A Summary of ACT WorkKeys® Validation Research

Mary LeFebvre

ACT®

Mary LeFebvre is a principal research scientist specializing in workforce research, policy evaluation, and competency supply/demand analysis.

Acknowledgments

Contents

Abstract

ACT WorkKeys® assessments and the ACT National Career Readiness Certificate™ (ACT NCRC®) are measures of cognitive foundational workplace skills and are used for a variety of purposes, some of which involve high-stakes decisions. This report summarizes validity evidence to date for these uses in a manner that is responsive to the *Uniform Guidelines on Employee Selection Procedures* (1978), the *Standards for Educational and Psychological Testing* (2014), and the *Principles for the Validation and Use of Personnel Selection Procedures* (2003). For organizational purposes, the report is divided into five main sections: (1) an overview of the ACT WorkKeys assessments and the ACT NCRC, (2) construct validity evidence, (3) content validity evidence, (4) criterion validity evidence, and (5) discussion. Attempts are also made to include both published and unpublished sources of ACT WorkKeys validity research. Additional information can be found in specific research reports and technical manuals for each assessment.

Context for Validation

ACT WorkKeys assessments and the ACT NCRC are measures of cognitive foundational workplace skills. The individual assessments and associated credential are used for a variety of purposes, some of which involve high-stakes decisions. This report summarizes validity evidence to date for these uses.

Conceptually, validity is the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests (American Educational Research Association [AERA] et al., 2014). Although the *Uniform Guidelines on Employee Selection Procedures* (1978), the *Standards for Educational and Psychological Testing* (2014), and the *Principles for the Validation and Use of Personnel Selection Procedures* (2003) describe validity in slightly different terms, this report summarizes three traditional sources of evidence—construct, criterion, and content.

There are three ACT WorkKeys assessments which were primarily developed for use in employment selection or employee development. Additional uses of the individual assessments include (1) establishing cutoffs for entry into occupational/career training programs and (2) measuring the effectiveness of education and workforce training programs in developing foundational work readiness skills. The primary use of the ACT NCRC is to document an individual's level of foundational work readiness important for success in a variety of careers. Employers and industry associations use this credential as a measure of foundational skills. In this way, the ACT NCRC serves as one of several qualifications for entry in a specific job or career path. The credential has also been used as an aggregate measure of a population's readiness for "work" or "career" (e.g., an unemployed adult population receiving career services, a census of high school students taking the credential to meet state or local college and career readiness standards).

The *Uniform Guidelines* (Equal Employment Opportunity Commission [EEOC], 1978) were developed to establish a common set of principles to assist employers, labor organizations, employment agencies, licensing and certification boards, and others to comply with federal law related to the use of tests and other selection procedures. Specifically, they address the use of selection procedures when adverse impact exists. While validity studies are not mandated when adverse impact is absent, they encourage the use of procedures with appropriate validation evidence to support the proposed use of tests and other selection procedures.

ACT assessments, including ACT WorkKeys, have continued to be developed, used, and supported in a manner that strives to adhere to the professional standards and best practices reflected in the *Standards* (AERA et al., 2014) and the *Principles* (Society for Industrial and Organizational Psychology [SIOP], 2003).

It is the responsibility of the test developer to identify the appropriate use of its tests, as well as to warn test users against inappropriate uses when these can be anticipated. The test developer and test users have additional responsibilities, which are addressed in the *Uniform Guidelines*, the *Standards*, and the *Principles*. These responsibilities include providing evidence and a theoretical basis to support the intended uses and interpretation of scores (AERA et al., 2014). The *Uniform Guidelines* were last revised nearly forty years ago, and there have been significant changes in the scientific methods and professional practices related to the design, development, use, and validation of employment tests. When tests are used explicitly for employment decisions, the *Uniform Guidelines* are relevant and are in all instances the professional and technical standards to be practiced and used. As noted in the *Standards*, not each Standard is applicable to all tests and all uses; professional judgment is essential for determining and documenting how test developers and test users have addressed the relevant standards. A rationale and evidence may be presented when addressing Standards or, similarly, when there is a determination that one or more Standards are not applicable. Tests can often be used for additional purposes beyond the original intended use. For example, the ACT® test has been primarily used as a national college admissions test, but additional evidence has been provided to support its use in college course placement, as an indicator of college readiness, and in contributing to the evaluation of schools in preparing students for college. Each additional purpose needs additional lines of evidence, which are often provided by parties other than the test developer (e.g, test user, other researchers) (ACT, 2014a).

This research review summarizes validation evidence in a manner responsive to the *Uniform Guidelines*, the *Standards*, and the *Principles*. The *Uniform Guidelines* define three types of validity studies that are acceptable for providing evidence of validity for a test used in selection procedures: construct, content, or criterion-related validity. The *Standards* and the *Principles*, however, consider validity to be a unitary concept; they do not focus on the distinction between the types of validity, but rather on the sources of validity evidence for a particular use. For organizational purposes, this summary is divided into five main sections: (1) an overview of the ACT WorkKeys assessments and the ACT NCRC, (2) construct validity evidence, (3) content validity evidence, (4) criterion validity evidence, and (5) discussion. Attempts are also made to include both published and unpublished sources of ACT WorkKeys validity research. When unpublished sources are cited, detailed results are provided when available. Additional data for published studies can be found in the referenced sources.

Overview of ACT WorkKeys Assessments

The ACT WorkKeys cognitive and non-cognitive assessments currently comprise Reading for Information, Applied Mathematics, Locating Information, Applied Technology, Listening for Understanding, Teamwork, Business Writing, Fit, Performance, Talent, and Workplace Observation assessments, as well as the ACT NCRC. Many of the current assessments are under revision or review by ACT.

Seven of the cognitive assessments are exclusively multiple-choice exams and one is a constructed response assessment. One of the cognitive assessments, Listening for Understanding, presents audio information via speakers. Two of the three ACT WorkKeys behavioral assessments are norm-referenced assessments in which an examinee receives percentile rank scores. Each cognitive assessment includes scale scores that correspond to a set of defined skill levels in a content domain.

In addition, the ACT WorkKeys cognitive assessments (with the exception of Workplace Observation) include levels which were conceptualized as an independent definition of the construct to be measured, a definition not based on the psychometric properties of the assessment (McLarty & Vansickle, 1997). In this way, job analysis can be linked with the assessment of individuals so that both could address the same skill level. Table 1 provides a summary of technical aspects for each assessment.

ACT WorkKeys assessments can help employers identify a pool of qualified applicants who have achieved the levels of proficiency needed to perform a job as determined through job analysis. The assessments should be used in combination with additional measures (e.g., tests, interviews, other selection procedures) that the employer deems appropriate and relevant for pre-employment selection or other employment decisions (ACT, 2012a).

Table 1. Summary of ACT WorkKeys Assessments' Characteristics

Test Title	Construct	Content	Item Type	Total Number of Items	Score Type	Timing	Format	Languages
ACT WorkKeys Cognitive Assessments								
Applied Mathematics	Applying mathematical reasoning, critical thinking, and problem-solving techniques to work-related problems.	The test questions require examinees to set up and solve problems and do the types of calculations that occur in the workplace.	Multiple Choice	33	Levels (3–7) Scale Score (65–90)	45 min. (Paper) 55 min. (Online)	Paper and Online	English and Spanish
Reading for Information	Reading and using written text to do a job.	The written texts include memos, letters, directions, signs, notices, bulletins, policies, and regulations.	Multiple Choice	33	Levels (3–7) Scale Score (65–90)	45 min. (Paper) 55 min. (Online)	Paper and Online	English and Spanish
Locating Information	Working with workplace graphics.	Examinees are asked to find information in a graphic or insert information into a graphic. They also must compare, summarize, and analyze information found in related graphics.	Multiple Choice	38	Levels (3–6) Scale Score (65–90)	45 min. (Paper) 55 min. (Online)	Paper and Online	English and Spanish
Business Writing	Writing an original response to a work-related situation.	Components include sentence structure, mechanics, grammar, word usage, tone and word choice, organization and focus, and development of ideas.	Constructed Response	1 prompt	Levels (1–5) Scale Score (50–90)	30 min. (Paper) 30 min. (Online)	Paper and Online	English Only

Table 1. (continued)

Test Title	Construct	Content	Item Type	Total Number of Items	Score Type	Timing	Format	Languages
Workplace Observation	Observing, following, understanding, and evaluating processes, demonstrations, and other on-the-job procedures.	The test measures examinees' ability to focus and notice what they are observing. Examinees may also be asked to follow, interpret, synthesize, analyze, or evaluate what they have observed.	Multiple Choice	35	Levels (1–5)	55 min.	Online	English Only
Applied Technology	Solving problems with machines and equipment found in the workplace.	Test content is related to four areas of technology: electricity, mechanics, fluid dynamics, and thermodynamics. Examinees are asked to analyze and solve a problem and apply existing tools, materials, or methods to new situations.	Multiple Choice	34	Levels (3–6) Scale Score (65–90)	45 min. (Paper) 55 min. (Online)	Paper and Online	English and Spanish
Listening for Understanding	Listening skills and the ability to understand and follow directions in the workplace.	Examinees listen to spoken information and answer multiple-choice questions about the information they heard. The test includes spoken information using short, simple sentences and information that is longer, more complex, and indirectly stated.	Multiple Choice	28	Levels (1–5) Scale Score (50–90)	45 min.	Online	English Only
Teamwork	Choosing behaviors that both lead toward the accomplishment of work tasks and support the relationships between team members.	The test contains twelve video teamwork scenarios, each accompanied by three multiple-choice items. Examinees must identify the most appropriate teamwork responses to specific situations.	Multiple Choice	36	Levels (3–6) Scale Score (65–90)	64 min.	Paper	English Only
ACT WorkKeys Non-Cognitive Assessments								
Fit	Self-reported work-related interests and values of examinees to the corresponding activities and characteristics of occupations.	The test contains Likert Scale items asking examinees to provide information on their level of interest and the importance of work-related values.	Likert Scale: Interest Inventory (Dislike, Indifferent, Like); Work Values 5-point scale (Not Important—Extremely Important).	100	Fit Index using percentile scores from 1–99 based on a norm sample. Also provides Level of Fit Index (High, Moderate, Low). Values (High, Moderate, and Low). Interest Inventory from 20–80 based on a normative sample.	20 min.	Online	English Only

Table 1. (continued)

Test Title	Construct	Content	Item Type	Total Number of Items	Score Type	Timing	Format	Languages
Performance	General work behaviors that might be problematic (e.g., absenteeism, theft, violation of work rules, hostility in the workplace, general work attitude, and conduct issues). It includes an overall index score and two scale scores.	The test contains Likert Scale items asking examinees to indicate the extent to which they agree that a behavior, attitude, or personality-related item describes them.	Likert Scale: 6-point scale (Strongly Disagree—Strongly Agree).	55	Performance Index score from 1–99.[1] Level of Desirability (High, Moderate, and Low) based on Index Score Range. General Work Attitude and Risk Reduction subscale scores (1–99).	15 min.	Online	English Only
Talent	Twelve work-related personality characteristics. The assessment is based on the facets of the Five Factor Model of personality, as well as concepts from the emotional intelligence literature. The four indices are intended to more directly target prediction of important work outcomes.	The test contains Likert Scale items asking examinees to indicate the extent to which they agree that a behavior, attitude, or personality related item describes them.	Likert Scale: 6-point scale (Strongly Disagree—Strongly Agree)	165	Percentile rank scores for four indices and twelve scales.[2] Indices: ▪ Teamwork ▪ Work Discipline ▪ Managerial Potential ▪ Customer Service Orientation Scales: ▪ Carefulness ▪ Cooperation ▪ Creativity ▪ Discipline ▪ Goodwill ▪ Influence ▪ Optimism ▪ Order ▪ Savvy ▪ Sociability ▪ Stability ▪ Striving	35 min.	Online	English Only

[1] The Performance Index is based on a combination of scores from General Work Attitudes and Risk Reduction subscales. Percentiles were derived from raw scores, which range from 1–99, using norms obtained from the field study sample when the test was developed.

