that the Applied Math assessment is unidimensional based on the same method used in Chapter 9, Section 9.2.5 (the eigenvalue comparisons of the first three factors from the EFA). Table 13.12 presents the EFA results for the CBT 1 form. As shown in Table 13.12, the FDRI value is significantly greater than 3, and the first factor explains 24% of the total variance for the full set of operational items. These findings indicate an underlying single-factor structure of the Applied Math assessment.

Table 13.12. Eigenvalues and Factor Difference Ratio Index (FDRI) for Applied Math—CBT 1 Form

Factor	Eigenvalue	Difference	FDRI
1	7.36 (23.74%)	—	—
2	2.03 (6.56%)	5.32	—
3	1.11 (3.57%)	0.93	5.72

Note. The percentage in each set of parentheses is the percentage of the total variance accounted for by that factor.

In summary, the results in this chapter that were obtained using recent operational data consistently support the findings from the field study and also provide strong evidence of the sound psychometric quality of the Applied Math forms. As additional Applied Math forms are developed according to the WorkKeys Assessment blueprint and statistical guidelines, ACT researchers continue to conduct similar analyses to review and monitor test form and item quality.

13.4 Graphic Literacy

13.4.1 Examinees

This section summarizes the results of analyses of different gender and racial/ethnic groups from the examinees who took at least one ACT® WorkKeys® Graphic Literacy assessment between May 1, 2021, and April 30, 2022.

A total of 455,099 examinees who took one of the Graphic Literacy forms between these dates and had valid scores were included in the analyses. Based on the gender and racial/ethnic group distributions, the assessment samples shown in Table 13.13 are consistent with samples from previous test administrations, as shown in Chapter 9, Table 9.3. Current demographic trends consistent with demographic trends seen in past assessments are (a) more male examinees than female examinees and (b) White, Black, and Latinx examinees at approximate proportions of 40%, 20%, and 10%, respectively. As evident in Table 13.13, the average scale score earned by male examinees (78.3) is identical to that earned by female examinees (78.3). Among the four largest racial/ethnic groups, the average scale score, ordered from highest to lowest, was earned by Asian, White, Latinx, and Black examinees. The average scale score for this large sample of 455,099 examinees is 78.2 with a standard deviation of 4.5.

Table 13.13. Score Summary for Different Gender and Racial/Ethnic Groups for WorkKeys Graphic Literacy Assessment (5/1/2021–4/30/2022)

				Scale score		Percentage distributions for level scores					
Group		*N*	%	Mean	SD	Below 3	3	4	5	6	7
Full group		455,099	—	78.2	4.5	7	23	16	30	18	6
Gender	Female	180,278	39.6	78.3	4.3	5	23	17	32	18	4
	Male	207,501	45.6	78.3	4.8	8	22	15	29	19	7
	Missing	67,320	14.8	77.7	4.5	9	25	17	28	15	5
Race/ ethnicity	White	188,824	41.5	79.4	4.5	4	16	14	33	24	8
	Black	87,811	19.3	76.2	3.9	11	34	20	25	8	1
	Latinx	45,111	9.9	77.5	4.3	8	26	18	30	14	4
	Asian	7,737	1.7	80.2	4.8	4	14	12	28	29	13
	American Indian/ Alaska Native	4,135	.9	76.9	4.3	11	29	18	28	12	3
	Native Hawaiian/ Pacific Islander	1,045	.2	76.9	4.3	11	30	18	25	13	2
	Two or more races/ ethnicities	15,465	3.4	78.5	4.4	6	21	16	32	19	6
	Missing	104,971	23.1	77.9	4.6	9	24	16	29	17	5

Note. Results are based on test records with valid scale scores. Missing groups include the response category “prefer not to respond” for gender and racial/ethnic groups. Percentages of CBT and paper test administrations are 66% and 34%, respectively.

13.4.2 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 13.4.1. The results in Table 13.14 include the sample sizes, gender and racial/ethnic group distributions, test completion rates, and scale score means and standard deviations. There are three CBT forms (CBT 1, CBT 2, and CBT 3) and one paper form. The CBT forms were administered from September 2021 to December 2021, and the paper form was administered from May 2021 to March 2022.

