

# WorkKeys<sup>®</sup>

## PERFORMANCE ASSESSMENT

User and Technical Guide



**ACT<sup>®</sup>**



## Contents

<b>1 Introduction</b> .....	1
What Is Integrity Testing? .....	1
Brief History of Integrity Testing. ....	2
<b>2 About the WorkKeys Performance Assessment</b> .....	3
What is the WorkKeys Performance Assessment? .....	3
Decision Making Based on the WorkKeys Performance Assessment. ....	3
Using the WorkKeys Performance Assessment for Selection: Examples .....	3
What Does the Performance Assessment Measure? .....	5
Benefits of Using the Performance Assessment .....	7
<b>3 Construction and Evaluation of the WorkKeys Performance Assessment</b> .....	10
Reliability .....	10
Validity .....	11
Fakeability .....	13
Applicant Reactions .....	13
Adverse Impact. ....	13
Compliance with Guidelines and Standards. ....	14
<b>4 Administration of the WorkKeys Performance Assessment</b> .....	15
Administration of the WorkKeys Performance Assessment. ....	15
Testing Environment .....	16
Examinee Setup in the Validus Virtual Test Center .....	16
Confidentiality Agreement. ....	16
Test Instructions for Examinee. ....	17
Accommodations for Examinees for Whom English is a Second Language ...	18
Reporting the Results of the WorkKeys Performance Assessment .....	18
<b>5 Interpreting Employer and List Reports</b> .....	19
Employer Reports .....	19
Sample Employer Reports. ....	20
High Desirability .....	21
Moderate Desirability .....	23
Low Desirability .....	26
Other Examples of Employer Reports .....	28
Sample List Reports. ....	29
<b>Appendix: Development of the WorkKeys Performance Assessment and Research Findings</b> .....	31
<b>References</b> .....	51
<b>Support and Customer Service</b> .....	54

## Figures

Figure 1	Welcome Screen for the Performance Assessment. . . . .	15
Figure 2	Confidentiality Agreement . . . . .	16
Figure 3	Demographics . . . . .	17
Figure 4	Instructions for the Performance Assessment . . . . .	17
Figure 5	Employer Report Reflecting High Scores. . . . .	21
Figure 6	Employer Report Reflecting Moderate Scores. . . . .	23
Figure 7	Employer Report Reflecting Uneven Subscale Scores . . . . .	25
Figure 8	Employer Report Reflecting Low Scores . . . . .	26
Figure 9	Employer Report with Inconsistent Reporting Flag . . . . .	28
Figure 10	List Report Sorted by Performance Index Recommendation Category . . . . .	29
Figure 11	List Report Sorted by Name . . . . .	30
Figure A1	Frequency Distribution of General Work Attitudes Raw Scores. . .	36
Figure A2	Frequency Distribution of Risk Reduction Raw Scores. . . . .	37
Figure A3	Frequency Distribution of Performance Index Scores. . . . .	37

## Tables

Table 1	Return on Investment Approximations under Various Scenarios . .	9
Table 2	Validity Estimates for the Development Sample of the Performance Assessment . . . . .	12
Table 3	Levels of Desirability and Associated Score Ranges . . . . .	20

## Appendix Tables

Table A1	Intercorrelations of the Supervisor Performance Rating Scales . .	33
Table A2	Demographic Characteristics of the Normative Sample . . . . .	34
Table A3	Descriptive Statistics at the Item Level. . . . .	36
Table A4	Descriptive Statistics at the Scale Level . . . . .	36
Table A5	Observed Validity Correlations and Internal Consistency Reliability for Development and Cross-validation Samples . . . . .	38
Table A6	Validity Correlations for the Development Sample. . . . .	40
Table A7	Validity Correlations for the Cross-validation Sample. . . . .	41
Table A8	Confidence Intervals of Validity Estimates for the Development Sample. . . . .	42
Table A9	Confidence Intervals of Validity Estimates for the Cross-validation Sample . . . . .	44
Table A10	Percentiles for the Performance Assessment Scales. . . . .	46
Table A11	Classification Rates for the Normative Sample Using the Performance Index . . . . .	47
Table A12	Performance Correlations with Demographics . . . . .	48
Table A13	Classification Rates for Caucasians Using the Performance Index . . . . .	49
Table A14	Classification Rates for Racial/Ethnic Minorities Using the Performance Index . . . . .	49

# 1 Introduction

Organizations are constantly looking for tools that will help them acquire the best candidates for a job and avoid the negative consequences of hiring the wrong person. Employers need individuals who have the skills required to perform well on their assigned duties. However, good performance also includes honesty, dependability, and trustworthiness. Employees are also expected to follow company rules and procedures. ACT, a leader in test development and assessment, has added a suite of personal skills tools as part of the WorkKeys® line. These assessments can assist employers with critical personnel decisions. For example, the WorkKeys Performance Assessment provides information on the integrity of an applicant. Integrity testing provides employers with a time- and cost-efficient method of screening candidates on their attitudes toward work, safety, and overall performance in the workplace.

## What Is Integrity Testing?

Integrity testing is used in employment selection procedures. Employers administer tests to prospective employees to assess behaviors related to work performance. The tests identify less desirable candidates, particularly those who are more likely to engage in counterproductive work behaviors. Employers also use the tests to predict the quantity and quality of employee productivity. The use of integrity tests has increased as employers recognize the importance of considering work attitudes and personality in hiring decisions. Responding to a recent survey, more than 90% of human resources and recruiting professionals indicated that personality is “important” or “critically important” when making a hiring decision (Hiredesk, 2006). For instance, two-thirds of large companies (\$1 billion or more in annual revenue) increased the use of prescreening tests, and integrity assessments have seen a 60% increase as part of an employers’ selection process in the past five years (Spherion, 2006).

Integrity can be measured in two ways—directly and indirectly. The more direct way, also referred to as “overt,” is to ask questions related to an individual’s attitudes and beliefs, and to ask direct questions about behavior. For example, a question may address past illegal or dishonest behavior. The indirect way, frequently referred to as “covert,” is to ask questions that tap an individual’s personality on dimensions that broadly predict counterproductive behaviors. These questions are used to *disguise* the purpose of the questions in an effort to get more accurate information from the examinee. Essentially, overt content measures attitudes, beliefs about behavior, and direct admissions of behavior while covert items measure latent personality traits (i.e., conscientiousness, agreeableness, emotional stability) (Sackett & Wanek, 1996; Ones, Viswesvaran, & Schmidt, 1993). Integrity tests measure both overt and covert aspects of an individual’s personality in order to broadly capture the true scope of each person and to predict, as accurately as possible, the potential for counterproductive work behaviors.

While valuable in many respects, integrity tests are most important for the associated cost-savings realized by organizations that adopt these assessment tools as part of their selection process. Employers rely on integrity tests to identify good, reliable potential employees and make smart hiring decisions that can save valuable company resources in the long run.

In general, the prevailing belief is that job performance is a function of several components, including the knowledge and skills necessary to perform job tasks, the attitude and general acceptance of the job, and the absence of counterproductive work behaviors (Tett, Jackson, & Rothstein, 1991; Sackett, 1994; Rotundo & Sackett, 2002). That is:

$$\text{Job performance} = f(\text{task ability} + \text{general work attitude and citizenship} - \text{counterproductive work behaviors})$$

### **Brief History of Integrity Testing**

Integrity tests have been in existence since the 1940s, but two events have contributed to their increased use. The 1988 Employee Polygraph Protection Act effectively banned the use of the polygraph in employment settings, and the 1991 Civil Rights Act requires demonstration of the job relatedness of a test when there is evidence of adverse impact. Although integrity tests are broadly used in a wide variety of settings, they are most commonly used for screening applicants for jobs where employees have access to money and/or merchandise, as well as construction, manufacturing, retail, and financial operations (Sackett, 1994). Recent estimates have placed the use of integrity tests at over five million administered per year.

## 2

# About the WorkKeys Performance Assessment

### What is the WorkKeys Performance Assessment?

The WorkKeys Performance Assessment is an integrity test used to screen potential employees for work behaviors that might be problematic. The Performance Assessment can also identify personality and behavior characteristics that are important in the workplace. The report generated from each assessment provides a performance index that helps human resources staff in their selection decisions. This assessment identifies candidates who might be prone to counterproductive work behaviors that may include:

- Absenteeism
- Theft
- Violation of work rules
- Hostility in the workplace
- General work attitude and conduct

### Decision Making Based on the WorkKeys Performance Assessment

The purpose of the Performance Assessment is to provide an easily administered and scored self-report measure of integrity that can be used in employee screening and selection. Employers can use the Assessment results in concert with other selection procedures, including WorkKeys Foundational Skills tests, job interviews, work samples, and reference checks, to make final decisions about whether to hire a job candidate. The Performance Assessment is one of several integrity tests that have been validated in research studies. Like other integrity tests, the results of the Performance Assessment yield the most benefits when used in a “top-down” approach. That is, qualified job candidates with the highest scores are considered more likely to be successful in the job and are given more serious consideration for it. The Performance Assessment scores are used as part of a “multiple-hurdle” approach to selection. In this approach, applicants must achieve a certain cutoff score on the Performance Assessment before they are moved to the next stage of Foundational Skills testing or interviews.