[2] Percentile rank scores range from 1–99 for each of the twelve Talent scales. Percentiles derived from raw scores using norms obtained from an operational sample collected after the test was released. Subsets of the development group for Talent were used to develop the Talent index scores.

ACT NCRC

The ACT NCRC is a portable credential that measures essential foundational work readiness skills. Foundation skills are defined as the fundamental, portable skills that are critical to training and workplace success (Clark, 2015). The ACT NCRC comprises Applied Mathematics, Locating Information, and Reading for Information. These three skills have been consistently identified as important for success in a broad range of jobs, making them "essential" foundational skills for career readiness.[3] The ACT NCRC is issued in four levels based on examinee level scores on the component tests:

Certificate Level	Level Score
Platinum	6 or above on all three tests
Gold	5 of above on all three tests
Silver	4 or above on all three tests
Bronze	3 or above on all three tests

In 2003, several states originated the ACT NCRC's predecessor, the Career Readiness Certificate (CRC) (Bolin, 2005). The decision to include the Reading for Information, Applied Mathematics, and Locating Information ACT WorkKeys assessments for the state-issued CRCs, as well as defining the certificate levels as being any combination of the lowest of the three assessment level scores, was made independently by the states. In 2007, ACT began issuing a national version of the certificate using the same ACT WorkKeys assessments and levels.

The ACT NCRC is intended to be used in a variety of ways:

- Employers may recommend that applicants provide evidence of their ACT NCRC in addition to traditional criteria (e.g., employment application, credentials, interview) as part of the job application process. In these instances, employers may then use the ACT NCRC and the other criteria to screen prospective applicants for hiring. In this scenario, the employer is using the ACT NCRC and other criteria to identify a qualified pool of applicants[4] and is not requiring a specific level of the ACT NCRC.

- Employers may use the ACT NCRC to make employment decisions and require a specific level (e.g., gold). In such instances, a formal job profile should be conducted. When available, such evidence may also be transported from job analysis studies of similar positions requiring the same skills. In this situation, the ACT NCRC level is used to partially determine whether an applicant is qualified, and additional evidence that the specified level represents skills required on the job should be documented.

- States, communities, and schools may use the ACT NCRC to document an individual's level of essential work readiness skills. Specifically, state and local workforce and education agencies often provide ACT NCRC testing for individuals to document their work readiness for a potential job opportunity or career path. In this scenario, states, communities, or institutions use ACT NCRC testing to assist individuals with improving their work readiness skills.

---

[3] Analysis of the ACT JobPro® database has found that the three ACT WorkKeys cognitive skill areas—Reading for Information, Locating Information, and Applied Mathematics—are most often determined via the job profiling process to be important for job and task performance (ACT, 2011a).

[4] As noted above, reading, math, and locating information are included in the ACT NCRC based on evidence that the majority of profiled careers and career clusters require these skills. However, organizations employing the ACT NCRC for pre-employment screening or selection decisions should verify that these essential skills are important for their positions or careers. This can be done using a variety of methods and does not mandate a formal job analysis or job profiling study.

- Finally, states, communities or schools may use the ACT NCRC to document the aggregate "work readiness" of a community, region or state. As part of a Work Ready Community initiative, education, economic, and workforce partners provide ACT NCRC testing to gauge the work readiness skill levels for different segments of the local labor market. Used in this manner, the ACT NCRC can help workforce and education stakeholders identify skill gaps for a region or state for individuals in both the emerging and current labor force.

Construct Validation Evidence

Overview

"Construct" refers not only to the structure of a test but, more importantly, refers to the overarching attribute that it claims or purports to measure, such as reading ability or conscientiousness. Construct validity evidence often focuses on the test score as an accurate and thorough measure of a designated attribute. The process of compiling construct evidence starts with test development. It continues with research conducted to observe the relationships between test scores and other variables, until a pattern is detected that clearly indicates the meaning of the test score. In other words, the evidence confirms that the test effectively measures the construct—or attribute—it was meant to measure (AERA et al., 2014; EEOC, 2000; SIOP, 2003).

Embedded within the concept of construct validity are convergent and discriminant validity. Convergent validity evidence investigates the relationship between test scores and other measures designed to measure similar constructs. Discriminant validity, on the other hand, investigates the relationship between test scores and measures of different constructs. For example, a strong correlation between a new test developed to measure spatial ability and an existing measure of spatial ability would provide convergent evidence, as would the correlation between ACT mathematics scores and grades in college mathematics courses. Divergent evidence would include studies that show a test of spatial ability has a stronger relationship with an existing test of spatial ability than it does with tests of unrelated constructs such as manual dexterity, reasoning, or strength. Divergent evidence is produced when research finds the correlation of ACT mathematics test scores to college mathematics grades is stronger than the correlation to English course grades. Alignment studies, or crosswalks, provide additional construct evidence by evaluating the correspondence between a skill or learning standard and the content outline and its items.

Creation of ACT WorkKeys Skills

In developing the ACT WorkKeys assessments, ACT consulted with workforce developers, employers, and educators to identify foundational workplace skills that (1) are used in a wide range of jobs, (2) could be taught in a short period of time, and (3) could be determined through job analysis (ACT, 1992). Initial ACT WorkKeys skills were selected using data from high demand skills identified by employers (Agency for Instructional Technology, 1989; ACT, 1987; Bailey, 1990; Carnevale et al., 1990; Center for Occupational Research and Development, 1990; Conover Company 1991; Educational Testing Service, 1975; Electronic Selection Systems Corporation, 1992; Greenan, 1983; Secretary's Commission on Achieving Necessary Skills, 1990).

ACT gathered survey data from employers and educators in seven states (Illinois, Iowa, Michigan, Ohio, Oregon, Tennessee, and Wisconsin) and from several community colleges in California.

Collectively, they served as charter members in defining development efforts for ACT WorkKeys.[5] Charter members also assisted ACT in both the design and review of plans and materials, as well as providing examinee samples for prototype and field-testing.

Based on extensive reviews of the literature and empirical data collected from hundreds of educators and employers, eleven skills were initially selected for the ACT WorkKeys battery of assessments (McLarty, 1992). As with all assessments and the content and skills they measure, the ACT WorkKeys assessments need to be continually reviewed to confirm that the content assessed remains relevant and aligned to the foundation workplace skills, which are evolving over time, and that the construct measured represents the essential skills and behaviors required across jobs.

Test Development

Understanding the relationship of ACT WorkKeys level scores to skill definitions is critical for understanding both ACT WorkKeys test development and interpreting ACT WorkKeys scores. ACT subject matter experts (SMEs) established the skill levels for the cognitive skills before the assessments and scale scores were developed. The level scores are interpreted similarly to performance levels or achievement levels reported on educational tests. Further, skill levels are used both to classify examinees into performance categories and to identify the skill requirements for specific jobs (McLarty & Vansickle, 1997). ACT designed the level scores to be easily interpretable and to facilitate proper use in selection, promotion, development, or classification of individuals or groups. ACT WorkKeys assessments also report a more granular scale score. Scale scores provide finer distinctions between examinees and are useful for program evaluation, group comparisons, and research studies. Scale scores may also be more effective for selection decisions if adequate validation evidence supports such use.

Panels of employers and educators initially developed the ACT WorkKeys skill levels, and they began by defining the level of performance required for each skill based on available data and expert judgment. Next, the panel identified a list of exemplar tasks within the skill domain and selected those tasks most critical for performance across a variety of jobs. The SMEs then ordered the tasks by difficulty and complexity, building a hierarchy of skills with the least difficult and complex tasks forming the lowest skill levels and the more difficult and complex tasks forming the highest skill levels. Using information provided by the skill hierarchies, ACT developed initial test blueprints (McLarty & Vansickle, 1997). ACT developed the ACT WorkKeys cognitive skill levels to confirm that (1) the skill was assessed in a manner consistent with how it was used in the workplace, (2) the lowest level of skill was set at approximately the lowest level for which an employer would wish to set a standard, (3) the highest level of the skill was set at the highest level that entry-level jobs would require, and (4) the steps between levels were large enough to be distinguishable and small enough to be meaningful (McLarty, 1992).

Skill levels also are critical for item development. Based upon the defined skill levels, ACT first developed item prototypes and then utilized the prototypes as models for item development. Item writers receive training and coaching from ACT content specialists and develop items aligned to skill levels (e.g., a writer may develop an item to measure Reading for Information skills at Level 4). ACT provides item writers with guides that carefully define the skill levels and state that each item must align to an identified skill.

---

[5] Other charter members included the American Association of Community and Junior Colleges, the National Association of State Directors of Vocational Technical Education Consortium, and the National Association of Secondary School Principals.

When writers complete and submit items, ACT also requires them to submit workplace justifications ensuring that each item assesses the skill in a manner generally consistent with the way it is used in the workplace. Items developed and selected for use in the ACT WorkKeys assessments go through several reviews. ACT utilizes the reviews as a means of confirming that items are job-related and fair to candidates. External item reviewers carefully scrutinize each item for job relatedness and examinee fairness. They determine if the item is acceptable based on the following:

- Does the item assess skills needed in the workforce?

- Is the item applicable across a wide range of jobs and occupations?

- Does the item give an unfair advantage to examinees in some occupations or penalize examinees in other occupations?

- Does the item contain any content that might be offensive from a cultural, racial, or gender perspective?

- Are the knowledge, skill, and information required to answer the item equally available to all demographic groups?

Beyond the external reviews, ACT also conducts differential item functioning (DIF) analyses (a statistical procedure) on items contained within forms to identify possible differences in responses among racial groups and between males and females. The test development processes are conducted so that items are relevant to the workplace and are fair to examinees. As a result, ACT WorkKeys items align to defined workplace skills, are consistent with the construct, and are workplace-relevant.

Test blueprints articulate the skill and complexity level required for each level of items. All ACT WorkKeys foundational assessments incorporate hierarchies of increasing complexity with respect to the tasks and skills assessed. Utilizing the skill hierarchies, ACT originally based the scoring rules on a theoretical Guttmann model (Guttmann, 1950; McLarty, 1992). Because of the limitations of Guttmann scaling and resulting complexities in applying it to actual test data, ACT transitioned in 1996 from an item-based pattern scoring (Guttmann model) to a probabilistic model based on total score (IRT model).

ACT was able to transition to an IRT scoring model because of two important characteristics of ACT WorkKeys assessments. First, test items were based on a hierarchy of difficulty and complexity. Second, the assessments were unidimensional; that is, test blueprints and items were developed to measure a single domain. For these two reasons, all test items (not just the items representing a specific mastery level) could be used to infer an examinee's mastery of a defined level. To illustrate, an examinee's response pattern to all thirty items on the Reading for Information assessment could be used to determine whether he or she has mastered the skills of Level 3. IRT methods are used to analyze item responses and provide an estimate of the examinee's ability, which is not dependent on the specific test form or set of items administered. Using the IRT ability estimate, it is possible to estimate an examinee's probability of correctly answering any item in the pool. After estimating an examinee's ability on the IRT scale, ACT was able to estimate the expected performance (percentage of items answered correctly) with respect to the entire pool of items at each level. ACT used this information to establish cutoff scores for each level that represented an examinee being able to answer 80% or more of the items correctly within the level (Schulz, Kolen, & Nicewander, 1997).

The advantage of using the IRT model total score method over the Guttmann item-based pattern scoring is that it provides higher levels of reliability. Because the Guttmann item-based pattern scoring was based on only the items administered at a given level (e.g., six, eight, or nine items), score

reliability was relatively low. The IRT model total score method utilizes all items administered on the form (e.g., thirty or thirty-two items) to estimate the examinee's mastery. Utilizing more information to determine test scores normally provides better score reliability (Schulz, Kolen, & Nicewander, 1997).

Equivalent forms of cognitive assessments have been developed to maintain test security. As ACT develops new forms of the cognitive assessments, each form is constructed to adhere to the test blueprint. All ACT WorkKeys test blueprints define both content requirements and statistical requirements. To control for the inevitable small differences in form difficulty, and to confirm the accuracy and meaningfulness of scores, ACT equates and scales each newly developed form to a base form. Test equating is a statistical process used to verify that the scores obtained through the administration of a new form have the same meaning and interpretation as scores earned on earlier forms (Kolen & Brennan, 2004). ACT's application of equating methods review whether examinees' scores are independent of the test form. It should be a matter of indifference to the test takers which form of the assessment they take (Lord, 1980). More specifically, examinees should have the same probability of earning a specific score regardless of whether they take Form X or Form Y.