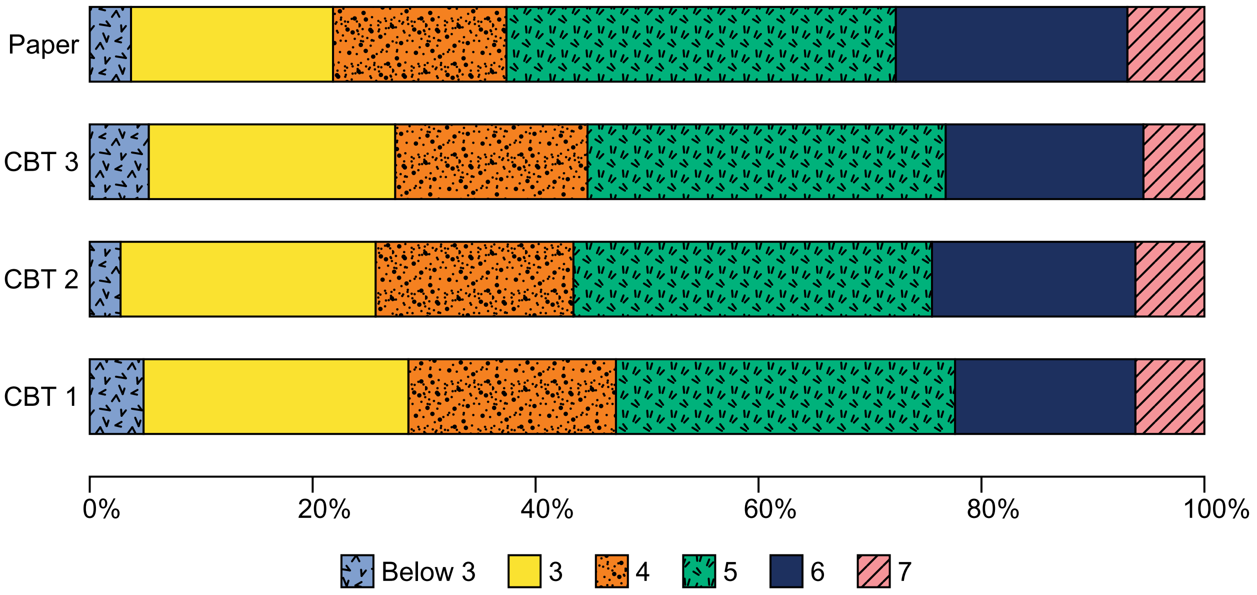
As shown in Table 13.14, the demographic proportions for each of the four examinee groups are similar to the demographic proportions given in Table 13.13 for the total sample of 455,099 examinees. One prominent exception is a higher percentage of male examinees who took the paper form (62.0%). As shown in Table 13.14, the CBT forms have larger sample sizes than the paper form (approximately 32,000 and 4,500, respectively). The demographic percentages of the three largest racial/ethnic groups are similar to those reported in Table 13.13.

Table 13.14. Summary Statistics for Four Graphic Literacy Forms

Form	*N*	Female (%)	Male (%)	White (%)	Black (%)	Latinx (%)	Test completion (%)	Scale score mean	Scale score SD
CBT 1	32,268	37.5	40.7	39.4	19.2	11.2	97.6	78.32	4.31
CBT 2	32,744	37.8	41.3	39.8	19.6	11.1	97.4	78.74	4.24
CBT 3	32,338	38.2	40.4	39.9	19.4	11.0	96.7	78.35	4.26
Paper	4,481	25.7	62.0	47.9	16.9	7.6	87.6	79.03	4.26

Test completion rates are over 96% for the three CBT forms and over 87% for the paper form. The average scale scores ranged from 78.32 to 79.03, which are slightly higher than the targeted mean scale score of 77.9 (see Chapter 8, Section 8.4, Procedures for Establishing the Score Scale). The SEM is consistent with the targeted SEM of 1.7 (see Table 13.15 for the SEM for each form). Figure 13.7 presents the level score distributions for these forms (note that for all four forms, the percentages for both the Below Level 3 and the Level 7 groups are lower than 7%).

Figure 13.7. Level Score Distributions for Graphic Literacy



The TCCs and TIFs for the four forms are presented in Figure 13.8 and Figure 13.9, respectively. For comparison, the scaling form is included as the base form (identical to that in Chapter 8, Figure 8.13). Note that these forms were built to meet the same test blueprint as presented in Chapter 3, Section 3.3. The TCCs are placed tightly across the forms, as shown in Figure 13.8.

Figure 13.8. Test Characteristic Curves (TCCs) for Graphic Literacy—Base Form and Four Operational Forms