### Using the WorkKeys Performance Assessment for Selection: Examples

Businesses need employees who are honest, dependable, and trustworthy, and they need an efficient way to hire them. The WorkKeys Performance Assessment provides employers with a powerful tool that helps identify the right candidate for the job. This assessment measures work behaviors that may serve to reduce productivity, such as theft, absenteeism, and violation of work rules. It can also identify attitudes and behavior characteristics that are important to the job, such as conscientiousness, agreeableness, and emotional stability.

Businesses have the flexibility to determine how to best use the Performance Assessment, depending on such factors as job requirements, employer needs, and market forces. For example, employers can use the test as a prescreening device for applicants or with other components of a selection system. In order to assist employers with this decision, ACT has created case scenarios. Employers can review these for guidance on how to use the Performance Assessment to meet their hiring needs.

### ***Case 1: Prescreening***

Prescreening job applicants is an excellent use of the Performance Assessment. In this context, employers can use the screening services of independent testing sites and use the test results to establish a pool of desirable applicants. Businesses can also administer the Assessment on site. Here is an example of how this works:

1. Applicants complete the Performance Assessment.
2. Top-scoring applicants are referred to the employer and continue the selection process. A cutoff score ensures that only those applicants who are at or above the cutoff point are referred to the employer.
3. The applicant proceeds through the remaining steps of the employer's selection system, which might include an application review, a knowledge test, and an interview.
4. Applicants with the highest scores in the employer's selection system are hired.

### ***Case 2: Assessment Set with Specific Cut Scores***

In this example, multiple tests are used and job candidates are required to meet or exceed the cutoff score on each test. This approach is most appropriate when the job requires a minimal amount of a certain set of key characteristics. For example, if an employer is selecting candidates for an accounting job, strong reading skills cannot compensate for poor work habits and weak math skills. Here is an example of how the Performance Assessment can be used with other WorkKeys tests:

1. The employer conducts application review for job candidates for such things as minimum experience or education requirements.
2. Applicants who pass the application review take the required tests, such as *WorkKeys Performance*, *Applied Mathematics*, and *Reading for Information*.
3. Only those applicants who meet or exceed the required scores on **all** three of the tests are scheduled for the last step of the selection system, such as an interview.
4. Applicants with the highest scores from the interview are hired.

### ***Case 3: Multiple Tests in Hurdles***

Employers may best meet their needs by setting up stages—called hurdles—in their selection system. In these cases, the employer arranges the selection system components to require job candidates to pass hurdles in succession. Candidates can only progress to the next hurdle if they pass the requirements of the prior hurdle. This process uses tests to progressively narrow the pool of applicants to the most qualified candidates. Here is an example of how it works:

1. Applicants complete the Performance Assessment, the first hurdle.
2. Only top-scoring applicants move to the second hurdle. If a cutoff score is used, only those who score at or above the desired cutoff continue to the second hurdle.
3. Applicants in the second group complete the tests, such as WorkKeys *Applied Mathematics* and *Reading for Information*, in the second hurdle.
4. Only those applicants who meet the cutoff scores for *both* skills continue to the third hurdle.
5. Applicants in the third group complete the last hurdle in the selection system, such as an interview.
6. Applicants from the third group with the highest scores are hired.

### **What Does the Performance Assessment Measure?**

The Performance Assessment contains 60 self-report questions written at the 5th-grade reading level and requires approximately 10 to 15 minutes to complete. Employers may administer the assessment to examinees as part of the screening and selection hiring process. It yields an overall score—the Performance Index—based on two subscale scores: General Work Attitudes and Risk Reduction.

#### ***General Work Attitudes***

General Work Attitudes refer to individuals' overall approach to their work and their work environment. This subscale focuses on individuals' overall knowledge of their job, their ability to communicate and relate to others, and their level of productivity. This subscale contains a mixture of direct and indirect items. The direct or “overt” content assesses an individual's attitude toward corporate misconduct, while the indirect or “covert” content includes items that tap personality traits and work-related behaviors. Poor work attitudes, as measured by this subscale, have been shown to predict a range of counterproductive work behaviors and poor job performance. Much of the covert content is a combination of three broad personality traits: conscientiousness, agreeableness, and emotional stability.

*Conscientiousness* refers to the tendency to be purposeful, dependable, determined, punctual, and reliable. Higher scores identify individuals who are dependable, are most likely to complete work on time, and can be relied upon to get their job done. Scores at the low end identify individuals who may have a more difficult time completing their work, getting to work on time, and may not always be dependable.

*Agreeableness* refers to the tendency to be altruistic, helpful, and friendly to coworkers. High scores on items that measure agreeableness identify individuals who are likely to trust their fellow coworkers, are generally good team players, and are likely to comply with company regulations. Scores at

the low end reflect individuals who may have a more difficult time interacting with their fellow employees, may be mistrustful, and unfriendly, or hostile to their coworkers.

*Emotional stability* refers to the ability to maintain one's composure and rationality in situations of actual or perceived stress. High scores reflect a greater ability to handle stress, remain confident in existing abilities to cope, and handle negative situations. Low scores reflect a lower ability to handle stressful situations, reduced confidence in abilities, and may suggest weaker coping skills in the work environment.

Scores from the General Work Attitudes subscale can be related to:

- Productivity
- Supervisor ratings
- Aggression
- Resilience to work-related stress
- Team orientation

Sample items, which measure an examinee's attitudes about their work and fellow coworkers, include:

- *During heated arguments, I may become so agitated that I start yelling at coworkers.*
- *I am easily irritated by coworkers.*
- *I don't prioritize work activities.*
- *I have worked slowly on purpose to get overtime.*

### ***Risk Reduction***

This subscale contains items that focus on compliance with safety rules and procedures, as well as other expectations to avoid work-related accidents and unnecessary risk-taking in a work environment. Thus, the risk reduction subscale focuses on an individual's attitudes about safety procedures or improper operation of machinery.

Scores can be related to the following:

- Work-related accidents
- Respect for professional protocol
- Outbursts of physical and/or verbal aggression or hostility
- Coworker complaints about conduct and/or harassment

Sample items, which measure an examinee's approach to safety rules and regulations, include:

- *Some safety regulations are overprotective and should not be followed.*
- *Most job-related accidents just can't be prevented.*
- *I couldn't care less what happens to others around me.*

## **Benefits of Using the WorkKeys Performance Assessment**

The WorkKeys Performance Assessment can be a valuable tool for organizations that wish to improve the efficiency of their personnel selection procedures and save thousands of dollars from costly personnel mistakes. Using an objective and validated test, such as the Performance Assessment, results in fair selection procedures since subjective biases are largely avoided. The Performance Assessment is inexpensive and, since it can be completed online in 10-15 minutes, does not slow down the interviewing process. Organizations can expect large returns on their investment in the Performance Assessment, while improving quality and reducing risk associated within their workforces.

### ***The Utility of the WorkKeys Performance Assessment***

Most organizations incur enormous costs recruiting, screening, interviewing, and hiring individuals. In addition to these initial costs, organizations will incur additional costs when employees choose to leave, are terminated due to poor performance or counterproductive behaviors (e.g., absenteeism, theft), or experience accidents in the workplace caused by failure to comply with safety procedures. Organizations can reduce the costs associated with personnel selection and employee turnover by adopting efficient selection processes that identify individuals who demonstrate the highest likelihood of successful and sustained employment.

To help companies achieve this goal, ACT has developed a WorkKeys Performance Assessment tool, a survey that is administered to job candidates to measure their overall work attitude, as well as their tendency to engage in risky behaviors in the workplace. An expert team of researchers at ACT completed a national field study of the Performance Assessment. The study demonstrated that Performance Assessment scores are reliable predictors of which individuals are most likely to be successful on the job. The overall Performance Assessment score had a correlation of .434 with supervisor ratings of job performance. Higher Performance Assessment scores were associated with positive work attitudes, good organizational citizenship behaviors, and compliance with safety procedures. Organizations should use the Performance Assessment as part of a multiple hurdle selection system, which involves other selection activities including application reviews, interviews, work samples, and other skills tests. As discussed, Performance Assessment scores can be used to make personnel selection more effective, resulting in improved identification of potentially successful employees. Since the Performance Assessment can be administered to many applicants quickly and inexpensively, it serves to help organizations select applicants at lower costs to the company.

### ***Example: XYZ Manufacturing***

To explain the financial and practical benefits of selecting applicants using the Performance Assessment, let us assume a fictitious organization, XYZ Manufacturing, received applications from 200 individuals, and—based on their interviews, applications, and work samples—chose 100 candidates. Of these 100 candidates, assume that a quarter of the candidates (25) would ultimately become successful employees if hired at XYZ Manufacturing. We refer to these candidates as “future successes.” A sound selection system will ensure that the selected pool has a larger proportion of future successes than the original pool of applicants. If the Performance Assessment is used as the basis for selection, we can predict how well the selected pool will be improved with respect to percentage of future successes.