To enhance score interpretations and to decrease potential score misuse, standardized assessments develop a score scale that is independent of the number of items or percentage of items answered correctly (i.e., the raw score). The score scale provides a common metric through which performance on different forms of the assessment are compared. By establishing a common metric and then applying IRT equating methods to place raw scores on the common metric, scores earned on different forms have the same meaning.

ACT WorkKeys multiple-choice cognitive assessments have a 26-point score scale, which ranges from 65 to 90. The assessments provide level scores with either four or five levels. Through the equating and scaling process, for each assessment (e.g., Reading for Information, Applied Mathematics), test users are able to interpret and compare scores achieved on one form to scores earned on a different form.

In addition to equating and scaling, ACT continually evaluates the reliability of ACT WorkKeys assessment scores using a variety of techniques. These techniques include estimating the internal consistency of each test form, conducting generalizability analyses, computing scale score reliability estimates, and estimating classification consistency. Classification consistency refers to the extent to which classifications of examinees agree when obtained from two independent administrations of a test or two alternate forms of a test.

Testing accommodations are available for individuals with disabilities taking the ACT WorkKeys tests, as required by the Americans with Disabilities Act. Accommodations are authorized by the test supervisor, following ACT guidelines and with proper documentation, and may include the use of special testing materials provided by ACT, such as large-print test booklets, large-print answer documents, captioned videotapes, braille versions of the tests, and reader scripts. Also offered are the use of a sign-language interpreter to sign test items and response choices in Exact English Signing (usually by signing from a regular-print test booklet), assistance in recording responses (may include a large-print answer document), the use of word-to-word foreign language dictionaries, and extended testing time. Scores on ACT WorkKeys assessments for examinees who take those assessments under testing conditions that do not meet ACT standards will not be considered eligible for the ACT NCRC (ACT, 2007).

Summary of Convergent and Discriminant Validity Evidence

In addition to the construct and content validity evidence related to test development, the following evidence related to convergent and discriminant validity has been assembled.

ACT WorkKeys Reading for Information—Measures skill in reading and using written text to do a job

To support the construct-related validity of Reading for Information test scores, ACT examined the relationship between Reading for Information and the ACT reading and English tests, which measure the language skills identified as prerequisites to successful performance in entry-level college courses in reading and English (ACT, 2008a). ACT collected score data from two test administrations in a midwestern state, one from a sample of 121,304 students in spring 2002 and another from a sample of 122,820 students in spring 2003. Test takers who received higher scale scores on the ACT reading and English tests generally received higher level scores on Reading for Information (ACT, 2008a).

ACT WorkKeys Applied Mathematics—Measures skill in applying mathematical reasoning, critical thinking, and problem-solving techniques to work-related problems

Researchers examined the relationship between ACT WorkKeys Applied Mathematics and the ACT mathematics test to provide construct evidence (ACT, 2008b). The ACT mathematics test measures the mathematical reasoning skills needed to solve practical problems in mathematics identified as prerequisite to successful performance in entry-level courses in college mathematics. ACT collected score data from two test administrations in a midwestern state, one from a sample of 121,304 students in spring 2002 and another from a sample of 122,304 students in spring 2003. The results indicate that ACT WorkKeys Applied Mathematics scores are highly correlated with ACT mathematics test scores. In general, test takers who received higher scores on Applied Mathematics also received higher scale scores on the ACT mathematics test (ACT, 2008b).

Adverse Impact

When ACT WorkKeys tests are used for pre-employment screening or other essential employment decisions, employers should ensure that a well-documented job analysis or other available evidence links the skills required on the job to the skills measured through the assessment. If cutoff scores are needed, they should be established at the appropriate levels, and the process for determining that level needs to be clearly documented (AERA et al., 2014; SIOP, 2000).

When sufficient numbers of test takers are available, ACT uses DIF analyses to evaluate and flag operational items that could be unfair to any group of test takers. Items found to be fair in earlier qualitative reviews can still function differently for specific population subgroups. DIF detects any statistical differences in terms of item responses between a specific population (the focal group) and subgroup of equal ability. DIF procedures take background group differences into account and indicate whether an item may perform differentially for a specific group of test takers (e.g., females, Hispanics) as compared to all test takers. ACT uses the standardized difference in proportion correct (STD) and the Mantel-Haenszel common-odds-ratio (MH) statistics to detect the existence of DIF in items on ACT WorkKeys test forms. Items found to exceed critical values for DIF are reviewed singly and overall. The results of this review may lead to the removal of one or more items from a form (ACT, 2008a).

ACT has conducted fairness reviews comparing group means on the ACT WorkKeys assessments. To examine item level fairness, ACT conducted a preliminary DIF analysis on operational items (two forms for each of Reading for Information, Applied Mathematics, and Locating Information). The findings indicated that the majority of items in the investigated forms do not perform differently between females and males, Blacks and Whites, or Hispanics and Whites (ACT, 2008a).

ACT WorkKeys Levels and Demographic Group Analysis

Tables 2–4 present supplemental analyses of data comparing mean ACT WorkKeys Level scores by race/ethnicity, gender, and age. In some comparisons, mean scores were statistically different despite the fact that the mean level score rounded to the same score. For all three demographic group analyses, the standardized mean effect (d) is reported to provide insight regarding the magnitude of the effect of the difference in group mean scores (Cohen, 1977). By reporting the magnitude of the effect of the mean differences, the results can be viewed in terms of their practical significance in interpreting differences in level scores. The meaning of effect size varies by context, but the standard interpretation offered by Cohen (1988) is that .80 = large effect, .50 = moderate effect, and .20 = small effect.

The fact that statistically significant differences in cognitive ability test performance are typical between a majority and a minority group has been thoroughly researched and documented (Ryan, 2001). Performance on the ACT WorkKeys cognitive assessments is consistent with these findings.

Table 2. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Race/Ethnicity

ACT WorkKeys Assessment	White			Minority			t	df	95% CI	Cohen's d
	n	M	SD	n	M	SD				
Applied Mathematics	1,631,837	4.95	1.41	1,539,375	4.15	1.67	457.63**	3,015,975	[.794, .800]	0.52
Locating Information	1,427,486	4.058	0.96	1,013,784	3.589	1.23	312.56**	2,370,323	[.421, .427]	0.43
Reading for Information	1,626,558	5.069	1.18	1,192,107	4.531	1.29	357.53**	3,098,849	[.497, .502]	0.43

Note: ** p < .001

Table 3. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Gender

ACT WorkKeys Assessment	Male			Female			t	df	95% CI	Cohen's d
	n	M	SD	n	M	SD				
Applied Mathematics	1,651,372	4.70	1.61	1,483,814	4.43	1.58	153.25**	3,111,549	[.272, .279]	0.17
Locating Information	1,422,567	3.86	1.17	1,227,836	3.87	1.06	-5.43**	2,643,152	[-.010, -.005]	-0.01
Reading for Information	1,646,941	4.76	1.35	1,486,163	4.87	1.18	-74.29**	3,128,851	[-.109, -.103]	-0.08

Note: ** p < .001

Table 4. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Age

ACT WorkKeys Assessment	Less than 40 Years of Age			40+ Years or Older			t	df	95% CI	Cohen's d
	n	M	SD	n	M	SD				
Applied Mathematics	2,899,385	4.56	1.60	507,016	4.57	1.54	-3.52**	713,234	[-.013, -.004]	0.00
Locating Information	2,408,959	3.88	1.11	511,006	3.73	1.17	86.27**	718,231	[.150, .157]	0.14
Reading for Information	2,895,981	4.78	1.27	506,027	5.07	1.25	-156.09**	700,574	[-.301, -.294]	-0.24

Note: ** p < .001

City of Albuquerque Study—ACT WorkKeys Fairness Analysis

ACT examined test bias for Reading for Information, Locating Information, and Workplace Observation with applicants for the position of motorcoach operator for the City of Albuquerque (n = 92). No differences in scores on Reading for Information, Locating Information, or Workplace Observation were observed for comparisons of Hispanic/Latino examinees to White examinees or for female examinees to male examinees. The only difference found for protected groups was in Locating Information. Examinees less than 40 years of age scored significantly higher than examinees aged 40 or over (M = 4.20 to M = 3.94).

Alignment Studies

Crosswalks, or alignment studies, which examine the overlap of skills and test content provide additional evidence of construct validity. Various alignment studies of ACT WorkKeys assessments have been conducted over the years to compare the skills included in secondary education standards and business competency models.

ACT WorkKeys Assessments

In 2014, ACT conducted an alignment study of ACT WorkKeys assessments with a preliminary version of an industry competency model developed by the National Network of Business and Industry Association (NNBIA) (ACT, 2014b). The competency model, or Blueprint, was developed to describe the common employability skills necessary for jobs across all industries and sectors. The NNBIA Blueprint includes four competency buckets—Personal Skills, People Skills, Applied Knowledge, and Workplace Skills—with core skills highlighted within each bucket. Eleven ACT WorkKeys cognitive and non-cognitive assessments were included in the alignment study. The analysis found a high level of correspondence between the NNBIA Blueprint and the ACT WorkKeys assessments. Roughly 90% (70 out of 79) of the common employability skills identified in the Blueprint were measured by the ACT WorkKeys assessments. The greatest deficiency areas are science and technology knowledge and business fundamentals.

National Assessment of Education Progress (NAEP) and ACT WorkKeys

Separate alignment studies were conducted between ACT WorkKeys assessments and the 8th and 12th grade NAEP, a national assessment of what students know and can do in various academic subjects. Both studies examined alignment of ACT WorkKeys Reading for Information and Applied Mathematics and were conducted to determine whether NAEP could support inferences about career readiness. ACT WorkKeys was selected by the National Assessment Governing Board (NAGB) because it was considered to be a national indicator of career readiness (NAGB, 2010a). "While NAEP has been designed to provide evidence of what students in the United States know and can do in a broad academic sense, WorkKeys assessments provide information about job-related skills that can be used in the selection, hiring, training, and development of employees" (NAGB, 2010a, p. i).

The alignment of ACT WorkKeys Applied Mathematics with the NAEP Grade 12 Mathematics assessment was conducted in 2010 by the NAGB (NAGB, 2010a). The study investigated the alignment of ACT WorkKeys Applied Mathematics test items and levels against the items and five content strands in the NAEP Grade 12 Mathematics framework.[6] Alignment of the NAEP Mathematics items to the ACT WorkKeys Applied Mathematics items was found for 75% of the

---

[6] Number Properties and Operations; Measurement; Geometry; Data Analysis, Statistics, and Probability; and Algebra.

NAEP items, while 40% of ACT WorkKeys Applied Mathematics items aligned to the NAEP standards. The NAEP Mathematics items with the strongest alignment to Applied Mathematics included problem-solving applications of number operations and measurement. NAEP Mathematics items for which no ACT WorkKeys Applied Mathematics items were aligned were related to geometry, data analysis, statistics, probability, and algebra.

The alignment study of ACT WorkKeys Reading for Information and NAEP Reading found some degree of alignment with NAEP standards on literacy and informational reading and integrating/ interpreting NAEP standards (NAGB, 2010b). Additionally, ACT WorkKeys Reading for Information items to which no NAEP Reading items aligned included applying complex, multistep, conditional instructions to similar and new workplace situations; determining the meaning of work-related acronyms, jargon, and technical terms; and figuring out and applying general principles contained in informational documents to similar and new workplace situations.

ACT WorkKeys Cognitive Assessments: Reading for Information, Applied Mathematics, Locating Information, and Applied Technology

In 2014, the NAGB conducted another alignment study to compare ACT WorkKeys assessment content with the NAEP Mathematics and Reading items and frameworks (NAGB, 2014). The study sought to expand upon the prior NAEP alignment studies by including additional ACT WorkKeys assessments (Applied Technology and Locating Information) that might align to NAEP Reading and Mathematics. Items and Targets for Instruction for ACT WorkKeys Reading for Information were paired with Locating Information for analysis while those for Applied Mathematics were paired with Applied Technology. The results found that NAEP items did not represent the content domain. ACT WorkKeys Applied Mathematics and Reading for Information content domains (52% and 72%, respectively) were not matched to NAEP items. Sixteen of the twenty-four content strands in the NAEP Mathematics framework and one of the three cognitive targets in the NAEP Reading framework could not be matched to any of the ACT WorkKeys items.