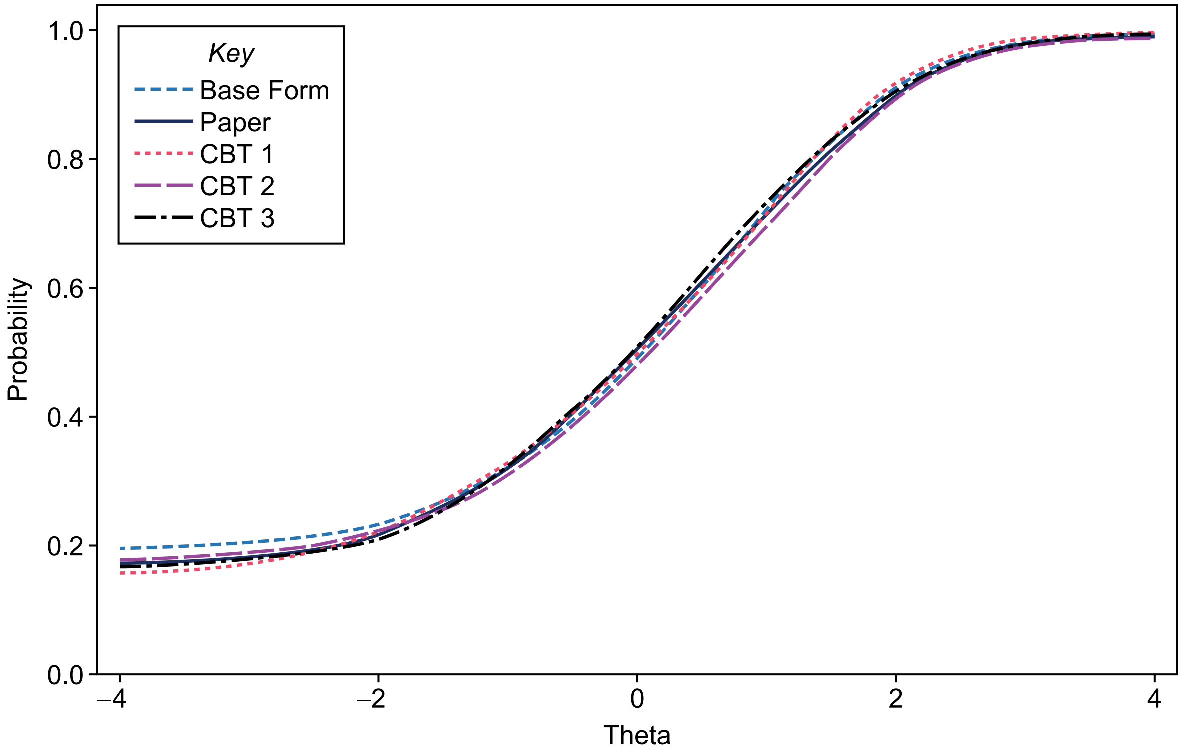
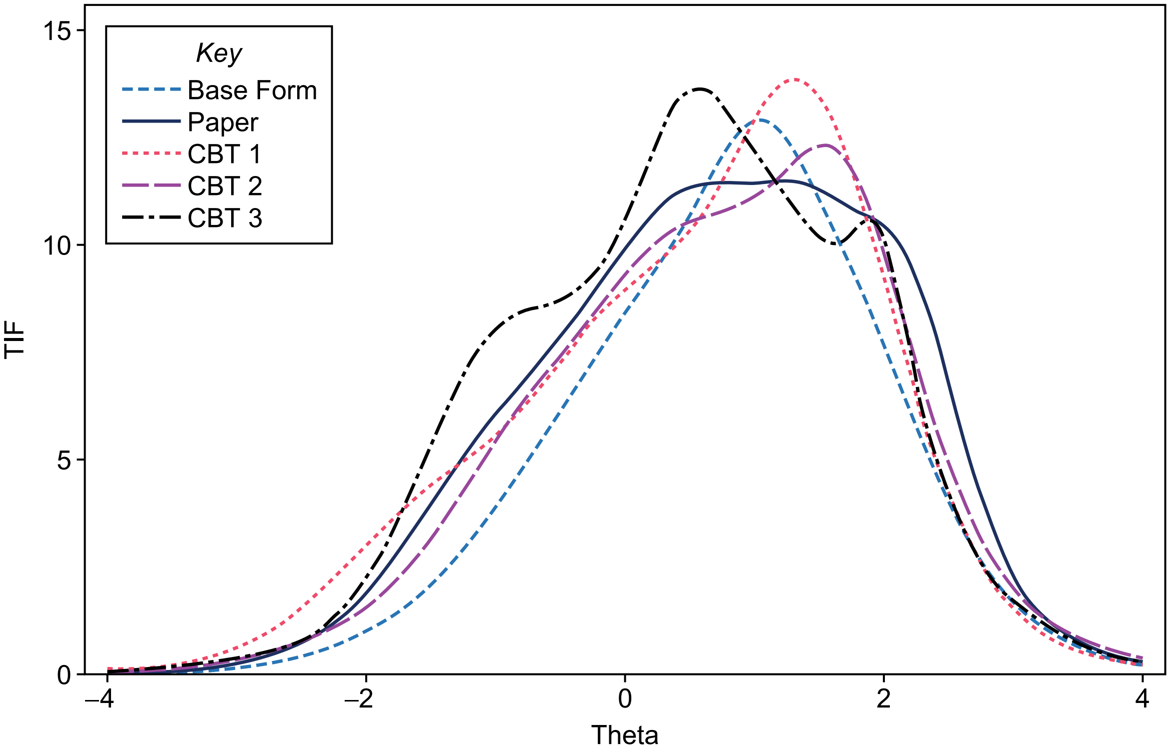


Figure 13.9. Test Information Function (TIF) Curves for Graphic Literacy—Base Form and Four Operational Forms



ACT researchers also monitor DIF for both pretest and operational items using the method presented in Chapter 12, Section 12.4. For the four forms, no operational items were flagged as Group C DIF based on comparisons between female and male, Black and White, and Latinx and White.

13.4.3 Reliability Analyses

The reliability analyses were divided into two parts. The first was based on computing familiar estimates of reliability, including Cronbach's alpha, scale score reliability, and SEMs for the scale scores from the four forms. Cronbach's alpha estimates were .83 or .84 (see Table 13.15), which is slightly lower than .85 for the scaling form (reported in Chapter 10, Section 10.2). For the four forms, the reliability estimates for the scale scores were .84 or .85, and the scale score SEM values ranged from 1.72 to 1.75, which are slightly higher than 1.71 for the scaling form.

Table 13.15. Reliability and SEM Results for Four Graphic Literacy Forms

Form	Cronbach's alpha	Scale score reliability	Scale score SEM
CBT 1	.83	.85	1.72
CBT 2	.84	.85	1.72
CBT 3	.84	.84	1.75
Paper	.84	.85	1.72

The second part of the reliability analyses was based on computing the classification consistency of each WorkKeys level for the four forms. Classification consistency analysis (described in Chapter 10, Section 10.4) was computed for the four forms using item parameter estimates that were used in pre-equating. By comparing Table 13.16 to Chapter 10, Table 10.9, one can observe that the classification consistency results are very stable.

Table 13.16. Estimated Classification Consistency Indices for Level Scores for Four Graphic Literacy Forms

	CBT 1		CBT 2		CBT 3		Paper	
Level	*P* (%)	κ (%)	*P* (%)	κ (%)	*P* (%)	κ (%)	*P* (%)	κ (%)
Exact	51	38	52	39	50	37	51	38
3	93	50	94	44	93	47	93	48
4	85	66	85	66	85	65	86	67
5	84	68	84	67	84	67	84	68
6	88	65	88	64	87	63	87	63
7	95	54	95	55	95	53	95	52

13.4.4 Dimensionality Evaluation

This section provides evidence that the Graphic Literacy assessment is unidimensional based on the same method used in Chapter 9, Section 9.2.5 (the eigenvalue comparisons of the first three factors from the EFA). Table 13.17 presents the EFA results for the CBT 1 form. As shown in Table 13.17, the FDRI value is significantly greater than 3, and the first factor explains 17% of the total variance for the full set of operational items. These findings indicate an underlying single-factor structure of the Graphic Literacy assessment.