Let us assume that XYZ Manufacturing wishes to hire 10 individuals from its candidate pool of 100. Using the results of the national field study, we estimate that 54% of the selected individuals will be future successes. (For details on how this estimate is derived, see the notes at the end of this chapter, page 9.) Therefore, the likelihood of the new hires becoming successful employees has improved from 25% selection to 54% through the use of the Performance Assessment.

Depending on the average cost incurred for each unsuccessful employee relative to each successful employee (due to recruiting and training costs, lack of productivity, worker's compensation for injury, unemployment payouts, severance costs, etc.), the potential return on investment associated with using the Performance Assessment is significant. To continue with our example, assume that XYZ Manufacturing loses an average of \$5,000 for each unsuccessful employee, relative to each successful employee. In this hypothetical scenario, XYZ Manufacturing would save \$14,500 [ $10 \times (54\% - 25\%) \times \$5,000$ ] by screening the original 100 applicants using the Performance Assessment. This translates to a savings of \$145 per original applicant! Given that the per-applicant price of the Performance Assessment is \$15 in this hypothetical scenario, the assumed net return on investment is \$130 per applicant or \$13,000 per 100 applicants.

### ***Return on Investment from the Performance Assessment***

In the hypothetical example previously discussed, we showed that XYZ Manufacturing could save thousands of dollars by using the Performance Assessment to select from a large applicant pool. We made certain assumptions about the size and quality of their applicant pool, hiring practices, and average cost of hiring an unsuccessful employee. Naturally, the actual return on investment of the Performance Assessment will depend on the actual values of these variables, which vary across organizations.

To gain a more general understanding of the expected return on investment (ROI), we calculated the per-applicant ROI under several scenarios. The first row of Table 1 shows the ROI for the example involving XYZ Manufacturing. Subsequent rows report the ROI under other plausible scenarios. We assumed average costs of selection failure to be \$5,000 or \$10,000. In reality, these costs may be even greater; some experts estimate the average cost of selection failure to be at least an employee's full year's salary, while others put the costs as high as 2.5 times an annual salary. Besides the direct costs of selection, other indirect costs can be attributed to such factors as decreased satisfaction of successful incumbent employees, lower employee morale, tainting of the organization as a reputable employer, and loss of clientele.

Table 1 shows that the ROI per 100 candidates varies considerably, depending on the nature of the candidate pool, the organization's hiring practices, and the average cost for each unsuccessful employee. Across all scenarios, however, the ROI is considerable and suggests that administration of the Performance Assessment could benefit many types of organizations in many different ways.

**Table 1**  
**Return on Investment Approximations under Various Scenarios**

<i>Selection %</i>	<i>Candidate Success %</i>	<i>Selected Success %</i>	<i>Cost per failure</i>	<i>ROI per 100 candidates</i>
10%	25%	54%	\$5,000	<b>\$13,000</b>
			\$10,000	<b>\$27,500</b>
	50%	80%	\$5,000	<b>\$13,500</b>
			\$10,000	<b>\$28,500</b>
25%	25%	45%	\$5,000	<b>\$23,500</b>
			\$10,000	<b>\$48,500</b>
	50%	72%	\$5,000	<b>\$26,000</b>
			\$10,000	<b>\$53,500</b>
50%	25%	36%	\$5,000	<b>\$26,000</b>
			\$10,000	<b>\$53,500</b>
	50%	64%	\$5,000	<b>\$33,500</b>
			\$10,000	<b>\$68,500</b>

**Notes:** *Selection %* = the percentage of the candidate pool selected for hire, *Candidate Success %* = the percentage of the candidate pool that would be successful if hired, *Selected Success %* = the percentage of the selected candidate pool that will be successful (see NOTES, below, for derivation), *Cost per failure* = the average cost for each unsuccessful employee relative to each successful employee, *ROI per 100 candidates* = the average return on investment for the selection procedure assuming a \$15 fee per candidate.

**NOTES TO TABLE 1, ABOVE**

To derive the percentage of the selected candidate pool that will be successful (*Selected Success %*, Table 1), we made the following assumptions:

- 1) The candidates' Performance Assessment scores ( $X$ ) follow a standard normal distribution.
- 2) The candidates' level of success, if hired, ( $Y$ ) follow a standard normal distribution.
- 3) The correlation of Performance Assessment scores ( $X$ ) and level of success ( $Y$ ) is .434. This is the correlation of overall Performance and overall supervisor rating observed in the national field study.

We then estimated the percentage of the selected candidate pool that will be successful as a function of  $x_c$  (the proportion of the candidate pool selected for hire) and  $y_c$  (the proportion of the candidate pool that would be successful if hired) using laws of probability. The estimated percentage is obtained by integrating over the probability density function of the bivariate normal distribution as follows:

$$\Pr(Y > z_{y_c} | X > z_{x_c}) = \frac{\Pr(Y > z_{y_c}, X > z_{x_c})}{\Pr(X > z_{x_c})}$$

$$= \frac{\int_{z_{x_c}}^{\infty} \int_{z_{y_c}}^{\infty} (2\pi\sqrt{1-.434^2})^{-1} \exp\left[-\frac{1}{2} \left( \frac{1}{1-.434^2} \right) (x^2 - 2[.434]xy + y^2)\right] dy dx}{x_c}$$

In this expression,  $z_{x_c}$  represents the 100(1- $x_c$ ) percentile of the standard normal distribution.

### 3

## Construction and Evaluation of the WorkKeys Performance Assessment

Employers must consider several issues prior to selecting an appropriate integrity test. They must weigh evidence that supports the *validity* of the test—whether the test measures what it claims to measure, and whether the test can be influenced by answers that are not necessarily true. Employers must also be informed about the *reliability* of the test, which reflects evidence about the consistency of test scores. Further, employers are frequently concerned about the examinees' *reaction* to integrity assessments. Finally, employers must consider the *fairness* of their overall hiring practices to understand the impact of an integrity test. These factors are explained in more detail below.

The factors guiding the construction of the Performance Assessment, and the extent to which it follows existing testing guidelines, are presented in this section. The six main areas to consider related to the construction, evaluation, and use of the Performance Assessment for selection are:

1. Reliability
2. Validity
3. Fakeability
4. Applicant Reaction
5. Adverse Impact
6. Compliance with Guidelines

### Reliability

The reliability of a test reflects the stability of test results over time and across diverse settings. Thus, employers should select a test that yields consistent results for each individual, indicating the test is dependable. Essentially, reliability refers to the consistency of test results. Reliability is measured in two ways:

- Internal Consistency is the most popular measure of reliability and refers to how well items measuring the same concept relate with each other.
- Temporal Stability, also known as *test-retest* reliability, assesses whether results and responses on items from a test are consistent over time.

Research has reported moderate to high internal consistency (mean coefficient alpha = .81, *SD* = 0.10) and stability (mean test-retest = .85, *SD* = 0.38; mean time interval 111 days) across overt and covert integrity tests (Coyne & Bartram, 2002; Wanek, 1999). In a field study conducted by ACT, the WorkKeys Performance Assessment scales demonstrated good to excellent internal consistency reliabilities across multiple samples (range = .79 to .84). Further, the Performance Index, which is a combination of both scales, demonstrated excellent internal consistency reliability (range = .89 to .90) (More details on the reliability estimates of the Performance Assessment are provided in the Appendix.).

## Validity

Validity refers to the ability of a test to measure what it is intended to measure. Meta-analyses of industrial/organizational psychology literature have repeatedly documented the validity of integrity tests for predicting overall job performance, counterproductive work behaviors, and work safety/risk-taking behaviors [range of correlation = .18 to .57; validity estimates (after correcting for measurement error and range restriction) = .26 to .77; mean validity = .41] (Coyne & Bartram, 2002; Ones, Viswesvaran, & Schmidt, 1993). The capability of integrity tests to predict workplace and counterproductive behaviors has been demonstrated across a variety of occupations, work settings, and employee status.

Integrity tests explain (or predict) a significant portion of an average person's work behavior (range = 10% to 20%) (Furnham, 2001). Other important factors, which explain a large portion of work behavior, include the variety of work settings, differences between job applicants and incumbent employees, and differences between overt and covert types of integrity tests (Coyne & Bartram, 2002).

Adding integrity tests to a selection system that already includes cognitive assessments, such as WorkKeys Foundational Skills tests, can significantly improve the ability to predict job success. This is because integrity tests appear to measure aspects of job behavior that are different from those measured by cognitive ability. Additionally, research supports the notion that integrity tests provide the most incremental validity gains (27%) over cognitive ability. In contrast, other selection tools provide smaller gains: structured interviews (24%), work samples (24%), reference checks (12%), and biodata (4%) (Schmidt & Hunter, 1998). Using both kinds of measures—WorkKeys Foundational Skills and the Performance Assessment—provides a better indication of whether the candidate is likely to be successful on the job.