The NAGB concluded that there was weak alignment between the ACT WorkKeys and NAEP assessments, which was expected due to the differing purposes of the two suites of assessments (i.e. NAEP is not a career or work assessment, but rather a measure of academic skills). Based on the low level of alignment and the evidence ACT provided concerning the relationship of ACT WorkKeys content to job preparedness, the study concluded that results suggest NAEP is not an appropriate measure of academic preparedness for job training (NAGB, 2014).

Overall Claims and Interpretative Argument of ACT WorkKeys Test Scores

ACT WorkKeys assessments can be used for (1) pre-employment screening to identify individuals who have achieved levels of skill proficiency needed, (2) pre-employment screening to identify less desirable candidates based on behaviors associated with job performance, (3) employee development, and (4) determining the appropriate level of fit with occupations in terms of interests.

When ACT WorkKeys tests 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 is linked to work behaviors and job tasks either through job profiling or through research that links the test to job performance. When ACT WorkKeys tests 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 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 or other employment decisions.

The ACT NCRC can be used by (1) an employer who uses the ACT NCRC and other criteria to identify a qualified pool of applicants[7] and does not require a specific level of the ACT NCRC, (2) an employer who uses the ACT NCRC to make employment decisions and requires a specific level (e.g., gold), (3) states, communities, and schools who use the ACT NCRC to document an individual's level of essential work readiness skills, and (4) states, communities or schools who use the ACT NCRC to document the aggregate "work readiness" of a community, region, or state.

When employers recommend that applicants provide evidence of their ACT NCRC, they should do so in addition to other traditional criteria (e.g., employment application, credentials, and interview) as part of the job application process. In instances where employers use the ACT NCRC and the other criteria to screen prospective applicants for hiring, a formal job profile should be conducted. When available, such evidence may also be transported from job analysis studies of similar positions requiring the same skills. Evidence that supports the use of ACT WorkKeys or ACT NCRC to predict one outcome may not necessarily be generalized to other outcomes.

# Content Validation Evidence

Overview

Content evidence comprises one source to establish validity of test scores in support of the interpretative argument (AERA et al., 2014). Content evidence often comprises the first line of evidence to support employment selection practices. The *Uniform Guidelines* (EEOC, 2000), the *Standards* (AERA et al., 2014), and the *Principles* (SIOP, 2003) all describe the need to demonstrate that knowledge and skills in pre-employment measures should be demonstrably linked to work behaviors and job tasks. Both the *Standards* (2014) and the *Principles* (2003) suggest that expert judgment can be used to determine the importance and criticality of job tasks and to relate such tasks to the content domain of a measure. This process is commonly conducted through a job analysis which identifies the tasks required for performance on a job and subsequently for development of the content blueprint and item development to ensure content validity (Cascio, 1982; Dunnette & Hough, 1990). ACT WorkKeys assessments were designed to assess generalizable skills and skill levels associated with many jobs. As such, the content-related validity evidence for ACT WorkKeys assessments was originally established by SMEs across numerous jobs that linked ACT WorkKeys skills and skill levels to specific tasks and job behaviors for a particular job.

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

---

[7] As noted previously, reading, mathematics, and locating information are included in the ACT NCRC based on evidence that the majority of profiled careers and career clusters require these skills. However, organizations employing ACT NCRC pre-employment screening or selection decisions should verify these essential skills are important for their positions or careers. This can be done in a variety of methods and does not mandate a formal job analysis or job profiling study.

The job profiling process involves several steps to establish a link between the ACT WorkKeys skill level definitions and the requirements of a particular job. Ideally, the SMEs identified comprise a representative sample across a variety of factors (e.g., race, ethnicity, gender, geographic location). Incumbent SMEs with disabilities also participate in focus groups as part of achieving a representative sample. In such cases, profilers are encouraged to contact ACT for recommendations for accommodations during the profiling process. Figure 1 provides an overview of the job profiling process.
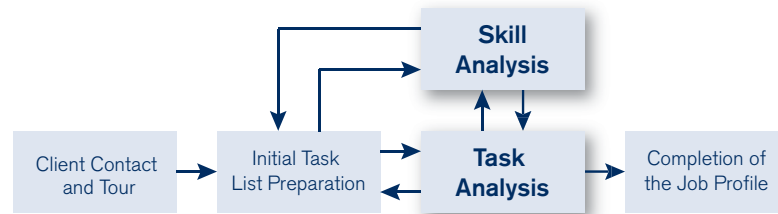
Figure 1. Job profiling process

This process begins with a task analysis where the group of SMEs is asked to generate a task list that accurately represents the job at an organization and to rate each task for importance. Figure 2 identifies the outcomes of the task analysis as a part of the complete job profile process.
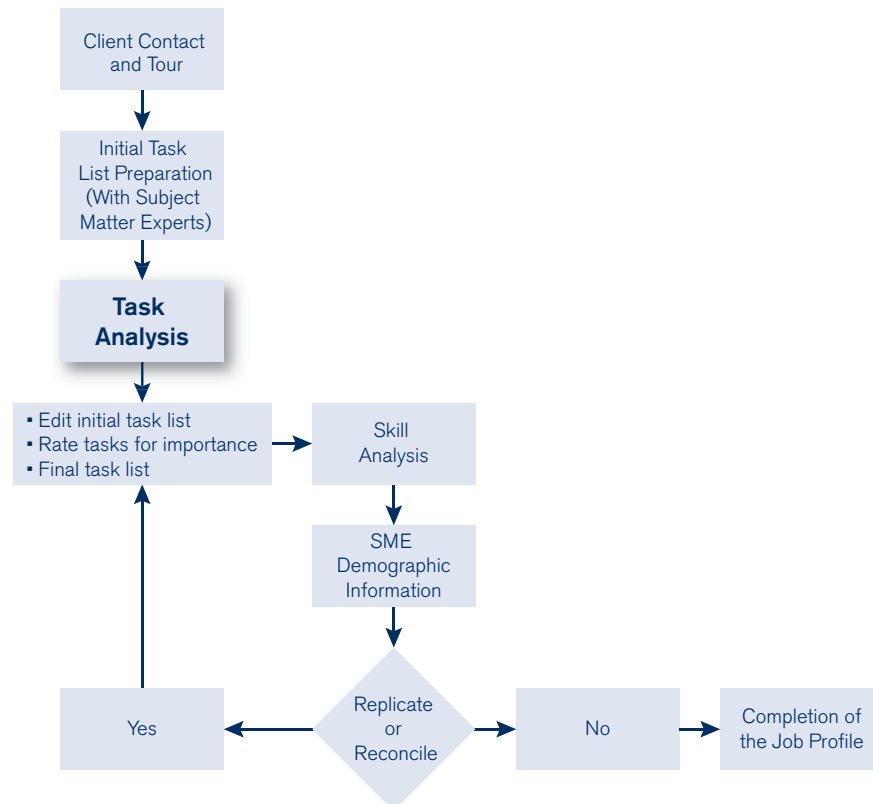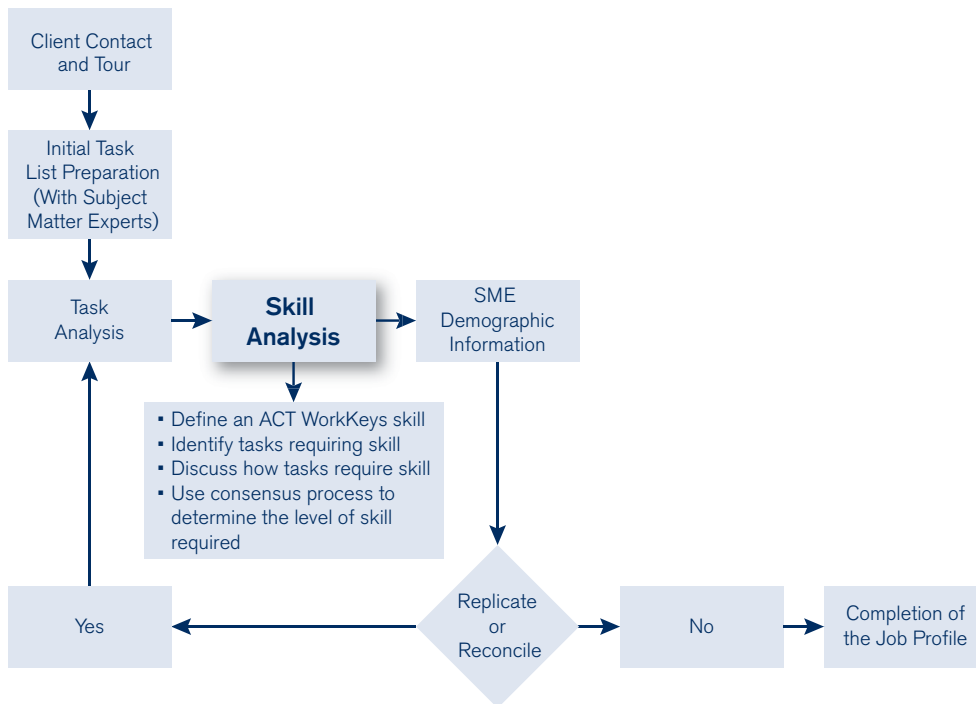

Figure 2. Task analysis process

Equally important is the skill analysis where the SMEs review each skill measured by ACT WorkKeys assessments. Once the SMEs understand the definition of an ACT WorkKeys skill and have determined its relevance to the job, they independently identify the important tasks on the Final Task List that require the skill. They also identify how the tasks specifically use that skill. After discussing the relationship between the skill and the tasks, only those tasks identified as important by a majority of SMEs are included in the subsequent discussion, 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 phase, the SMEs use successive approximation to determine the skill level required for that final set of tasks. Each skill level denotes a level of difficulty, with the lowest level representing the simplest of tasks related to the skill construct and the highest level representing the most complex. The SMEs typically begin with the lowest skill level. They then determine whether the job requires skills at, above, or below the level described. If the SMEs determine that the skills required for the job are higher than skills described in a level, they proceed to the next higher level; if they determine the required skills are lower, they review the next lower level. If they determine that the skills are about the same as the level they are reviewing, they are still shown both the next higher level before confirming agreement between skills and a designated level to confirm their initial judgment.

No decision is reached until the SMEs have considered a range of skill levels: those skills they have identified at the required level, at least one level above it, and at least one level below it (unless they have chosen the highest or lowest level available). Occasionally, the SMEs find that the level required is below or above the levels measured by ACT WorkKeys. Figure 3 identifies the outcomes of the skill analysis as a part of the complete job profile process.