Table 13.17. Eigenvalues and Factor Difference Ratio Index (FDRI) for Graphic Literacy—CBT 1 Form

Factor	Eigenvalue	Difference	FDRI
1	5.37 (17.31%)	—	—
2	1.65 (5.34%)	3.71	—
3	1.10 (3.56%)	0.55	6.75

Note. The percentage in each set of parentheses is the percentage of the total variance accounted for by that factor.

In summary, the results in this chapter that were obtained using recent operational data consistently support the findings from the field study and also provide strong evidence of the sound psychometric quality of the WorkKeys Graphic Literacy forms. As additional Graphic Literacy forms are developed according to the WorkKeys Assessment blueprint and statistical guidelines, ACT researchers continue to conduct 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—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 et al., 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 and moving to the skill levels needed to be successful in a career pathway or for specific occupations. Career readiness is defined as having the levels of knowledge, skills, abilities, and other characteristics (KSAOs) needed to be successful in a typical job in a typical organization (see Figure 14.1). Within the context of career readiness, foundational skills are the fundamental, portable skills that are critical to training and workplace success (Symonds et al., 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 needed to be successful in a typical job within a career pathway.

Figure 14.1. General Conceptualizations of College Readiness and Career Readiness

COLLEGE READINESS

DEFINITION: *KSAOs and level of KSAOs needed to succeed in typical courses students take in the first year as a typical college or university*

USE CASES: *Setting national, state, and local educational policies; accountability purposes*

EXAMPLES: *ACT College Readiness Benchmarks*

CAREER READINESS

DEFINITION: *KSAOs and level of KSAOs needed to succeed in a typical job at a typical organization*

USE CASES: *Setting national, state, and local educational and workforce training policies; accountability purposes*

EXAMPLES: *ACT WorkKeys National Career Readiness Certificate levels*

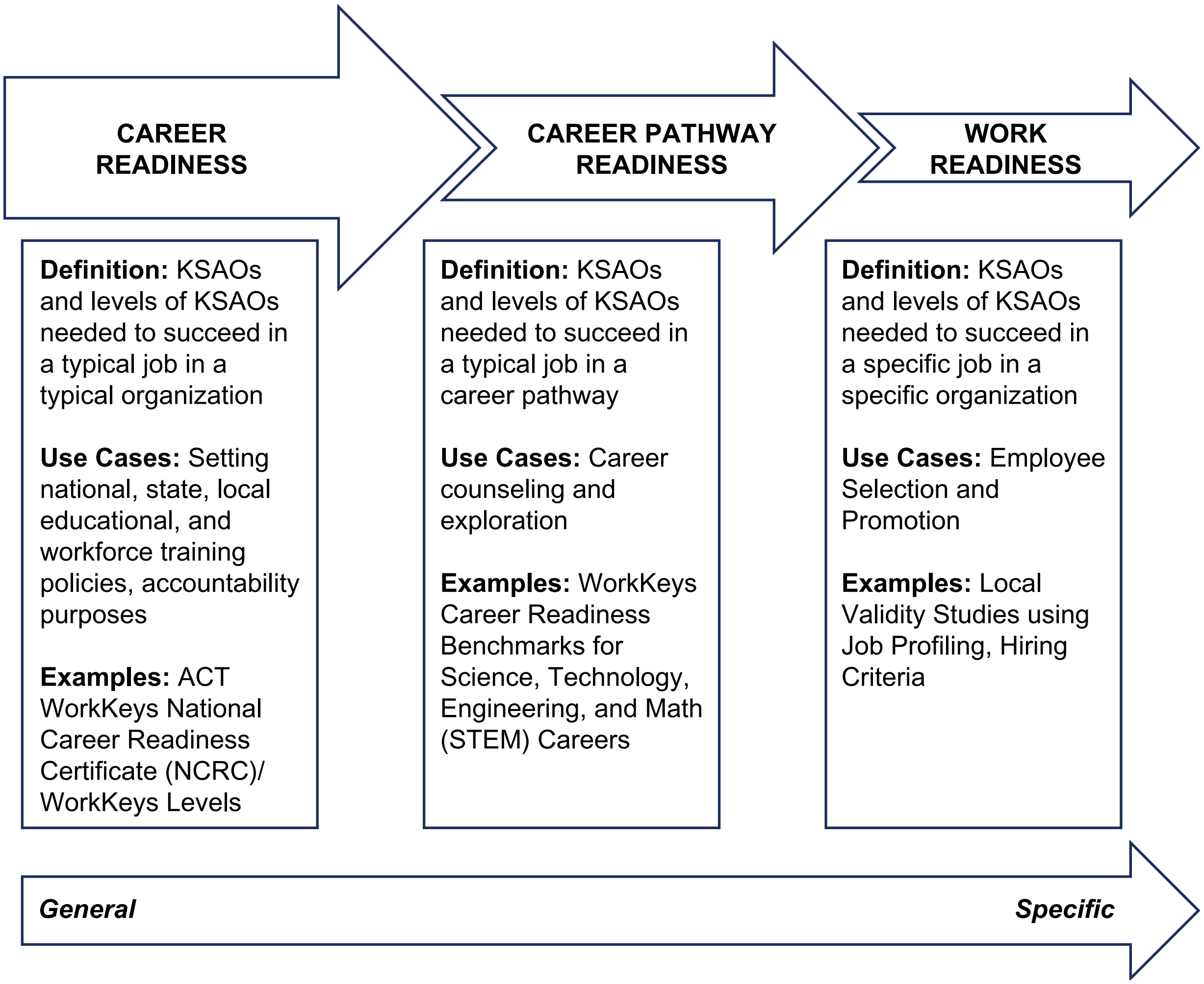
Note. From "A Hierarchical Education and Workplace Readiness Framework" by M. LeFebvre and K. Mattern, 2018, *Ready for What? Development of an Empirical Framework: Linking College Readiness and Career Readiness*, p. 10. Copyright 2018 by ACT, Inc.