Although available integrity tests purportedly tap as many as 23 thematic constructs—such as admissions of theft, impulse control, attitudes toward supervisors, and orderliness—research has shown that broad, general scales are as predictive of job performance and counterproductive work behaviors as more targeted subscales (Murphy, 2000; Wanek, Sackett, & Ones, 2003). In addition, subscales measuring work safety/risk-taking behavior contributed independent and unique information that is different from the job performance or counterproductive work behaviors domain (Schmidt, Thoresen, Le, Ilies, & Holland, 2001). Thus, the Performance Assessment is based on a Risk Reduction scale, which captures regard toward safety procedures and risk-taking, as well as a General Work Attitudes scale, which captures general work behaviors.

The capability of the WorkKeys Performance Assessment to predict workplace and counterproductive behaviors has been demonstrated across a variety of occupations, work settings, and examinee characteristics. For example, in a field study conducted by ACT that sampled incumbents across a range of industries and occupations, observed (uncorrected) validity estimates ranged from .16 to .30 (median = .27) when the Performance Index was correlated with supervisor ratings of job performance. These ratings included task performance, prosocial/organizational citizenship behaviors, counterproductive behaviors, and safety/risk-taking behaviors.

*(continued on page 13)*

**Table 2**  
*Validity Estimates for the Development Sample of the Performance Assessment*

Job Performance Criteria	Performance Index				General Work Attitudes				Risk Reduction				
	Operational Validity				Operational Validity				Operational Validity				
	Obs	r	cME	IRR	Obs	r	cME	IRR	Obs	r	cME	IRR	
Task	.16		.21	.26	.27	.22	.28	.34	.36	.08	.10	.12	.13
Prosocial	.29		.40	.47	.49	.31	.43	.50	.53	.22	.30	.36	.38
Counter	.27		.37	.43	.46	.22	.31	.37	.39	.26	.36	.43	.45
Safety	.24		.33	.39	.41	.18	.25	.30	.32	.26	.35	.41	.43
Task, Prosocial, & Counter*	.26		.36	.42	.44	.31	.43	.50	.52	.16	.22	.27	.28
Counter & Safety*	.30		.41	.48	.50	.24	.33	.39	.41	.30	.41	.48	.51
All Supervisor Ratings*	.27		.38	.44	.47	.31	.44	.51	.53	.18	.25	.30	.31

*Note.* N = 484, Obs r = observed (uncorrected) correlation, cME = corrected only for measurement error in the criterion measure (i.e., supervisor unreliability), DRR = cME further corrected for direct range restriction, IRR = cME further corrected for indirect range restriction. Correlations  $\geq .10$  are significant ( $p \leq .05$ ).

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

After correction for measurement error and range restriction, the validity estimates (a.k.a. operational validities) for the Performance Index ranged from .27 to .50 (median = .46) (Refer to Table 2, page 12). These operational validities are comparable to those reported in the literature, where the mean operational predictive validity of integrity tests has been noted as .41 (Ones et al., 1993). (More details on the validity estimates of the Performance Assessment are provided in the Appendix.).

### **Fakeability**

Examinees may realize that the Performance Assessment is an assessment of their integrity. Consequently, there may be individuals who want to present themselves in the best possible light. Some individuals might feel pressured to respond in ways they deem more socially desirable than their true inclinations. Research has shown that, in general, individuals who respond in more socially desirable ways—which may or may not reflect their true attitudes—do not affect the validity of integrity tests. The ability of the test to explain or predict individual behavior in work settings is unaffected by such answers (Ones et al., 1993; Ones & Viswesvaran, 1998a; Schmidt et al., 2001). Similarly, socially desirable answers do not impact the predictive power of the subscales or the overall validity of the Performance Assessment. In addition, the Performance Assessment is able to identify people who might try to manipulate the test results by responding inconsistently to the items (see Chapter 5 for more details).

### **Applicant Reactions**

Employers are frequently concerned about applicant and employee perceptions of hiring and firing decisions and, by extension, the processes that contribute to those decisions. Research from opinion surveys of job applicants has shown that the majority of respondents perceived integrity testing as an appropriate selection procedure. When asked to rank order their overall impression of various selection procedures from positive to negative, integrity tests consistently ranked in the middle, below interviews and above ability or cognitive testing (Coyne & Bartram, 2002).

### **Adverse Impact**

Adverse impact results when there is unfair discrimination against members of protected classes regardless of an employer's intent. One source of evidence for adverse impact is when members of a protected class are selected at rates that are less than four-fifths (80%) of the group with the highest selection rate. Research on integrity tests like the Performance Assessment has shown that these tests do not result in adverse impact. Evidence shows only small to insignificant differences in the results between demographic groups. For instance, women tend to score slightly higher than men ( $SD = 0.16$  standard deviation units), older applicants tend to score slightly higher than younger applicants ( $SD = 0.08$  standard deviation units), and the differences between Caucasian and other ethnic groups (including American Indian, Asian, African American, and Hispanic) has been described as negligible (Coyne & Bartram, 2002; Ones & Viswesvaran, 1998b). In contrast, cognitive ability tests show larger differences between some demographic groups, and their use can pose significant problems for employers when used for selection purposes without job analysis. As a result, the use of cognitive tests in some settings has been challenged under federal antidiscrimination laws (Sackett, 1994).

Results from the Performance Assessment have not shown significant differences between groups of individuals—including sex, age, or race/ethnic groups—who have completed the assessment (Please refer to the Appendix for more details). This means the test is unlikely to yield lower scores for any racial/ethnic groups or a particular gender or age group. With the scientific literature finding no meaningful differences between groups, integrity tests tend not to be challenged under federal antidiscrimination laws (Coyne & Bartram, 2002).

## **Compliance with Guidelines and Standards**

The Performance Assessment is in compliance with the test development guidelines recommended by the International Testing Commission (2006), the Association of Test Publishers (2002), Society for Industrial and Organizational Psychology (2003), and the guidelines recommended by the Joint Committee on Standards for Educational and Psychological Testing (consisting of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education) (1999). These standards address “criteria for the evaluation of tests, testing practices, and the effects of test use” (p. 2) including delivery formats, administration and hardware/software requirements, and the documentation of the validity and reliability of a test (Joint Committee on Standards for Education and Psychological Testing, 1999).

The Equal Employment Opportunity Commission (EEOC) provides detailed guidelines for employment testing (EEOC, 1978). Along with many other recommendations, the EEOC advises that tests showing adverse impact should generally be avoided. However, the business necessity of a test should be demonstrated if a test does show adverse impact against any demographic groups. The Performance Assessment has been designed to meet EEOC standards. It has also been found to bear no undue adverse impact on any racial/ethnic or gender groups. Employers can use this information to assist them in adopting lawful and appropriate hiring practices and to avoid legal challenges to their screening and hiring practices.

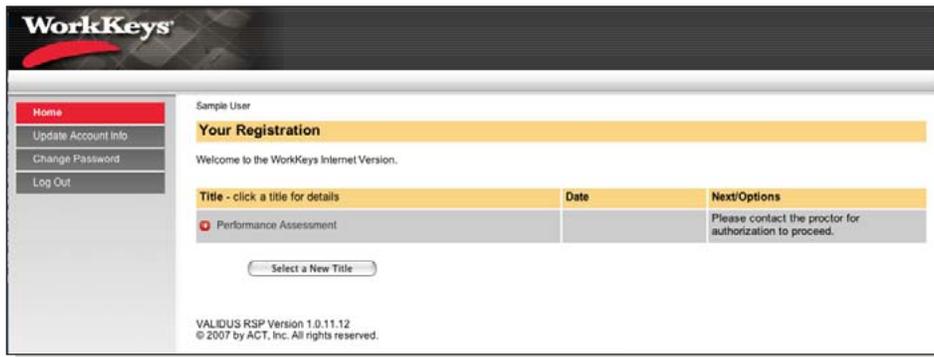
# 4 Administration of the WorkKeys Performance Assessment

## Administration of the WorkKeys Performance Assessment

The Performance Assessment is highly efficient and practical in terms of cost, length of time for test administration, type of equipment needed, and test user training. The assessment is administered entirely online through a Web-based platform, which saves costs normally associated with administering, scoring, and reporting results from traditional paper-and pencil tests. As a result, the Performance Assessment requires only basic computing skills and Internet hardware and software. The use of computers also creates an easy-to-use and cost-efficient account management system for employers. This section introduces the basic testing environment requirements, the instructions for the examinee, and the confidentiality agreement.

A screen shot of the assessment Web-delivered survey platform is shown below in Figure 1. Other documents, including frequently asked questions (FAQs) and test administration guides, are available at [www.act.org/workkeys/assess/personal.html](http://www.act.org/workkeys/assess/personal.html).

**Figure 1**  
*Welcome Screen for the Performance Assessment*



A comprehensive WorkKeys Internet Version User Guide with instructions for test administrators is at [www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf](http://www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf). This document details the steps required for setting up examinees in the online environment and managing company examinees for the entire WorkKeys line.

## Testing Environment

Remind users to turn off pagers, cell phones, and wristwatch alarms to avoid distracting other users. All testing staff, room supervisors, and proctors are to *remain attentive* to their testing responsibilities throughout the entire administration. To protect the validity of individual test scores and maintain the security of the test materials, testing staff should:

- Walk around the room during testing to ensure users are working on the correct assessment and to prevent prohibited behaviors.
- Not read or engage in any tasks not related to the administration of the assessment.
- Not engage in conversation during the assessment or allow unauthorized personnel into the testing room.
- Not leave the testing room unattended at any time.