Figure 3. Skill analysis process

Summary of Content Evidence

As of January 2015, approximately 20,000 ACT job profiles have been conducted, which comprise the content validity evidence for ACT WorkKeys. ACT's job profile database, JobPro, represents a wide cross-section of jobs, including 53% (584) of all O*NET codes (1,091). The database provides foundational skills data for 193 (50%) of the 387 Bright Outlook Occupations as defined by O*NET using US Bureau of Labor Statistics Occupational Projections data for 2012–2022.[8] JobPro data is representative of occupations across major occupational families and levels of education and training required for entry (ACT, 2014c). Figure 4 provides the distribution of a cohort of job profiles by major occupational family. When aggregated by occupation, the aggregate job profile data represents 86% of the total occupational employment in the US (US Bureau of Labor Statistics, 2013). Figure 5 provides the distribution of a cohort of job profiles and US employment by major industry sector.
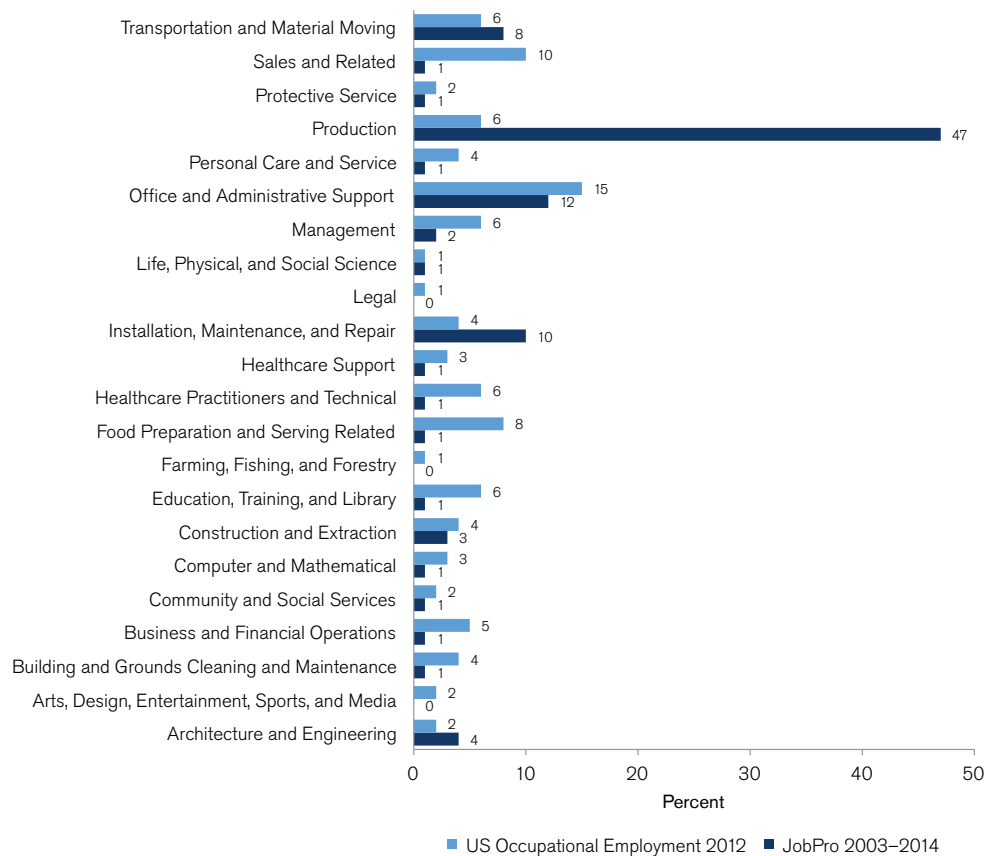
Figure 4. Distribution of job profiles and US employment by occupational family

---
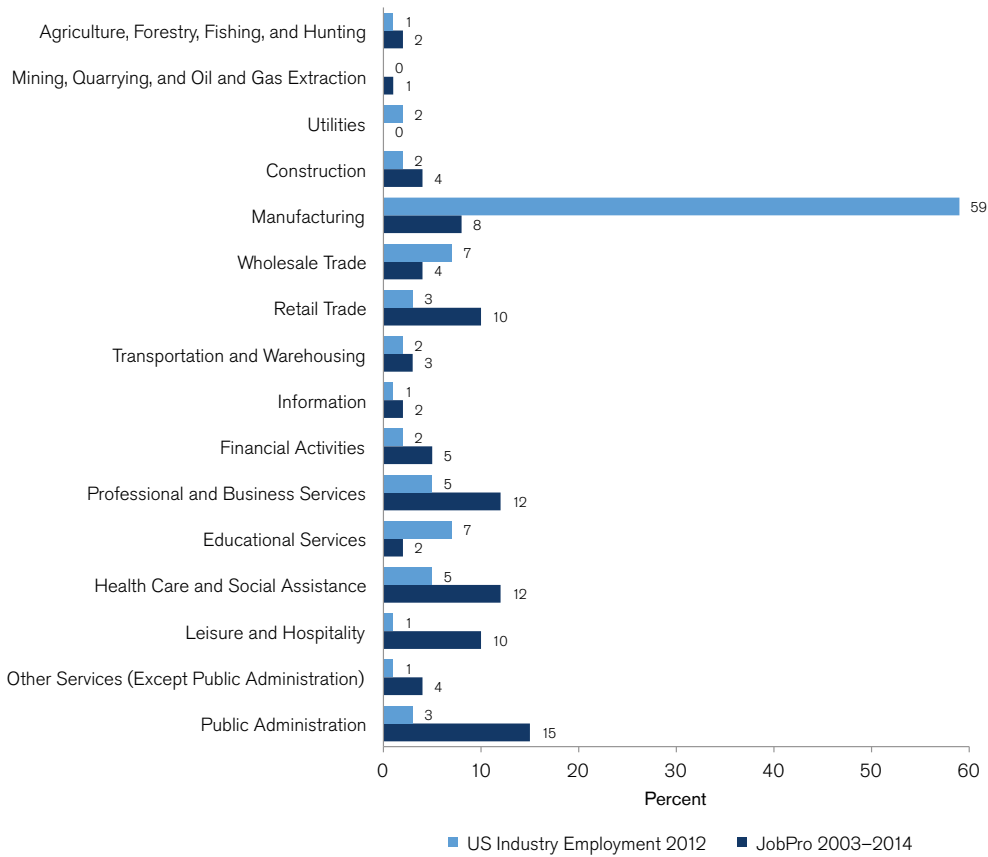
Figure 5. Distribution of job profiles and US employment by industry sector

Criterion Validation and Outcome-Based Evidence

Overview

Criterion validation evidence includes statistical studies that establish a relationship between a test score and an outcome or criterion relevant to the proposed purpose of the test (AERA et al., 2014; SIOP, 2003; EEOC, 1978). Different outcomes or criteria may be relevant for each specific use of ACT WorkKeys assessments and the ACT NCRC. For example, ACT WorkKeys assessments and the ACT NCRC are used for employment decisions where job performance is the primary outcome of interest. Job performance is typically measured via supervisor ratings of performance, but other criteria or employee outcomes such as attendance and safety incidents may be used in specific studies. When ACT WorkKeys tests 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, and training completion). Evidence that supports the use of ACT WorkKeys or the ACT NCRC to predict one outcome may not necessarily generalized to other outcomes.

There are two main types of criterion validation research studies. In a predictive study, a test is administered to a group of job applicants but is not used for selection decisions. Some of the applicants are hired, and some are not. After those who were hired begin their jobs, their supervisors rate their performance. Test scores are then compared to the performance data to determine if

there is a positive relationship between scores or level with job criteria. In a concurrent study, a test is administered to job incumbents, and the scores are compared to data describing the incumbents' current job performance. This comparison is based on test data and performance data collected for the same time period and the results are used to determine if scores and levels are positively related to outcomes (AERA et al., 2014; EEOC, 2000; SIOP, 2003).

Generalizability or validity generalization is the degree to which criterion validity evidence may be generalized to a new situation (e.g. other employers or jobs) without the need for conducting a separate study for the new setting and situation under some conditions. The *Principles* (2003) delineate three strategies for generalizing validity evidence:

- transportability
- synthetic validity/job component validity
- validity generalization using meta-analytics

Transportability refers to the use of a job selection procedure for a new situation based on validation evidence from a previous study. This method of generalizability relies on establishing similarity of job requirements and content as well as context and job applicant pools (SIOP, 2003).

Synthetic validity involves the use of a selection procedure based on the validity of inferences of the selection assessment scores to specific domain/components of the job (e.g., word processing, fluency in communicating in a foreign language). Such validity generalization is dependent upon the establishment of a documented relationship between the assessment scores and the job components for a job or multiple jobs (SIOP, 2003).

Regarding the third validity generalization strategy, the *Standards* (2014) state that statistical summaries of previous validation studies can be useful when estimating criterion validity for a new situation. Meta-analytic techniques should take into account the variability of sampling fluctuations and reliability of criterion measures.

Criterion validation studies can also serve as the basis for utility calculations to show return on investment of the Work Readiness system. Traditional return on investment research is conducted at the organizational or employer level where an employment selection system is implemented. The validity coefficients from a criterion validation study in addition to other organizational performance metrics are used to calculate the impact or "utility" of the intervention introduced (Brogden, 1949; Cabrera & Raju, 2001; Cronbach & Gleser, 1965).

The correlations between ACT WorkKeys tests and job performance ratings provide criterion-related evidence for the validity of using ACT WorkKeys assessments in relation to a specific job. A number of studies have been conducted across a range of organizations which examine the relationship of 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, 2015b). 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 A1 (ACT, 2015a). Several recent studies are discussed in more detail in the Appendix.

Summary of Criterion Validation Evidence

There has been and will continue to be a need to conduct local criterion validation studies or apply meta-analysis to support the use of the ACT WorkKeys system for pre-employment selection and other high-stakes purposes. This section briefly summarizes available criterion-related evidence for ACT WorkKeys assessments in terms of the work or educational setting, outcomes examined, and direction and strength of the findings.

Reading for Information—The results of studies on Reading for Information 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 (5 of 5 correlations), and a positive relationship with reduction in safety incidents and customer complaints (1 of 1 correlation).

Applied Mathematics—Research on Applied Mathematics shows a modest relationship with overall job performance and a positive relationship with education outcomes for GPA, course grades, and persistence (5 of 5 correlations).

Locating Information—The studies about Locating Information show a modest relationship with overall job performance, education outcomes such as course grades (2 of 2 correlations), and reduction in safety incidents, customer complaints, absenteeism, and turnover (4 of 4 correlations).

ACT NCRC—Research on the ACT NCRC shows a modest relationship with increase in earnings, employment attainment, and employment retention rates.

Discussion

ACT WorkKeys tests originally relied solely on content validation evidence in its inception. Content validity evidence—traditionally established through job analysis, which demonstrates a strong link between assessment content and knowledge, skills, and abilities required on the job—is often the first building block of a validity argument. In employment testing, content validity evidence often serves as the primary form of validity evidence. More recently, additional forms of evidence have supplemented the validation research to demonstrate efficacy and utility associated with important outcomes in work, educational, and employment training environments. 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 ACT WorkKeys 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 these ACT WorkKeys tests and the ACT NCRC.

Construct-Related Evidence

Evidence demonstrating that the ACT WorkKeys assessments do in fact measure the skills they purport to (construct validity) is critical. Previous sections of this paper have described the test development process and initial studies used to provide construct validity evidence. Research demonstrates strong relationships between ACT WorkKeys assessments and other cognitive tests and outcomes (e.g., ACT, grades). Additional research examining the relationships between the ACT WorkKeys cognitive test scores and other assessments (e.g., both convergent and divergent validity) commonly used with adult populations in work settings would be a useful supplement to the existing body of research, so that a clear indication of the meaning of the test score is determined.

Specifically, there have been multiple attempts to crosswalk the ACT WorkKeys/ACT NCRC skills to measures of secondary education standards and to postsecondary education entry-level requirements. The degree to which the work-contextualized cognitive skills measured by ACT WorkKeys overlap with skills developed in an academic setting is largely posited to an assumption that ACT WorkKeys cognitive skills tap into a larger measure of general cognitive ability. Schmidt and Sharf argue that any three of the ACT WorkKeys/ACT NCRC tests are measuring general cognitive ability (Schmidt & Sharf, 2010). Additional research studies that investigate the relationship of the ACT NCRC skills to reliable measures of general cognitive ability would be useful in providing a better understanding of the degree of overlap of constructs measured by the ACT WorkKeys cognitive assessments and cognitive skills needed in non-work-related settings.

Content-Related Evidence

ACT WorkKeys tests were developed using a content validation approach so that test items and cut scores were linked to data collected across job profiles. Results from the ACT job profile database provide essential content validation evidence across ACT WorkKeys tests and skill levels (i.e. cut scores) that are important to job performance.

Hiring applicants using profile recommendations (pass/fail) *should* result in higher job performance and better outcomes (e.g., less turnover after using this system compared to before using ACT WorkKeys). However, more recent studies are needed to provide more contemporary evidence of content validity since job knowledge and skills change over time.