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 (Clark et al., 2013). The skills needed for work readiness must (a) be 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 (Clark et al., 2013). 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 to be considered work ready for that occupation. The standards and benchmarks ensure that current and prospective employees' skills are aligned with employers' 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 to determine work readiness by providing individuals with a snapshot of skill requirements for different career pathways (LeFebvre, 2015). Figure 14.2 provides a summary of the work and career readiness definitions and corresponding examples of use cases.

Figure 14.2. Summary of Work and Career Readiness



Definition: KSAOs and levels of KSAOs needed to succeed in a typical job in a typical organization

Use Cases: Setting national, state, local educational, and workforce training policies, accountability purposes

Examples: ACT WorkKeys National Career Readiness Certificate (NCRC)/ WorkKeys Levels

Definition: KSAOs and levels of KSAOs needed to succeed in a typical job in a career pathway

Use Cases: Career counseling and exploration

Examples: WorkKeys Career Readiness Benchmarks for Science, Technology, Engineering, and Math (STEM) Careers

Definition: KSAOs and levels of KSAOs needed to succeed in a specific job in a specific organization

Use Cases: Employee Selection and Promotion

Examples: Local Validity Studies using Job Profiling, Hiring Criteria

The ACT® WorkKeys® National Career Readiness Certificate® (NCRC®) 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 NCRC 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 NCRC assessments and the WorkKeys NCRC.

14.1.1 Personnel Selection and Development

WorkKeys NCRC 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 on the basis of behaviors associated with job performance, (c) employee development, and (d) development of the appropriate level of fit with occupations in terms of interests (LeFebvre, 2016).

When WorkKeys NCRC 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 NCRC 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 NCRC 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.

14.1.2 Workforce and Economic Development

The WorkKeys NCRC assessments and the WorkKeys NCRC are widely used in workforce and economic development programs, such as by (a) an employer who uses the WorkKeys NCRC assessment or the WorkKeys NCRC and other criteria to identify a qualified pool of applicants and requires a specific level of WorkKeys NCRC or WorkKeys NCRC 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 who use the WorkKeys NCRC to document an individual’s level of essential work readiness skills; and (d) states, communities, or schools who use the WorkKeys NCRC to document the aggregate career readiness of a community, region, or state.

ACT® Work Ready Communities 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 NCRC assessments and the WorkKeys NCRC to measure foundational workplace skills with goals established for the current, emerging, and transitioning workforce. 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 populations of the workforce (ACT, 2018). Skill gaps across various sectors of the workforce can be identified and addressed by state or local community policies and programs.

14.1.3 Accountability

State accountability systems, such as Career and Technical Education programs, have incorporated WorkKeys NCRC assessments and the WorkKeys NCRC as a measure of employability skills or career readiness (McMurrer & Frizzell, 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, 2013). Some states also report using WorkKeys NCRC assessment results as a requirement for graduation, for receipt of a career or technical diploma, as an endorsement on a standard diploma, or for scholarship eligibility.

As noted in Chapter 1, “the National Reporting System (NRS) is the accountability system for the federally funded, State-administered adult education program. It embodies the accountability requirements of the Workforce Innovation and Opportunity Act (WIOA, the Act) for the adult education and literacy program (Title II) and reporting under WIOA” (Division of Adult Education and Literacy et al., 2021, p. 1). WorkKeys Applied Math and Workplace Documents assessments are approved for use by NRS, and authorized by WIOA.

References

ACT. (2008). *WorkKeys® Locating Information technical manual*.

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

ACT. (2013). *Skills credentials aid displaced manufacturing workers in Ohio: Case study*. https://www.workreadycommunities.org/resources/case_studies/MSSC_case_study.pdf

ACT. (2014). *Foundational skills: What makes a skill "foundational"?* https://www.act.org/content/dam/act/unsecured/documents/WK-Brief-KeyFacts-FoundationalSkills.pdf

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

ACT. (2016b). *ACT WorkKeys® administration manual: Online testing*. https://www.act.org/content/dam/act/unsecured/documents/IV_UserGuide.pdf

ACT. (2018). *ACT® Work Ready Communities: Common criteria*. https://www.workreadycommunities.org/resources/WRC0003_Work_Ready_Communities_Common_Criteria_WEB.pdf

ACT. (2021). ACT® WorkKeys® NCRC® crosswalk to college and career readiness standards for adult education. https://www.workreadycommunities.org/files/pdf/WorkKeys%20NCRC%20Crosswalk%20to%20CCRSAE%20Report%20&%20Appendices%201.6.22.pdf

ACT. (2022a). *ACT® WorkKeys® administration manual: Paper testing*. https://www.act.org/content/dam/act/unsecured/documents/WorkKeysAdministrationManualforPaperTesting.pdf

ACT. (2022b). *Accessibility supports guide for ACT WorkKeys National Career Readiness Certificate (NCRC)*. https://www.act.org/content/dam/act/unsecured/documents/WorkKeysAccessibilitySupportsGuide.pdf

ACT. (2023). *Occupational profile search* [database]. http://jobprofiles.act.org/

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

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 technical report*. National Center for Education Statistics.