More information about the testing environment and administrator guides is available in the WorkKeys Internet Version User Guide at: [www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf](http://www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf)

## Examinee Setup in the Validus™ Virtual Test Center

The administrator will need to register the examinee in the Validus™ Virtual Test Center prior to the examinee beginning the test. Please refer to the WorkKeys Internet Version User Guide for instructions and further reference materials: [www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf](http://www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf)

## Confidentiality Agreement

All examinees must agree to the confidentiality agreement prior to starting the Performance Assessment (see Figure 2). The agreement states the confidential nature of the contents of the test. Examinees who do not agree to the confidentiality agreement will not be able to take the assessment. However, the assessment will be counted as used even if an examinee does not agree to the confidentiality statement.

**Figure 2**  
**Confidentiality Agreement**



The screenshot shows a web-based form titled "Confidentiality Agreement" from WorkKeys. The form is divided into two main sections: "Performance Assessment" and "Confidentiality Agreement". The "Performance Assessment" section contains the following text: "I understand that the content of this examination is confidential and is protected by copyright, trademark, and intellectual property rights. No part of the materials conveyed during this examination may be copied, downloaded, reproduced, photographed, stored, distributed, used, or transferred to any individual or group." Below this, it states: "I also certify that I am the person authorized to take this assessment." The "Confidentiality Agreement" section contains two radio button options: "Agree: I have read the above statements and AGREE to the terms." and "Disagree: I have read the above statements and DISAGREE with the terms." A note below the options states: "Note: Selecting 'Disagree' will cause you to exit this website and you will not be able to complete the assessment." A footer note reads: "\*\*More information about ACT's use of data can be found in Policies and Guidelines for Uses of Data from ACT-Administered Assessment. ©2007 by ACT, Inc. All rights reserved." A "Submit" button is located at the bottom right of the form. The WorkKeys logo is visible in the top left corner of the page. The text "VALIDUS SURVEY Version 1.0.4.34" is visible at the bottom left of the page.

## Test Instructions for Examinee

Examinees will be required to provide demographic information and the occupation for which the examinee is applying (see figure 3). Examinees are allowed to select up to five occupations, with the first one being required and serving as the primary occupation. A lookup table function to locate occupational titles is built into this portion of the assessment. The associated O\*NET codes (occupation codes based on the 10th version of the O\*NET) appear for human resources purposes, and also appear at the end of the score reports generated for the Performance Assessment.

The examinee will receive the instructions shown in Figure 4 prior to taking the assessment.

**Figure 3**  
*Demographics*

The screenshot shows the 'Demographics' section of the WorkKeys assessment. On the left is a navigation menu with 'Demographics' selected. The main content area is titled 'Demographics' and contains several dropdown menus for 'Gender', 'Ethnicity', 'What language do you know best?', and 'What is your current employment status?'. Below these is a section for 'What is your current occupation?' with instructions and three dropdown menus for 'Select a category', 'Select a subcategory', and 'Select an occupation title'. Each dropdown menu has a question mark icon for help.

**Figure 4**  
*Instructions for the Performance Assessment*

The screenshot shows the 'Assessment' section of the WorkKeys assessment. On the left is a navigation menu with 'Assessment' selected. The main content area is titled 'Assessment' and contains instructions for the performance assessment. It includes a rating scale from 'Strongly Disagree' to 'Strongly Agree' and a list of instructions for the examinee. The instructions include: 'To select the appropriate answer, click on the circle that corresponds to your level of agreement. You can only select one answer per question.', 'Do not spend too much time deciding on any one answer. This assessment should take about 10 to 15 minutes to complete.', 'Do not worry if some items seem similar. Simply answer each item.', 'If you do not respond to all items, your assessment may not be scored.', and 'The sidebar can be used to navigate through the assessment, and you will have a chance to review your responses before you submit your answers.' A note at the bottom states: 'Please note that once you exit the assessment, you may not return to the questions.'

## **Accommodations for Examinees for Whom English is a Second Language**

Examinees for whom English is a second language may bring and use a foreign language dictionary. The test administrator must check the dictionary, before and after testing, to ensure that it does not contain any of the test items or responses to test items.

## **Reporting the Results of the WorkKeys Performance Assessment**

Scoring and reporting for the Performance Assessment is immediate. After the examinee has completed the assessment, the system creates a PDF document and stores it electronically for access by the client/employer for up to a year after the test session. Two different reports are generated. An explanation of the different components of these reports can be found in Chapter 5.

- The Employer Report provides details on each individual's performance profile, including a Performance Index, and scores on the General Work Attitudes and Risk Reduction subscales. The overall performance rating, as indicated by the Performance Index, is the general score that classifies the candidate as highly, moderately, or less desirable based on the predicted level of performance.
- The List Report catalogs all applicants who have been assessed during a given time period, the identification number associated with each person, the dates they were tested, and their performance, including overall Performance Index and subscale scores. The final column on a List Report reflects the recommended desirability level (i.e., highly, moderately, or less desirable), based on the Performance Index of each candidate.

## 5

# Interpreting Employer and List Reports

ACT recommends that decisions involving the Performance Assessment be based on the Performance Index—a composite score using the two subscales of the assessment—because it provides more information than either of the subscales alone. Further, ACT recommends that decision making based on the Performance Index proceed in a top-down or rank-ordered strategy with selection starting from the highest-scoring candidates. Employers will find that making decisions about who to hire and how far down the desirability scale the employer goes in making hiring decisions will be determined, in part, by the labor supply factors of their local markets.

Employers may decide to use a cut-score approach. They can set a “passing score” that indicates the minimal score they will accept from a job candidate. Only candidates with scores at or above the cut-score will continue in the selection process. ACT has provided cutoffs that include descriptions of high, moderate, and low desirability levels along a normal distribution of scores. (Refer to the Appendix for information on how desirability levels were determined.) ACT provides these cutoffs as guidelines for employers to use in their selection process. Such cutoffs will be most defensible when the organization can provide evidence that the score used to sort acceptable from unacceptable job candidates is based on job requirements as determined through a job analysis.

Some organizations may be interested in establishing their own cutoffs for the Performance Index alone or for the individual subscales, depending on their needs. For example, manufacturing or construction businesses might consider safety to be critical to their work and want to ensure that people recommended by the overall Performance Index meet a minimum safety threshold. ACT can generate initial cutoffs for both the General Work Attitudes and Risk Reduction subscales that companies can adopt for this purpose. ACT can also work with individual organizations to customize cutoffs to meet their specific workplace needs. This strategy will be most defensible when the scores are based on critical job requirements documented through a job analysis which supports the chosen cutoff scores and places priority on a particular subscale score.

## Employer Reports

This section provides sample Employer and List Reports that illustrate the range of scores an examinee may receive and what those scores mean. Employer Reports should reflect the appropriate test (upper left-hand corner). The following identifiers will appear at the top of the page:

- *Report for:* Your company name
- *Site:* Your company location (displayed if a company has multiple sites)
- *Test Date:* Date the particular test was completed
- *Examinee:* The name of the test taker
- *Examinee ID:* Last four digits of the unique identifier for each examinee

Following the identifying information, the WorkKeys Performance Assessment Profile box reflects bar graphs that show where the individual ranks relative to other test takers from the normative group. The bar chart provides a visual

representation of scores and demonstrates how individuals performed on the test compared to other examinees. The following are reflected in the bar graph:

- Performance: This is the Performance Index, a percentile score which shows how an individual compares to others who have taken the assessment. The Performance Index is based on the combination of scores generated by the General Work Attitudes and Risk Reduction subscales and ranges from 1 to 99.
- General Work Attitudes: This reflects the score of the individual on the General Work Attitudes subscale.
- Risk Reduction: This reflects the score of the individual on the Risk Reduction subscale.

Following this Performance Assessment Profile box, a section titled *What This Means* includes a score, which reflects the Performance Index, and a description of how the individual score can be interpreted. The level of desirability for the examinee is reported and explained in this section as well. The Employer Report details the performance of an individual and can represent three different levels of desirability—high, moderate, or low. The Performance Index is also color-coded to reflect the following levels of desirability:

Green – Highly desirable level of expected performance

Yellow – Moderately desirable level of expected performance

Red – Less desirable level of expected performance

The General Work Attitudes and Risk Reduction scores are also reflected with score-level descriptions of each of the subscales. The next section provides a more detailed description of different score ranges.

The final section of the Employer Report shows the occupations specified by the examinee. These are occupations the examinee selected as an area of interest or the specific positions for which they are applying. At the end of this section is the website employers can visit to view more information about the assessment.

## Sample Employer Reports

The following examples and descriptions are designed to help with the interpretation of Performance Employer and List reports.

There are three different levels of desirability, which are determined by the Performance Index. Table 3 reflects these levels and the score ranges that fall into each category. However, for additional assistance, please feel free to contact ACT at 1-800/WORKKEY (1-800/967-5539).