More recent studies have begun to analyze these assumptions by collecting both organizational and job performance outcome data in addition to conducting a job profile for specific job titles. If the goal of a study is to generate criterion-related validity results, then a job analysis is still necessary, but it is not necessary to conduct the detailed analysis done via job profiling that was developed for content validity purposes. The *Uniform Guidelines* indicate, "any method of job analysis may be used if it provides the information required for the specific validation strategy used" (EEOC, 1978, p. 129).

Once a job profile has been conducted for an occupation, additional evidence from a criterion-related validation study can be assembled at a later point. Of course, if the job requirements and skills changed from the time the job profile was established, more recent information is required. When recommending a test or cut score solely on the basis of content validity (resulting from a job profile), a high percentage of the tasks need to be related to the skill measured by the test.

Criterion-Related Evidence

Overall, criterion-related validity evidence is strong for the three ACT WorkKeys cognitive tests comprising the ACT NCRC and for the ACT NCRC itself. Additional studies for individual assessments provide some evidence of efficacy, primarily in predicting an overall measure of job performance, and occasionally one or more specific facets of performance. Studies showing the relationship of job performance to Reading for Information, Locating Information, and Applied Mathematics have been conducted primarily in manufacturing, healthcare, and government settings.

There is a continual need to collect evidence to support claims about the use of test scores (AERA et al., 2014). However, criterion-related studies are often difficult to conduct unless employers are willing to partner with assessment providers to conduct such studies. Criterion-related research requires adequate size and representative samples with either incumbents or applicants and reliable and relevant criteria. In addition, when validity is examined by subgroup (e.g., gender, ethnicity), location, or

specific jobs, larger samples are required to ensure there is sufficient "power" to examine effect sizes. ACT provides a number of incentives and strategies that can allow organizations wanting to use the assessments to conduct such research in advance of high-stakes uses.

In concurrent validation studies, statistical corrections can be made for the restricted range of both the ACT WorkKeys and outcome data. The criterion validity of the ACT WorkKeys assessments or the ACT NCRC may be influenced by other selection methods that employers are already using. However, such studies may still show the additional value or "incremental validity" that the ACT WorkKeys assessments and ACT NCRC can add to the employment selection process.

Reliable and relevant criteria are needed in order to conduct concurrent and predictive studies. Supervisor ratings are generally the most frequently used measure for such studies, and ACT developed the ACT Supervisory Survey to provide employers with a rigorous measure of relevant job performance to assist in such studies. More recent studies have utilized an ACT-developed measure of job performance, the ACT Supervisor Survey and the newer iteration of it—ACT WorkKeys Appraise. This newer measure relies on more generalized categories of job performance based on literature about common dimensions of job performance. ACT WorkKeys Appraise allows for the systematic collection of job performance ratings across settings and jobs. Employers should ensure that measures of job performance on criterion measures such as ACT WorkKeys Appraise are relevant to the job and job settings. This can be achieved in a number of ways, including, but not limited to, a job analysis or job profile study. Of course, other facets of job performance may be used in such studies when the purpose of the assessment is more narrowly focused (e.g., safety, integrity).

ACT NCRC

Employers and industry associations primarily use the ACT NCRC as a measure of essential foundational work readiness skills important for training and employment success. Positive results have been consistently reported across jobs with the composite certificate as well as the three ACT WorkKeys tests that comprise the ACT NCRC. Studies have investigated the relationship to employment, wage, and occupational training outcomes. Additional research utilizing rigorous quantitative measures that control for individual demographics and comparison groups is desirable. The ACT NCRC is intended for certifying initial competency in foundational skills but should not be used for employment selection purposes without detailed job profiling.

Due to the current design of the ACT NCRC levels (i.e. the lowest skill level across three assessments is determinative), there is a large degree of variability of skill levels within ACT NCRC levels. Previous studies have created composite scores of two or three of the ACT NCRC assessments to account for such variability but are problematic because composite scores are not an intended or actual use of the certificate. Composite scores are also problematic because of their compensatory nature and require specific studies to confirm that various combinations across test score levels are equivalent and related to the criterion.

Employers may use the ACT WorkKeys system for a variety of purposes. One likely use is to achieve a positive organizational impact, such as improved employee retention, increased efficiency or productivity, or reductions in accidents or disciplinary actions. Traditional approaches to utility analysis can be very time- and labor-intensive for both employers and researchers; however, newer approaches to estimate the organizational impact of the ACT WorkKeys system have been employed and are available.

Limitations

ACT's mission is to help individuals achieve education and workplace success. Therefore, it makes concerted efforts to collect evidence at the individual level showing that the ACT WorkKeys system promotes individuals' success.

However, there are a number of limitations in collecting validation evidence, especially in employment settings, that are important to recognize. First, employment and wage data for individuals is primarily found in governmental databases owned by state and federal entities. Second, there are numerous challenges in obtaining individual-level data required for in-depth research studies in many organizations.

For example, some institutions and agencies do not collect criterion data, some organizations conduct their own validation studies and results, some organizations prohibit test providers from publishing details or findings from studies, and others are reluctant to share such data for a variety of reasons.

Private sector organizations are generally more restrictive in allowing access to individual and aggregate outcome data. Moreover, state and federal databases primarily use personal identifiers (e.g., Social Security number) that ACT does not collect for ACT WorkKeys examinees. For both education and workforce outcome research, the ability to accurately match ACT WorkKeys examinee data to third-party outcome data is an issue that needs to be considered. Self-reported or estimated outcome data at the individual or organizational level are often used to provide validity evidence to support the use and interpretation of assessment results, and ACT has conducted many such studies, which are described in this report.

Criterion-related studies matching individual outcomes on the job with predicted performance provide an additional source of empirical research. However, no single study or line of evidence will sufficiently address questions of validity across jobs, organizations or settings, but together, multiple types of studies, with different populations in different settings, can collectively provide evidence to support the use of ACT assessments. ACT continues to employ a variety of validation strategies, including studies of test content, constructs, response processes, and criterion-related performance, to provide evidence which supports the intended purposes of its tests and interpretation of test scores.

References

ACT. (1992). *A strategic plan for the development and implementation of the WorkKeys system*. Unpublished report, Iowa City, IA: ACT.

ACT. (1987). *Study power* [Series]. Iowa City, IA: ACT.

ACT. (2007). *WorkKeys assessment technical bulletin*. Iowa City, IA: Author.

ACT. (2008a). *WorkKeys Reading for Information technical bulletin*. Iowa City, IA: ACT.

ACT. (2008b). *WorkKeys Applied Mathematics technical bulletin.* Iowa City, IA: ACT.

ACT. (2010). *WorkKeys value across the pipeline*. Unpublished report, Iowa City, IA: ACT.

ACT. (2011a). *JobPro skill analysis*. Unpublished report, Iowa City, IA: ACT.

ACT. (2011b). *E. & J. Gallo: WorkKeys—uncorking an employee's potential*. Unpublished report, Iowa City, IA: ACT.

ACT. (2012a). *Employer handbook: Talent acquisition*. Iowa City, IA: ACT.

ACT. (2012b). *WorkKeys American Health Care Association Study.* Unpublished report, Iowa City, IA: ACT.

ACT. (2013a & 2014d). *WorkKeys Concurrent Validation and Talent Benchmark Study for a Health Care Organization.* Unpublished report, Iowa City, IA: ACT.

ACT. (2013b). *WorkKeys as a predictor of critical thinking*. Unpublished report, Iowa City, Iowa: ACT.

ACT. (2013c). *Ohio MSSC CPT and NCRC®: A partnership for industry-recognized credentialing*. Iowa City, IA: ACT.

ACT. (2014a). *The ACT® Technical Manual.* Iowa City, IA: ACT.

ACT. (2014b). *ACT WorkKeys assessments and the National Network of Business and Industry Associations' common employability skills framework*. Iowa City, IA: ACT.

ACT. (2014c). *JobPro gap analysis*. Unpublished report, Iowa City, IA: ACT.

ACT. (2014e). *City of Albuquerque ACT WorkKeys validation study and comparison study*. Unpublished report, Iowa City, IA: ACT.

ACT. (2015a). *Summary of WorkKeys criterion validity research*. Unpublished report, Iowa City, Iowa: ACT.

Agency for Instructional Technology (AIT). (1989). *Workplace readiness: Education for employment, personal behavior, group effectiveness and problem solving for a changing workplace*. Unpublished report, Bloomington IN: AIT.

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

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

Bailey, L. J. (1990). *Working: Skills for a new age*. Albany, NY: Delmar.

Bolin, B. (2005). *The Career Readiness Certificate—An idea whose time has come*. Retrieved from www.ncrcadvocates.org.

Brogden, H. E. (1949). When testing pays off. *Personnel Psychology*, *2*(2), 171–183.

Cabrera, E. F., & Raju, N. S. (2001). Utility analysis: Current trends and future directions. *International Journal of Selection and Assessment*, *9*(1-2), 92–102.

Carnevale, A. P., Gainer, L. J., & Meltzer, A. S. (1990). *Workplace basics: The essential skills employers want.* San Francisco, CA: Jossey-Bass.

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

Center for Occupational Research and Development (CORD). (1990). *Applied mathematics* [Curriculum]. Waco, TX: CORD.

Clark, H. (2015). *Building a common language for career readiness and success: A foundational competency framework for employers and educators.* (ACT Working Paper Series WP-2015-02). Iowa City, IA: ACT.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* Routledge.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Conover Company. (1991). *Workplace literacy system* [Brochure]. Omro, WI: Conover Company.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological Tests and Personnel Decisions.* Urbana, IL: University of Illinois Press.

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

Educational Testing Service (ETS). (1975). *Cooperative assessment of experiential learning. Interpersonal learning in an academic setting: Theory and practice.* (CAEL Institutional Report No. 2). Princeton, NJ: ETS.

Electronic Selection Systems Corporation (ESSC). (1992). *AccuVision systems for personnel selection and development* [Assessment]. Maitland, FL: ESSC.

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

Greenan, J. P. (1983). *Identification of generalizable skills in secondary vocational programs* [Executive Summary]. Springfield, IL: Illinois State Board of Education.

Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. A. Lazarsfeld, S. A. Star, & J. A. Clausen (Eds.), *Measurement and Prediction* (pp. 60–90). Princeton, NJ: Princeton University.

Iowa Workforce Development. (2012). Skilled Iowa: Iowa NCRC 2012. Retrieved from portal.iowaworkforce.org/SkilledIowaLinks/.

John, O. P., & Srivastava, S. (1999). The big-five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and Research* (2nd ed.) (pp. 102–138). New York, NY: Guilford.

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

Lamb, R. R., & Prediger, D. J. (1981). *Technical report for the Unisex Edition of the ACT Interest Inventory (UNIACT).* Iowa City, Iowa: ACT.

Litwiller, B. J., Kyte, T., & LeFebvre, M. L. (2014, May). *Subject matter experts as a source of data for utility analysis. Poster presented at the meeting of the Society of Industrial Organizational Psychology. Honolulu, HI.*

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

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

McLarty, J. R. (1992, August). *WorkKeys: Developing the assessments.* In J. D. West (Chair), *WorkKeys supporting the transition from school to work.* Symposium presented at the meeting of the American Psychological Association, Washington, DC.

McLarty, J. R., & Vansickle, T. R. (1997). Assessing employability skills: The WorkKeys system. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 293–325). Mahwah, NJ: Lawrence Erlbaum Associates.

National Assessment Governing Board (NAGB). (2010a). *The alignment of the NAEP grade 12 mathematics assessment and the ACT WorkKeys applied mathematics assessment.* Unpublished report, Iowa City, IA: NAGB.

National Assessment Governing Board (NAGB). (2010b). *The alignment of the NAEP grade 12 reading assessment and the WorkKeys reading for information assessment.* Iowa City, IA: NAGB.

National Assessment Governing Board (NAGB). (2014). *The content alignment between the NAEP and WorkKeys assessments.* Iowa City, IA: NAGB.

Partners for a Competitive Workforce. (2012). *Work readiness collaborative.* Retrieved from www.competitiveworkforce.com/Work-Readiness.html.

Ryan, A. M. (2001). Explaining the black-white test score gap: The role of test perceptions. *Human Performance*, *14*(1), 45–75.