American Educational Research Association, American Psychological Association, & National Council for Educational Measurement (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Association for Career and Technical Education. (2018). *What is “career ready”?* https://www.acteonline.org/wp-content/uploads/2018/03/Career_Readiness_Paper_COLOR.pdf

Australian Association of Mathematics Teachers. (2014). *Workplace maths doesn’t add up*. https://aamt.edu.au/about-us/news-and-media/media-release/

Avgerinou, M. D., & Pettersson, R. (2015). Toward a cohesive theory of visual literacy. *Journal of Visual Literacy*, *30*(2), 1–19. 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.* University of Iowa.

Bartels, M., & Marshall, S. P. (2012, March 28–30). Measuring cognitive workload across different eye tracking hardware platforms. In *Proceedings of the 2012 symposium on eye tracking research and applications* [Symposium, pp. 161–164]. Eye Tracking Research & Applications 2012, Santa Barbara, CA, United States.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292.

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

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

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. (Eds.) (2015). *Beyond academics: A holistic framework for enhancing education and workplace success*. ACT. http://www.act.org/content/dam/act/unsecured/documents/ACT_RR2015-4.pdf

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

Center for Applied Special Technologies. (2011). *Universal design for learning guidelines version 2.0*. https://wvde.state.wv.us/osp/UDL/4.%20Guidelines%202.0.pdf

Clark, H., LeFebvre, M., Burkum, K., & Kyte, T. (2013) *Work readiness standards and benchmarks: The key to differentiating America's workforce and regaining global competitiveness*. ACT.

Conway, J. (2022). *WorkKeys Differential Prediction* [Manuscript in preparation]. ACT.

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. I. Braun (Eds.), *Test validity* (pp. 3–17). Lawrence Erlbaum.

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

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

Division of Adult Education and Literacy, Office of Career, Technical, and Adult Education, & U.S. Department of Education. (2021). *Technical assistance guide for performance accountability under the Workforce Innovation and Opportunity Act: National reporting system for adult education*. https://nrsweb.org/sites/default/files/NRS-TA-Mar2021-508.pdf

Djamasbi, S. (2014). Eye tracking and web experience. *AIS Transactions on Human-Computer Interaction*, *6*(2), 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). Lawrence Erlbaum Associates.

Dorans, N. J., Pommerich, M., & Holland, P. W. (Eds.). (2007). *Linking and aligning scores and scales*. Springer Science + Business Media.

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

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Adoption by four agencies of Uniform Guidelines on Employee Selection Procedures (1978). *Federal Register 43*(166), 38290–38315. https://www.ojp.gov/ncjrs/virtual-library/abstracts/employee-selection-procedures-adoption-four-agencies-uniform

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

Friel, S. N., & Bright, G. W. (1996, April 8–12). *Building a theory of graphicacy: How do students read graphs?* [Paper presentation]. Annual Meeting of the American Educational Research Association, New York, NY, United States. 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*(2), 124–158.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests.* Wiley.

Gottfredson, L. S. (1988). Reconsidering fairness: A matter of social and ethical priorities. *Journal of Vocational Behavior*, *33*(3), 293–319.

Greene, B. B. (2008). *Perceptions of the effects of the WorkKeys system in North Carolina*. Dissertation Abstracts International, *69*(12), 1373.

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

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications.* Springer Science + Business Media.

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

Hattie, J. A. (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 preemployment assessment. *Community College Journal of Research and Practice*, *35*(11), 895–908.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Routledge.

Human Resources and Skills Development Canada. (2004). *Comparing classroom and workplace reading* [Unpublished manuscript].

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).* 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*(1), 28–42.

Infosys. (2016). *Amplifying human potential: Education and skills for the fourth industrial revolution.* http://hdl.voced.edu.au/10707/396524

Institute for the Future. (2011). *Future work skills: 2020*. University of Phoenix Research Institute. https://legacy.iftf.org/futureworkskills/

International Organization for Standardization/International Electrotechnical Commission. (2013). *ISO/IEC 27001 information security management systems*. https://www.iso.org/isoiec-27001-information-security.html

International Organization for Standardization/International Electrotechnical Commission. (2018). *ISO/IEC 27005 Information technology—Security techniques—Information security risk management*. https://www.iso.org/standard/75281.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 J. Hyona, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–606). North-Holland.

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*. Springer.

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

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

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

Kolen, M. J. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement*, *25*(2), 97–110.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer.

Kolen, M. J., Hanson, B. A., & Brennan, R. L. (1992). Conditional standard errors of measurement for scale scores. *Journal of Educational Measurement*, *29*(4), 285–307.

Koomey, J. G. (2017). *Turning numbers into knowledge: Mastering the art of problem solving* (3rd ed.). Analytics Press.

Langenfeld, T. (2014). *ACT WorkKeys: Awarding college credit through the ACT® National Career Readiness Certificate.* ACT.

Lapan, R. T., Hinkelman, J. M., Adams, A., & Turner, S. (1999). Understanding rural adolescents' interests, values, and efficacy expectations. *Journal of Career development*, *26*(2), 107–124.

LeFebvre, M. (2015). *Career readiness in the United States 2015.* ACT.