**Table 3**  
*Levels of Desirability and Associated Score Ranges*

---

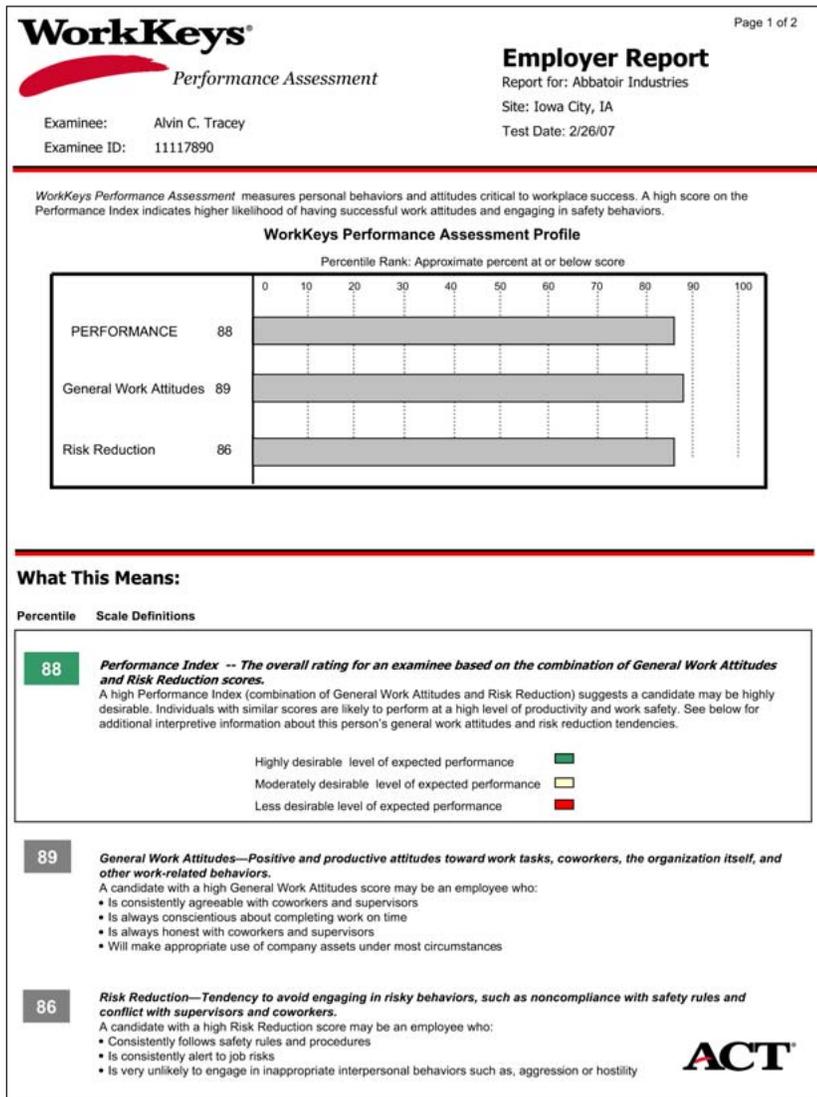
Desirability Level	Performance Index Score Range
High	76–99
Moderate	16–75
Low	1–15

---

## High Desirability

A high Performance Index (the combination of Risk Reduction and General Work Attitude scores) suggests a candidate may be **highly desirable**. Individuals with the highest scores in this category will perform at higher levels of productivity and work safety when compared to candidates whose scores are lower. For example, a score in the 88th percentile, as reflected by the highlighted number in Figure 5, indicates that the performance for this individual is the same as or higher than 88% of other individuals who took the test. Specific descriptive language for each of the subscales provides guidelines on the desirability of a particular candidate based on his/her Performance Index.

**Figure 5**  
*Employer Report Reflecting High Scores*



(Figure 5, continued)

WorkKeys® Performance Assessment

Employer Report

Page 2 of 2

Report for: Abbatoir Industries  
Site: Iowa City, IA  
Test Date: 2/26/07

Examinee: Alvin C. Tracey  
Examinee ID: 11117890

Examinee-specified primary occupation is in BOLD.

Code	Title
11-3071.02	<b>Storage and Distribution Managers</b>
11-3042.00	Training and Development Managers
11-3061.00	Purchasing Managers
11-3049.99	Human Resources Managers, All Other
11-3071.01	Transportation Managers

For more information go to <http://www.act.org/workkeys/assess/performance>

*General Work Attitudes:*

A candidate who demonstrates high performance on the General Work Attitudes portion of the assessment can be expected to exhibit positive work behaviors on the job. A score in the 89th percentile as shown in Figure 5 means that this individual scored the same as or higher than 89% of test takers at this subscale. Specifically, a candidate with a high score on the General Work Attitude subscale may have the following tendencies as an employee:

- Is consistently agreeable with coworkers and supervisors
- Is always conscientious about completing work on time
- Is always honest with coworkers and supervisors
- Will make appropriate use of company assets under most circumstances

*Risk Reduction:*

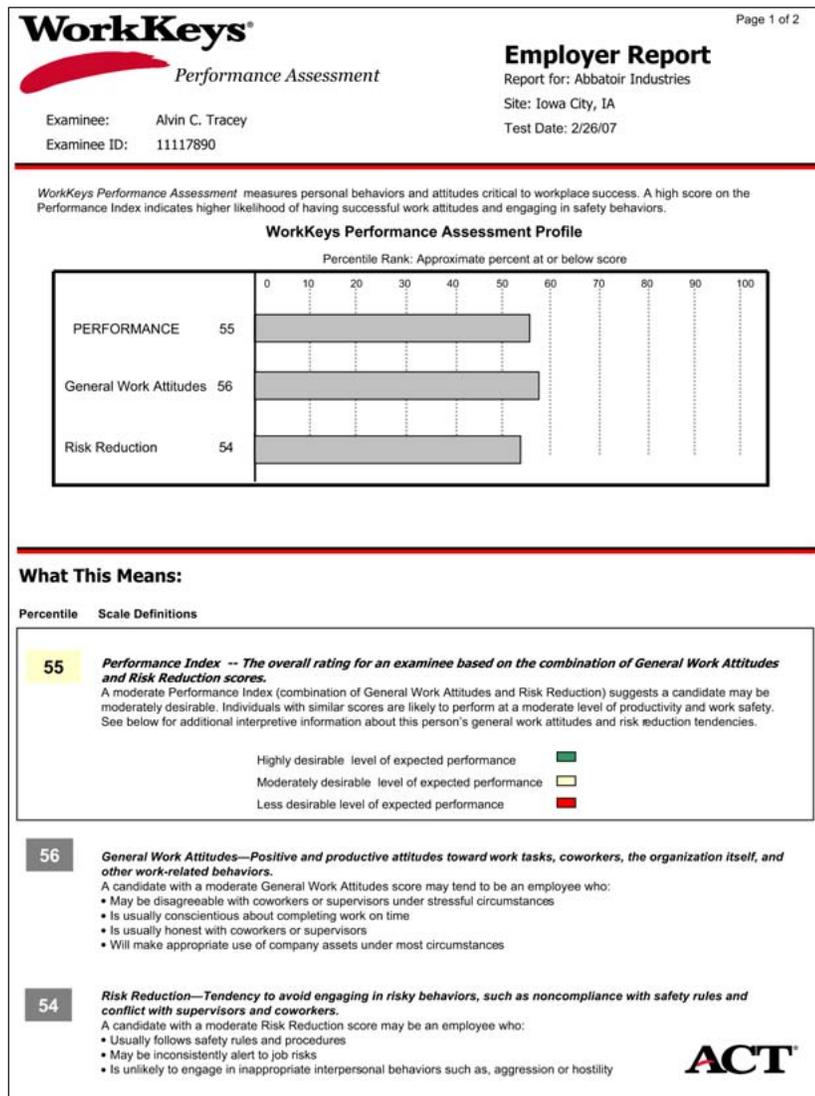
A candidate who performs very well on the Risk Reduction portion of the Performance Assessment suggests the candidate may be **highly desirable**. A Risk Reduction score in the 86th percentile as shown in Figure 5 means that this individual scored the same as or higher than 86% of the test takers at this subscale. Individuals with similar scores are likely to perform at higher levels of work safety. Score report language describes individuals with high Safety scores as employees who may have the following tendencies:

- Consistently follows safety rules and procedures
- Is consistently alert to job risks
- Is very unlikely to engage in aggressive, hostile, or other unsafe behaviors

## Moderate Desirability

A moderate Performance Index (the combination of General Work Attitude and Risk Reduction scores) suggests a candidate may be **moderately desirable**. Individuals with similar scores are likely to perform at reasonable levels of productivity and work safety in comparison to others. Selection of candidates in this score range is encouraged with recognition that they are good candidates with a few limitations. For example, a score in the 55th percentile as is shown in Figure 6 indicates that the performance for this individual is the same as or higher than 55% of other individuals who completed the assessment. Specific descriptive language for each of the subscales provides guidelines on the desirability of a particular candidate based on their performance.