Secretary's Commission on Achieving Necessary Skills (SCANS). (1990). Identifying and describing the skills required by work. Washington, DC: US Government Printing Office.

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

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 Research Report 97-7). Iowa City, IA: ACT.

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

US Bureau of Labor Statistics. (2013). *Occupational employment projections 2012–2022.* Retrieved from www.bls.gov.

Workforce Investment Board of Southwest Missouri (WIBSM). (2013). *Average earnings, employment, & retention by national career readiness certificate & education levels.* Joplin, MO: WIBSM.

Zimmer, T. E. (2012). Assessment of the ACT WorkKeys job skills test. *InContext*, *13*(6). Retrieved from www.incontext.indiana.edu/

Appendix

Tables A1 through A8

Table A1. Summary of ACT WorkKeys Criterion Validation Studies by Assessment and Outcome

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.

American Health Care Association Study and Critical Thinking

ACT Research partnered with the American Health Care Association (AHCA) in 2010 to conduct a study of certified nurse assistants (CNAs) to determine if and how ACT WorkKeys assessment scores relate to job performance. CNAs were asked to complete **Reading for Information**, **Applied Mathematics**, **Locating Information**, and **Talent** (ACT, 2012b). CNA job performance was measured using the ACT Supervisor Survey, a standardized supervisory rating inventory. ACT conducted two confirmatory studies of the relation between a composite of level scores on three different ACT WorkKeys assessments (Reading for Information, Locating Information, and Applied Mathematics) and supervisor ratings of critical thinking performance of incumbent employees as measured by the ACT Supervisor Survey. A composite level was computed by summing the individual level scores across the three ACT WorkKeys tests, and cutoffs were established at composite level scores of 15, 16, and 17. In addition, a floor or minimum level score of 4 was required for each test. The combination of both a composite and floor was desired in order to minimize but not eliminate the compensatory result of the assessments. A chi-square test of independence was performed to examine the relation between ACT WorkKeys composite level and supervisor ratings of critical

thinking performance on the job. Using the composite level cutoff of 15, the relation between these variables was significant, X^2 (1, N = 61) = 4.643, p < .05 meaning that higher composite ACT WorkKeys scores were better predictors of supervisor ratings of on-the-job critical thinking performance.

ACT WorkKeys Healthcare Study and Job Performance Ratings, Effort, Task Proficiency

In 2014, ACT Research concluded a multi-year study with a midwestern employer in the healthcare industry to investigate the use of the ACT WorkKeys system (ACT, 2013a and 2014d). First, a job analysis was conducted, which resulted in recommended skill levels as follows: Locating Information, Workplace Observation, and Listening for Understanding for registered nurses; Reading for Information and Workplace Observation for medical assistants; Listening for Understanding for patient care assistants; Locating Information and Workplace Observation for phlebotomists; and Locating Information, Workplace Observation, and Listening for Understanding for nutrition services (ACT, 2013a and 2014d). A concurrent validation study was conducted by job title to investigate how well the ACT WorkKeys cognitive assessments predicted various facets of job performance, as well as overall criterion of job performance. The results of the study are presented in Table A2 (ACT, 2013a and 2014d).

Table A2. Validity Correlations between ACT WorkKeys Cognitive Assessments and Job Performance

| Occupation | ACT WorkKeys Assessment | n | Job Performance Criteria | | | |
			Overall Performance	Effort	Task Proficiency	Critical Thinking
Patient Registration	Applied Mathematics	–	–	–	–	–
	Listening for Understanding	108	0.03	0.04	0.14	0.22*
	Locating Information	108	0.16	0.09	0.17	0.14
	Reading for Information	–	–	–	–	–
	Workplace Observation	110	0.08	0.10	0.21*	0.20*
Health Unit Coordinators	Applied Mathematics	–	–	–	–	–
	Listening for Understanding	66	0.38**	0.33**	0.27*	0.27*
	Locating Information	66	0.25*	0.41**	0.34**	0.37**
	Reading for Information	–	–	–	–	–
	Workplace Observation	66	0.10	0.13	0.09	0.22
Licensed Practical Nurses	Applied Mathematics	–	–	–	–	–
	Listening for Understanding	64	0.15	0.20	0.04	0.16
	Locating Information	65	0.21	0.32*	0.27*	0.23
	Reading for Information	–	–	–	–	–
	Workplace Observation	65	0.34**	0.35**	0.16	0.40**
Registered Nurses	Applied Mathematics	143	0.14	0.07	0.12	0.09
	Listening for Understanding	142	0.19*	0.13	0.14	0.20*
	Locating Information	143	0.21*	0.24**	0.23**	0.21*
	Reading for Information	142	0.1	0.01	0.09	0.04
	Workplace Observation	142	0.24**	0.17*	0.19*	0.22**

Note: * p ≤ .05, ** p ≤ .01.

ACT WorkKeys Healthcare Critical Thinking Study

A study was conducted with ACT WorkKeys assessment data and supervisor performance evaluations from a sample of 573 incumbent employees of a midwestern employer in the healthcare industry (ACT, 2013b). Four hundred ninety-six of the 573 incumbents completed three ACT WorkKeys tests (**Reading for Information**, **Locating Information**, and **Applied Mathematics**) and had completed the ACT Supervisor Survey. A chi-square test of independence was performed to examine the relation between ACT WorkKeys composite level scores and supervisor ratings of critical thinking performance on the job. Composite cutoffs at levels 15, 16, and 17 (the summed level scores across the three ACT WorkKeys assessments) were used as well as a minimum test performance level on each of the tests. The relation between these variables was significant, X^2 (1, N = 496) = 24.79, p < .01. Table A3 presents the frequencies of observed and expected counts for individuals. These results imply that individuals with an ACT WorkKeys composite level of 15 or higher were more likely to have above average supervisor ratings of critical thinking performance on the job than below average supervisor ratings (ACT, 2012c).

Table A3. ACT WorkKeys Composite Score and Critical Thinking Job Performance

		Below Average Critical Thinking	Above Average Critical Thinking	Total
Composite 14 or Lower	Count	157	125	282
	Expected Count	129.6	152.4	282.0
Composite 15 or Higher	Count	71	143	214
	Expected Count	98.4	115.6	214.0
Total	Count	228	268	496
	Expected Count	228.0	268.0	496.0

City of Albuquerque Study and Job Performance, Absenteeism, Violations, Injury Claims, Safety Incidents and Turnover

In 2013, ACT Research partnered with the City of Albuquerque, New Mexico to evaluate the relationship between **Reading for Information**, **Locating Information**, **Workplace Observation**, and job performance outcomes for its motorcoach operator (MCO) position (n = 92) (ACT, 2014e). ACT WorkKeys tests were administered as a component in the hiring process for these individuals and to incumbents hired for the MCO position immediately prior to the use of ACT WorkKeys for comparison purposes. This validation effort examined both the relationship between ACT WorkKeys test scores and various outcomes and compared differences on outcomes for both groups of MCOs. MCO supervisors completed a performance evaluation on MCOs hired using ACT WorkKeys. The city also provided data on employee absenteeism, substance abuse policy violations, personal injury claims, 311 calls, safety incidents, and turnover for both groups.

Reading for Information was significantly related to MCO task proficiency, following rules, overall job performance, safety incidents, and customer complaint calls to 311. Locating Information was significantly related to MCO absenteeism and turnover. Workplace Observation was significantly related to supervisor ratings of safety and turnover (see Tables A4 and A5).

Table A4. Associations between ACT WorkKeys and Job Performance Criteria

Job Performance Criteria	ACT WorkKeys Assessment		
	Reading for Information	Locating Information	Workplace Observation
Task Proficiency	0.27*	0.11	0.11
Extra Effort	0.14	0.12	0.13
Working With Others	0.11	0.10	0.17
Customer Service	0.17	0.10	0.10
Resilience	0.15	0.17	0.14
Learning	0.11	0.15	0.16
Critical Thinking	0.14	0.10	0.09
Following rules	0.18*	0.14	0.14
Safety	0.11	0.14	0.23*
Overall Job Performance	0.25*	0.18	0.14

Note: * $p < .05$. n = 68. Job performance as measured by ACT WorkKeys Appraise.

Table A5. Associations between ACT WorkKeys and HRIS Outcomes

HRIS Outcome	ACT WorkKeys Assessment		
	Reading for Information	Locating Information	Workplace Observation
Safety Incidents	-0.15*	-0.11	-0.11
311 Complaints	-0.24*	-0.11	0.10
Absenteeism	-0.13	-0.22*	0.13
Turnover	-0.12	-0.33*	-0.31*

Note: * $p < .05$. n = 51. HRIS = Human Resource Information System data collected by employer.

When comparing groups, the ACT WorkKeys group had fewer complaint calls made to 311, $x^2(1, n = 186) = 342$, $p < .05$. The ACT WorkKeys group had fewer preventable and unpreventable safety incidents, $x^2(1, n = 186) = 53.12$, $p < .05$. The ACT WorkKeys group also had fewer personal injury claims, $x^2(1, n = 186) = 4.16$, $p < .05$. There were no significant differences between substance abuse policy violations or turnover between the comparison and ACT WorkKeys groups.

In addition, there were no significant differences found for protected groups for the ACT WorkKeys assessments or ACT WorkKeys Appraise performance evaluation tool. With one exception, there were no differences in Reading for Information, Locating Information or Workplace Observation scores for examinees in protected groups (see Tables A6, A7, and A8). The only difference found for protected groups was in Locating Information. Examinees less than 40 years of age scored significantly higher than examinees aged 40 years or older. Re-analyzing using the continuous variable of age and Locating Information level scores and scale scores also demonstrated a slight difference in scores. Further exploration revealed this difference to be less than half a standard deviation and having a small effect.

Table A6. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Hispanic Status

ACT WorkKeys Assessment	Non-Hispanic			Hispanic				
	n	M	SD	n	M	SD	t	df
Workplace Observation	66	2.28	0.55	28	2.18	0.661	.78	91
Locating Information	65	4.07	0.51	28	4.04	0.611	.25	90
Reading for Information	64	5.26	0.89	28	5.12	0.696	.83	89

Table A7. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Gender

ACT WorkKeys Assessment	Male			Female				
	n	M	SD	n	M	SD	t	df
Workplace Observation	71	2.23	.66	23	2.22	.42	.05	92
Locating Information	71	4.07	.59	22	4	.54	.51	91
Reading for Information	70	5.19	.80	22	5.23	.75	-.22	90

Table A8. Results of t-test and Descriptive Statistics for ACT WorkKeys Mean Level Scores by Age

ACT WorkKeys Assessment	Less than 40 Years of Age			40+ Years or Older				
	n	M	SD	n	M	SD	t	df
Workplace Observation	40	2.33	.73	54	2.15	.49	1.40	92
Locating Information	41	4.20	.60	52	3.94	.60	2.21*	91
Reading for Information	40	5.33	.83	52	5.10	.75	1.39	90

Note: * $p < .05$

Manufacturing Skill Standards Council's Certified Production Technician Certificate and ACT WorkKeys

A 2012 research study was conducted in coordination with the Southwest Ohio Workforce Investment Board to evaluate the impact of a grant from the US Department of Labor to provide training for displaced manufacturing workers in several counties throughout Ohio (ACT, 2013c). The grant comprised a training program and certification process for the Manufacturing Skill Standards Council's (MSSC) Certified Production Technician (CPT) certificate. The purpose of the MSSC CPT program is to recognize through certification (entry-level through front-line supervisor positions) individuals who demonstrate mastery of the core competencies of manufacturing production at the front-line. Four independent assessments were employed in the MSSC CPT Program: Safety, Quality Practices & Measurement, Processes & Production, and Maintenance Awareness. The ACT WorkKeys assessments **Reading for Information** and **Applied Mathematics** were used as a screening tool for entrance into the MSSC CPT Program. A level of 4 or higher was required for entrance into the program.

When used in combination with ACT WorkKeys assessments, the pass rates for all four MSSC CPT assessments were over 90%. About 59% of individuals with an ACT WorkKeys Applied Mathematics score of 3 passed the CPT as compared with 87% of individuals with an ACT WorkKeys Applied Mathematics level of 5 (n = 439).