LeFebvre, M. (2016). *A summary of ACT WorkKeys® validation research*. ACT.

LeFebvre, M., & Mattern, K. (2018). *Ready for what? Development of a hierarchical framework linking college readiness and career readiness*. ACT. https://www.act.org/content/dam/act/unsecured/documents/Ready-for-What-May-2018.pdf

Levy, F., & Murnane, R. J. (2004). *The new division of labor: How computers are creating the next job market.* Princeton University Press.

Lewis, D. M., Mitzel, H., 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 ed., pp. 225–253). Routledge.

Linn, R. L. (1993). Linking results in distinct assessments. *Applied Measurement in Education*, *6*(1), 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). Routledge.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Routledge.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equating." *Applied Psychological Measurement*, *8*(4), 453–461.

Manpower Group (2015). *The talent shortage*. https://go.manpowergroup.com/talent-shortage

Marshall, S. P. (2002, September 15–19). The index of cognitive activity: Measuring cognitive workload. In J. J. Persensky, B. Hallbert, & H. Blackman (Eds.), *New century, new trends: Proceedings of the 2002 IEEE 7th conference on human factors and power plants* [Conference, pp. 75–79]. Scottsdale, AZ, United States.

Matos, R. (2010). Designing eye tracking experiments to measure human behavior [PowerPoint slides]. Tobil Technology. https://measuringbehavior.org/mb2010/files/tutorials/Tobii_MB2010_tutorial_handouts.pdf

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

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

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

McMurrer, J., & Frizzell, M. (2013). *Career readiness assessments across states: A summary of survey findings*. Center on Education Policy.

Mele, M. L., & Federici, S. (2012). Gaze and eye-tracking solutions for psychological research. *Cognitive processing*, *13 Suppl 1*, 261–265.

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

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

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

Mislevy, R. J., Almond, R. G., & Lukas J. F. (2004). *A brief introduction to evidence-centered design*. National Center for Research on Evaluation, Standards, and Student Testing. https://www.yumpu.com/en/document/view/40872213/a-brief-introduction-to-evidence-centered-design-cse-report-632

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design.* 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*(1), 194–206.

Mroch, A. A., Li, D., & Thompson, T. D. (2015, April 15–19). *A framework for evaluating score comparability* [Paper presentation]. Annual Meeting of the National Council On Measurement In Education, Chicago, IL, United States.

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

National Institute of Standards and Technology. (2017). *Computer security resource center* (Special Publication 800-37). https://csrc.nist.gov/publications/sp800

National Network of Business and Industry Associations. (2014). *Common employability skills: A foundation for success in the workplace: The skills all employees need, no matter where they work*. http://businessroundtable.org/sites/default/files/Common%20Employability_asingle_fm.pdf

Nicol, C. (2002). Where's the math? Prospective teachers visit the workplace. *Educational Studies in Mathematics*, *50*(3), 289–309.

Obersteiner, A., Moll, G., Beitlich, J. T., Cui, C., Schmidt, M., Khmelivska, T., & Reiss, K. (2014, July 15–20). Expert mathematicians' strategies for comparing the numerical values of fractions—Evidence from eye movements. In P. Liljedahl, S. Oesterle, C. Nicol, & D. Allan (Eds.), *Proceedings of the joint meeting of PME 38 and PME-NA 36*, *Vol. 4*. [Conference, pp. 338–344]. Vancouver, Canada.

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

PCI Security Standards Council. (2022). *Payment card industry data security standard (PCI DSS)*. (Version 3.2.1.) https://www.pcisecuritystandards.org/document_library/?document=pci_dss

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology*, *60*(2), 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). AAAI Press.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*(2), 297–330.

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*(4), 215–227.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, *49*(11), 929–954.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*(5), 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*(2), 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*(2), 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]. https://home.ubalt.edu/tmitch/645/session%204/Schmidt%20&%20Oh%20MKUP%20validity%20and%20util%20100%20yrs%20of%20research%20Wk%20PPR%202016.pdf

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].

Schultz, D., & Stern, S. (2013). *College and Career Ready? Perceptions of High School Students Related to WorkKeys Assessments*. https://www.ingentaconnect.com/content/acter/cter/2013/00000038/00000002/art00007

Schulz, E. M., Kolen, M. J., & Nicewander, W. A. (1997). *A study of modified-Guttman and IRT-based level scoring procedures for Work Keys assessments.* ACT.

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*(4), 347–362.

Schulz, E. M., & Mitzel, H. C. (2005, April 11–15). *The Mapmark standard setting method* [Paper presentation]. Annual Meeting of the National Council on Measurement and Education, Montreal, Canada.

Shah, P., & Freedman, E. G. (2011). Bar and line graph comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science*, *3*(3), 560–578. 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*(2), 77–105.

Smith, J. P., III. (1999). Tracking the mathematics of automobile production: Are schools failing to prepare students for work? *American Educational Research Journal*, *36*(4), 835–878.

Smith, M. C., Mikulecky, L., Kibby, M. W., Dreher, M. J., & Dole, J. A. (2000). What will be the demands of literacy in the workplace in the next millennium? *Reading Research Quarterly*, *35*(3), 378–383.

Society for Human Resource Management (2012). *Critical employee skills for the changing workforce*. 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 Resource Management (2022). *Avoiding adverse impact in employment practices*.

Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (5th ed). https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf

Steedle, J. T., & Hepburn, A. (2020). *The ACT WorkKeys NCRC as an indicator of skills needed for success in work-based learning*. ACT.