**Figure 6**  
*Employer Report Reflecting Moderate Scores*



(Figure 6, continued)

WorkKeys®  
Performance Assessment

Employer Report

Page 2 of 2

Report for: Abbotair Industries  
Site: Iowa City, IA  
Test Date: 2/26/07

Examinee: Alvin C. Tracey  
Examinee ID: 11117890

Examinee-specified primary occupation is in BOLD

**Occupations Specified by Examinee:**

Code	Title
11-3071.02	<b>Storage and Distribution Managers</b>
11-3042.00	Training and Development Managers
11-3061.00	Purchasing Managers
11-3049.99	Human Resources Managers, All Other
11-3071.01	Transportation Managers

For more information go to <http://www.act.org/workkeys/assess/performance>

*General Work Attitudes:*

A candidate who demonstrates moderate performance on the General Work Attitudes portion of the assessment can be expected to exhibit reasonable work behaviors while on the job. A General Work Attitudes score in the 56th percentile as is shown in Figure 6 indicates that this individual scored the same as or higher than 56% of test takers at this subscale. Specifically, a candidate with a moderate score on the General Work Attitudes subscale may have the following tendencies as an employee:

- May be disagreeable with coworkers or supervisors under stressful circumstances
- Is usually conscientious about completing work on time
- Is usually honest with coworkers or supervisors
- Will make appropriate use of company assets under most circumstances

*Risk Reduction:*

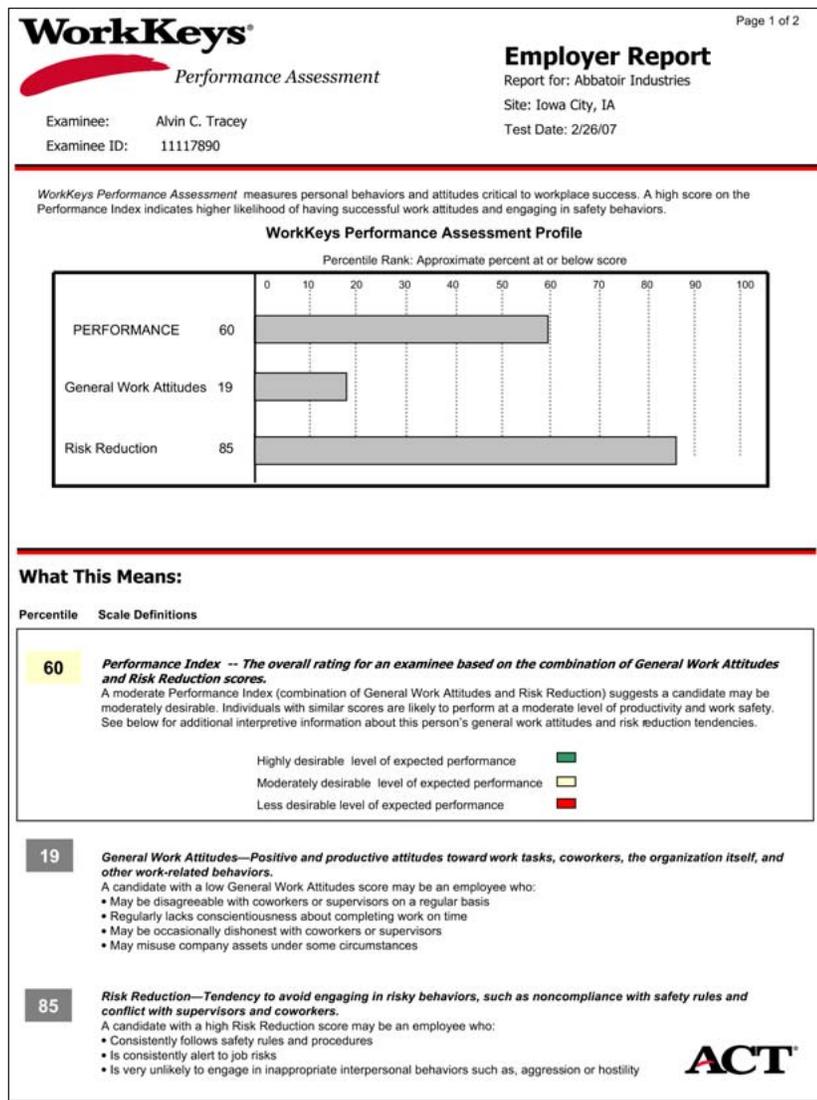
A candidate who performs moderately well on the Risk Reduction portion of the assessment can be expected to adhere to safety rules and procedures most of the time. A Risk Reduction score in the 54th percentile as is shown in Figure 6 means that this individual scored at or above 54% of other test takers at this subscale. Score report language describes individuals with moderate Risk Reduction scores as employees who may have the following tendencies:

- Usually follows safety rules and procedures
- May be inconsistently alert to job risks
- Is unlikely to engage in aggressive, hostile, or other unsafe behaviors

### Moderate Performance Index with Uneven Subscale Scores

Although a high Performance Index results from high subscale scores, and a low Performance Index is a result of low subscale scores, a moderate Performance Index may be the result of either (a) moderate subscale scores or (b) a high score on one subscale and a low score on the other subscale. If a dramatic difference between subscale scores is present (i.e., a difference greater than or equal to 41 percentile points), an employer may need to consider whether the low subscale score is acceptable for the job in question. For example, if job safety is paramount, a low score on the Risk Reduction subscale may not be acceptable to an employer.

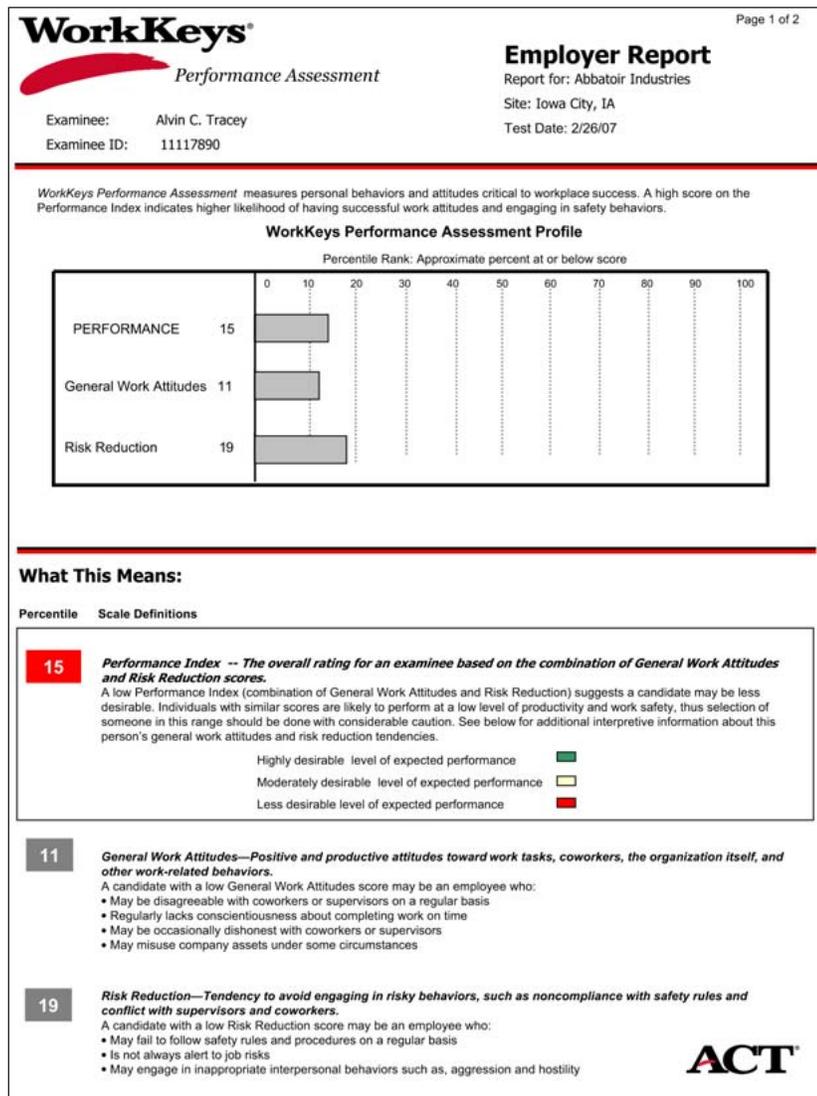
**Figure 7**  
*Employer Report Reflecting Uneven Subscale Scores*



## Low Desirability

A low Performance Index (the combination of Risk Reduction and General Work Attitude scores) suggests a candidate may be **less desirable**. Individuals with similar scores are likely to perform at a low level of productivity and work safety. Selection of an applicant in this performance range should be done with considerable caution and may be determined by labor supply factors in local markets. For example, a score in the 15th percentile as is shown in Figure 8 indicates that the performance score for this individual is the same as or higher than 15% of other individuals who took the test. Specific descriptive language for each of the subscales provides guidelines on the desirability of a particular candidate based on his/her performance.