The MSSC CPT pass rate for individuals with an ACT WorkKeys Reading for Information level of 4 was 68% as compared with 89% for those with an ACT WorkKeys Reading for Information

level of 5 (n = 436). Preliminary findings for an expanded statewide program show that individuals who completed both the ACT NCRC and MSSC CPT certificate had a 45% employment rate as compared to a 39% employment rate for those who did not complete both.

ACT WorkKeys and Secondary Education Grades and Course Selection

Two studies investigated the relationship between ACT WorkKeys assessments and secondary education outcomes (ACT, 2010). One study of secondary education outcomes focused on career technical center high school students in Grand Rapids, Michigan. Significant relationships were found between ACT WorkKeys scores and career technical course grades for **Reading for Information** (n = 2,223, r = .22, p ≤ .05), **Applied Mathematics** (n = 2,162, r = .21, p ≤ .05), Locating Information (n = 1,216, r = .21, p ≤ .05), and a **composite of all three assessments** (n = 951, r = .25, p ≤ .05).

The second study investigated the relationship between the study of core vs. non-core coursework and course selection with ACT WorkKeys **Reading for Information** and **Applied Mathematics** scores by matching ACT WorkKeys examinee data with ACT examinee data (n = 175,000). Study results indicated that students who took Algebra I, Algebra II, and Geometry were twice as likely to score at level 5 or higher on Applied Mathematics than were students who took fewer mathematics courses (63% vs. 34%). More than three-quarters of the students (79%) who took higher-level mathematics courses beyond Algebra I, Algebra II, and Geometry scored at level 5 or higher on Applied Mathematics; they were more than twice as likely to score a 5 or higher as students who took lower-level mathematics courses that did not include Algebra I and II and Geometry. Students taking Algebra I and II and Geometry, Biology, Chemistry, and two courses in history were twice as likely to score at level 5 or higher on Reading for Information than students who took fewer core courses (61% vs. 29%). Additionally, students taking more core curriculum courses were almost twice as likely to score level 5 or higher than were students who took fewer core courses (61% vs. 29%).

ACT WorkKeys and Postsecondary Education GPA, Course Grades, and Persistence

Three studies investigated the relationship between postsecondary education outcomes and ACT WorkKeys assessment scores (ACT, 2010).

In 2010, outcome data for postsecondary occupation-based programs were matched to ACT WorkKeys assessment data (ACT, 2010). The study focused on occupation-based programs requiring a two-year degree or less. Significant relationships were found between ACT WorkKeys assessment scores and cumulative GPA across programs for **Reading for Information** (n = 285, r = .15, p ≤ .05), **Applied Mathematics** (n = 296, r = .19, p ≤ .05), and a **composite of Reading and Math** (n = 278, r = .18, p ≤ .05). Over half of the students scoring at a level 5 or higher in Reading for Information (54.9%) or Applied Mathematics (55.8%) achieved grades at 3.0 or higher in their programs. A smaller percentage of students scoring at a level 4 or lower achieved similar grades (40.9%, 40.5%).

In 2010, ACT studied occupation-based program data from Wright College (Chicago, IL) by matching ACT WorkKeys examinee data to a matched Academic Advancement Program (AAP) Graduates/ Wright data file (ACT, 2010). Significant relationships were found between ACT WorkKeys scores and cumulative GPA overall for **Reading for Information** (n = 1,251, r = .25, p ≤ .05) and **Applied**

Mathematics (n = 1246, r = .28, p ≤ .05). Over twice as many students who scored at a level 5 in Reading for Information (43.5%) or Applied Mathematics (45.6%) achieved grades of 3.0 or higher in their programs than students who scored at level 3 (18.5%, 20.3%).

Other significant relationships with course grades were found with broad academic programs, such as Health Sciences for both Reading for Information (n = 391, r = .27, p ≤ .05) and Applied Mathematics (n = 391, r = .29, p ≤ .05); Business/Management/Marketing for both Reading for Information (n = 430, r = .30, p ≤ .05) and Applied Mathematics (n = 430, r = .31, p ≤ .05); and Human Services for both Reading for Information (n = 349, r = .19, p ≤ .05) and Applied Mathematics (n = 345, r = .24, p ≤ .05) (ACT, 2010).

In a third study, ACT WorkKeys assessment scores were matched to AAP Graduates data, matched to college outcomes data (COD) for two-year and four-year colleges, and ACT-matched National Clearinghouse Student data. Significant correlations were found between **composite** ACT WorkKeys scores for **Reading for Information** and **Applied Mathematics** and first-year college GPA (n = 10,744, r = .30, p ≤ .05). Students who achieved ACT WorkKeys scores of 5 or higher in both Applied Mathematics and Reading for Information obtained a mean first-year college GPA of 2.8. This is nearly a full grade higher than the GPA of 2.0 for students who achieved scores of 3 or lower. Significant correlations were also found between composite ACT WorkKeys scores and college persistence to second year (n = 277,631, r = .23, p ≤ .05). Almost 90% of students who achieved ACT WorkKeys scores of 5 or higher persisted to their second year of college, whereas 59% of students who achieved ACT WorkKeys scores of 3 or lower persisted to the second year of college (ACT, 2010).

Gallo/ACT WorkKeys and Return on Investment Study

In 2011, E. & J. Gallo Winery partnered with ACT to examine the utility of ACT WorkKeys assessments to predict task and safety performance (ACT, 2011b). Incumbent employees completed **Reading for Information**, **Applied Mathematics**, **Locating Information**, and **Talent** (n = 139). Incumbents' supervisors completed the ACT Supervisor Survey as a measure of job performance. The employer provided additional data regarding the number of disciplinary actions for incumbents.

Employee ACT WorkKeys scores significantly predicted overall job performance as measured by the ACT Supervisor Survey (n = 68): Reading for Information (r = .27, p ≤ .05), Applied Mathematics (r = .41, p ≤ .05), Locating Information (r = .20, p ≤ .05), and a composite of all three assessments (r = .29, p ≤ .05). Significant relationships were found with employee Talent scale scores and number of disciplinary actions (n = 46).

Significant relationship were also found for number of disciplinary actions with the Talent scales (n = 46) of Carefulness (r = -.29, p ≤ .05), Cooperation (r = -.30, p ≤ .05), Discipline (r = -.36, p ≤ .05), Order (r = -.24, p ≤ .05), as well as the Talent Work Discipline index (r = -.38, p ≤ .05).

Additional analysis was conducted 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 decreased hiring needs due to increased performance, and a 19.3% reduction in hiring needs due to increased employee safety.

New Options New Mexico Employer Return on Investment Survey

In 2012, researchers in the New Options New Mexico project surveyed a national sample of twelve medium- to large-size employers who used the **ACT WorkKeys** system (Mayo, 2012). Preexisting Return on Investment data for each employer was collected and outcomes compared both pre- and post-ACT WorkKeys implementation. The study reported 25–75% reduction in turnover, 50–70% reduction in time-to-hire, 70% reduction in cost-to-hire, and 50% reduction in training time.

Skilled Iowa ACT NCRC Employment and Wage Outcome Report

In 2012, the Iowa Department of Workforce Development partnered with ACT to investigate employment and wage outcomes for participants who earned the **ACT NCRC** as part of the Skilled Iowa program (Iowa Workforce Development, 2012). ACT NCRC recipient and wage data from 2010 and 2011 were included in the analysis (n = 9,346). The results indicated that over half (53.3%) to nearly two-thirds (63.4%) of the ACT NCRC recipients secured employment within the first quarter after earning the ACT NCRC. Wages increased for those employed within the first two quarters following the receipt of the ACT NCRC. Unemployed silver certificate recipients secured employment much more quickly than bronze certificate recipients. Most of the unemployed certificate recipients secured employment within the first three quarters following the award.

Indiana ACT WorkKeys Employment and Wage Outcome Study

In 2012, the Indiana Department of Workforce Development used state administrative unemployment insurance and wage records, public post-secondary records, Workforce Investment Act (WIA) case management records, and adult education records to examine the impact of ACT WorkKeys assessments on time-to-employment and wages-after-testing (Zimmer, 2012). ACT WorkKeys examinees included in the study were either enrolled in training education and training programs or actively looking for work. Time-to-employment was measured as the number of weeks from the date of the assessment until the individual appeared on a wage record (n = 200,044). Wages-after-testing was measured as the average wage collected in the second and third quarters from the date of initial employment (n = 265,229). Six ACT WorkKeys assessments were included in the study (**Reading for Information**, **Locating Information**, **Observation, Teamwork**, **Applied Mathematics**, and **Applied Technology**).

The analysis examined the influence of score on the assessments with regard to time to employment and wages, while controlling for gender and highest educational attainment. Two groups of individuals were examined separately: a younger group (aged 16 to 18) and an older group (aged 19 or older).

The time-to-employment analysis indicated a reduction of time to employment with higher ACT WorkKeys scores. For the older age group, a higher ACT WorkKeys Applied Mathematics level score resulted in five fewer days to employment, four fewer days for Reading for Information and Observation, three days for Locating Information, and two fewer days for Teamwork and Applied Technology.

Weaker effects were found for reduction of employment time for the younger age group. The wages-after-testing results showed a strong relationship between higher scores and higher wages for the older age group. For the older age group, a higher ACT WorkKeys level score resulted in a quarterly increase of $399 for Applied Technology, $373 for Teamwork, $350 for both Locating Information and Observation, as well as $275 and $204 for Applied Mathematics and Reading for Information, respectively. Weaker effects were found for wages-after-testing for the younger age group.

Ohio G*Stars ACT NCRC and Employment Outcomes Study

Analysis was conducted in 2012 by Partners for a Competitive Workforce (a partnership in the Ohio, Kentucky, and Indiana tri-state region), which examined employment and earnings outcomes for **ACT NCRC** earners (Partners for a Competitive Workforce, 2012). The study used the Ohio Unemployment Insurance Tax data that was matched with individuals in the G*Stars system (Workforce Investment Act case management system) who received workforce services during the first quarter of 2007 through the second quarter of 2012 (n = 275). Results indicated that the total annual estimated increase in potential earnings was $7,476 for ACT NCRC recipients compared to $2,916 for those with an associate's degree and $10,932 for those with an occupational license. The study found an earnings increase for ACT NCRC completers for four industry sectors.

Southwest Missouri Workforce Investment Board ACT NCRC and Workforce Outcomes

A research partnership between the Workforce Investment Board of Southwest Missouri, Mayo Enterprises, and the Missouri Division of Workforce Development examined the average earnings (n = 3,709), entered employment rate[9] (n = 6,968), and retention rates (n = 4,485) by **ACT NCRC** and education levels of adult clients enrolled in the workforce investment system over three years using state case management data (Workforce Investment Board of Southwest Missouri, 2013). The study found increased earnings for those with higher ACT NCRC scores across all levels of education. Among those with less than a high school diploma, there was a 24% increase in average earnings for those with a silver ACT NCRC compared to a bronze, and a 14% increase in earnings for ACT NCRC earners with a gold level compared to a silver certificate. For high school graduates, there was a 13% increase in wages for silver ACT NCRC holders compared to bronze and a 14% increase between silver and gold.

For associate's and bachelor's degree holders, the wage increase between silver and gold was 12% and 15%, respectively. Analysis of entered employment rates found that higher ACT NCRC levels results in increased likelihood to attain employment. Analysis of employment retention also indicated that higher ACT NCRC levels increased an individual's likelihood of staying in a job.

Schmidt & Sharf Review of ACT WorkKeys

In a 2010 commissioned review of ACT WorkKeys validation evidence, two external experts concluded that an employer's use of the ACT NCRC (or any three or more of the ACT WorkKeys skills assessments) could be shown to be valid under current professional standards and the *Uniform Guidelines*, without the need for a local validity study, based on meta-analytic validity generalization research and related research (Schmidt & Sharf, 2010). The report stated that the cumulative research findings of numerous professionals over many years establishes that the assessments should meet the requirements for criterion-related validity, content validity, and construct validity, through application of validity generalization research findings and related research.

---

[9] Entered employment rate is the number of adults who were employed within the first quarter after exiting the workforce investment system divided by the total number of adult exiters.

ACT®

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

For more information, visit www.act.org.