Steedle, J. T., & LeFebvre, M. (2018). *Income trends for ACT NCRC earners*. ACT. https://www.act.org/content/dam/act/unsecured/documents/R1714-income-trends-ncrc-2018-08.pdf

Steedle, J. T., Ndum, E., & Mattern, K. (2017). *The ACT® National Career Readiness Certificate® as a predictor of academic and workforce outcomes*. ACT. https://www.act.org/content/dam/act/unsecured/documents/pdfs/R1663-act-ncrc-outcomes-predictor-2017-12.pdf

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, *7*(2), 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. 157–180). 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). The Johns Hopkins University Press.

Symonds, W. C., Schwartz, R., & Ferguson, R. F. (2011). *Pathways to prosperity: Meeting the challenge of preparing young Americans for the 21st century*. Pathways to Prosperity Project, Harvard University Graduate School of Education. https://dash.harvard.edu/bitstream/handle/1/4740480/Pathways_to_Prosperity_Feb2011-1.pdf

Thomas, J., & Langenfeld, T. (2017, April 26–30). *Analyzing think-aloud and eye-tracking data to support score interpretations* [Paper presentation]. Annual Meeting of the National Council on Measurement in Education, San Antonio, TX, United States.

Van Aken, J. E., & Romme, A. G. L. (2012). A design science approach to evidence-based management. In D. M. Rousseau (Ed.), *The Oxford handbook of evidence-based management* (pp. 43–57). Oxford University Press.

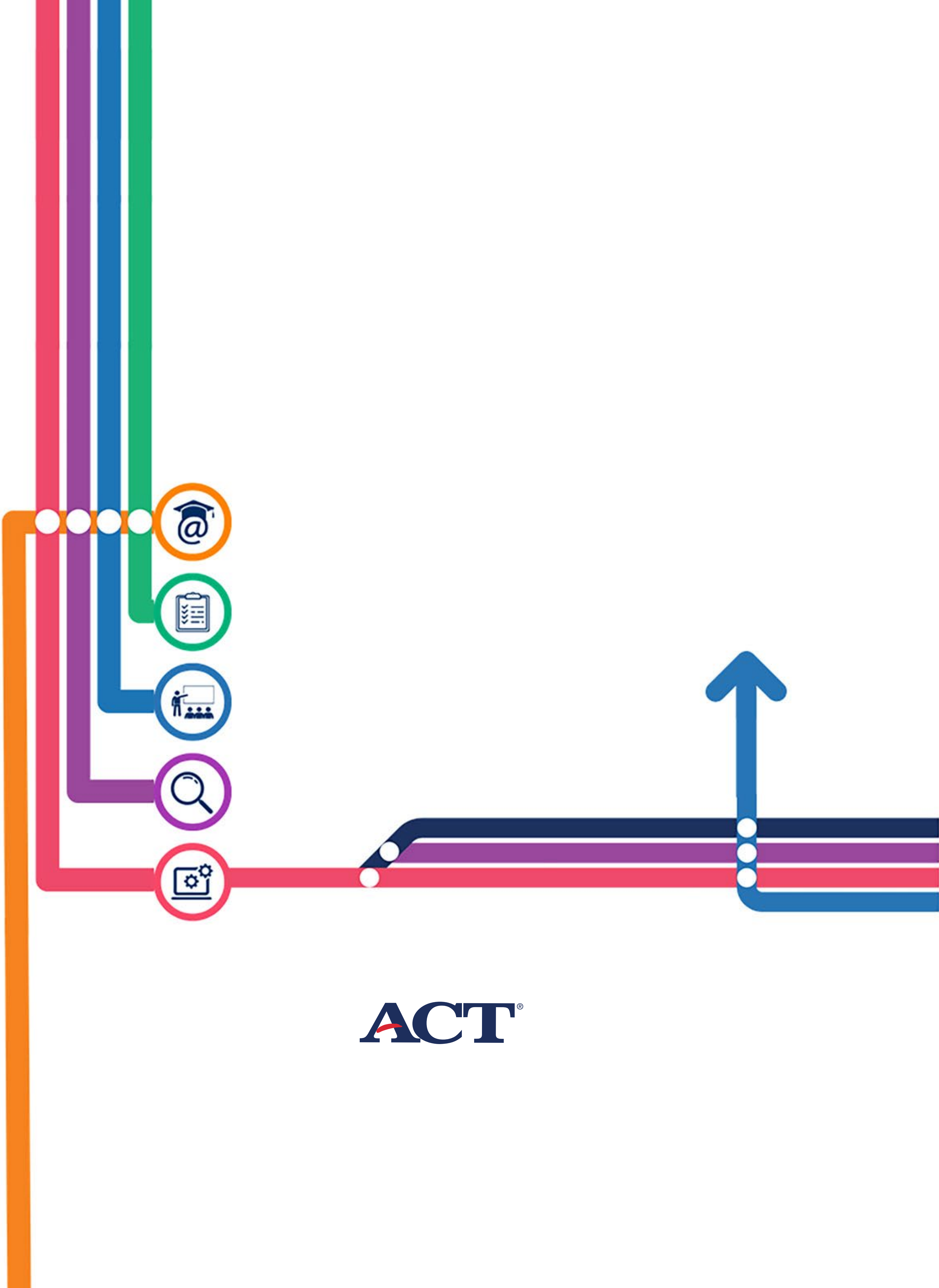
Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press.

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

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

Wolodtschenko, A., & Forner, T. (2007). Prehistoric and early historic maps in Europe: Conception of Cd-atlas. *ePerimetron*, *2*(2), 114–116. 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*(2), 125–145.



ACT®