**Figure 8**  
*Employer Report Reflecting Low Scores*



	<b>Employer Report</b>	<small>Page 2 of 2</small>
	Report for: Abbattoir Industries Site: Iowa City, IA Test Date: 2/26/07	
Examinee: Alvin C. Tracey Examinee ID: 11117890		

---

Examinee-specified primary occupation is in BOLD.

Code	Title
<b>11-3071.02</b>	<b>Storage and Distribution Managers</b>
11-3042.00	Training and Development Managers
11-3061.00	Purchasing Managers
11-3049.99	Human Resources Managers, All Other
11-3071.01	Transportation Managers

For more information go to <http://www.act.org/workkeys/assess/performance>

*General Work Attitudes:*

A candidate who performs poorly on the General Work Attitudes portion of the assessment may exhibit poor work behaviors while on the job. A General Work Attitudes score in the 11th percentile, as shown in Figure 8, indicates that this individual scored the same as or higher than 11% of test takers at this subscale. Specifically, a candidate with a low General Work Attitudes score may have the following tendencies as an employee:

- May be disagreeable with coworkers or supervisors on a regular basis
- Regularly lacks conscientiousness about completing work on time
- May be occasionally dishonest with coworkers or supervisors
- May misuse company assets under some circumstances

*Risk Reduction:*

A candidate who performs poorly on the Risk Reduction portion of the assessment may not recognize or adhere to safety rules and procedures. A Risk Reduction score in the 19th percentile as is indicated in Figure 8 means that this individual scored the same as or higher than 19% of test takers at this subscale. Score report language describes individuals with low Risk Reduction scores as employees who may have the following tendencies:

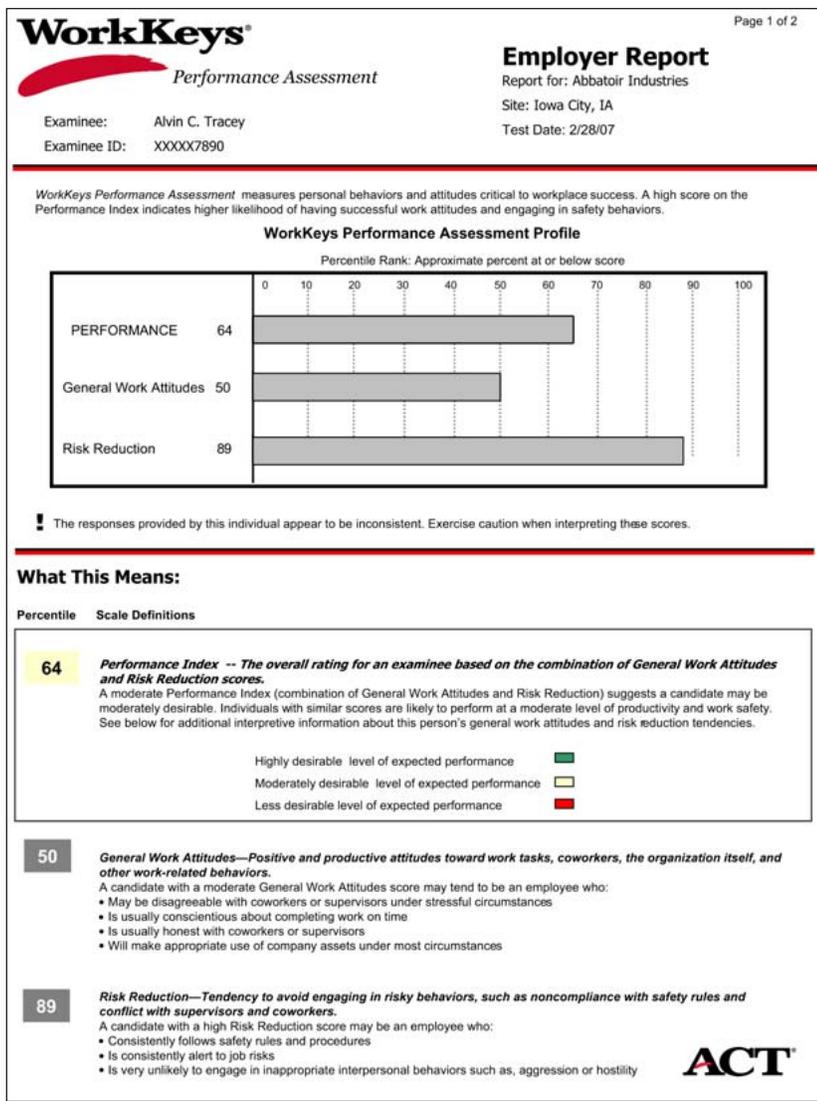
- May fail to follow safety rules and procedures on a regular basis
- Is not always alert to job risks
- Possibly engages in aggressive, hostile, or other unsafe behaviors in stressful situations

## Other Examples of Employer Reports

### Inconsistent Responding

Some individuals may respond to the items without careful consideration of the content, without reading them, or in a deliberate attempt to manipulate the results of the test. The responses generated may be inconsistent as a result. For instance, because some of the items are evaluated in such a way that low scores reflect positive answers (i.e., reverse-keyed), an individual who responds by selecting the same response (e.g., “strongly disagree”) to a large portion of items will trigger a warning. In the event that an individual responds to the items regardless of content, the score report will flag the Employer and List Reports with an exclamation mark. The report will also include a message warning that the responses require caution during interpretation by the employer.

**Figure 9**  
**Employer Report with Inconsistent Reporting Flag**



## Sample List Reports

The WorkKeys system produces a List Report for all candidates who have taken the WorkKeys Performance Assessment for the company during a specified period of time. The List Report includes identifier details for the company:

- *Site*: Name of the company
- *Report Date*: When the report was generated
- *Date Range*: Specified date range of assessment administration
- *Occupation Code*: Specific position codes within the company
- *Sort By*: Allows sorting by percentile score and examinee last name

**Figure 10**  
*List Report Sorted by Performance Index Recommendation Category*

Examinee	Examinee ID	Primary Occupation Code	Date Tested	Percentiles			Recommended Category Based on Performance Index
				General Work Attitudes (1-99%ile)	Risk Reduction (1-99%ile)	PERFORMANCE INDEX (1-99%ile)	
Smith, Bob	XXXXX9112	11-3042.00	11/14/2006	92	95	93	Highly desirable
Carter, Linda	XXXXX0733	11-3061.00	11/14/2006	83	88	86	Highly desirable
White, Tashon	XXXXX0944	11-3071.02	11/14/2006	83	88	86	Highly desirable
King, Billie	XXXXX9922	11-3061.00	11/14/2006	81	86	84	Highly desirable
Duff, Patrick	XXXXX0077	11-3071.02	11/14/2006	74	80	77	Highly desirable
Castillo, Leo	XXXXX8788	11-3042.00	11/14/2006	78 †	68 †	70 †	Moderately desirable
Murphy, Edie	XXXXX3400	11-3061.00	11/14/2006	69	59	66	Moderately desirable
Pak, Ruth	XXXXX8899	11-3071.02	11/14/2006	59	69	66	Moderately desirable
Villanegro, Mari	XXXXX6711	11-3061.00	11/14/2006	60	68	66	Moderately desirable
Grover, Arnold	XXXXX5433	11-3071.02	11/14/2006	50	53	51	Moderately desirable
Reynolds, Bart	XXXXX8722	11-3042.00	11/14/2006	50	51	50	Moderately desirable
Tezzini, Frank	XXXXX5644	11-3061.00	11/14/2006	41	44	43	Moderately desirable
Clark, Lewis	XXXXX8688	11-3071.02	11/14/2006	38	46	42	Moderately desirable
Black, Lois	XXXXX2166	11-3042.00	11/14/2006	42	38	40	Moderately desirable
Starr, Sam	XXXXX4100	11-3042.00	11/14/2006	35	41	38	Moderately desirable
Jones, Fred	XXXXX2907	11-3061.00	11/14/2006	31	34	32	Moderately desirable
Lester, Kelsey	XXXXX3768	11-3071.02	11/14/2006	14	13	13	Less desirable

Note: † = Inconsistent responses

The List Report details the applicant's name, a unique identification number associated with the applicant, the Standard Occupational Code (SOC code), and the date the individual completed the assessment. The percentiles for Risk Reduction, General Work Attitudes, and the Performance Index are displayed in individual columns. The final column features a recommendation for how desirable a candidate may be based on the Performance Index. The featured List Report (see Figure 10) is sorted by the Performance Index, ranging from highly desirable to less desirable.

List Reports can be generated through filtering and sorting functions built into the system. This includes two sorting functions, allowing a list of candidates to be sorted by name or by Performance Index (see Figure 11). A company may use filters to narrow a pool of examinees based on specific criteria from the larger pool of test takers in the company's database. For instance, the employer can choose to filter by a specific occupation code or by the date a group of individuals completed the assessment. When

occupation code is selected as the filter criteria, up to five occupation codes listed per individual will be scanned for any appearance of that code, and individuals with the specified code will subsequently be featured in the List Report.

**Figure 11**  
*List Report Sorted by Name*

WorkKeys <sup>®</sup>		Performance Assessment		List Report			Page 1 of 1
				Report for: Abattoir Industries Site: Iowa City Report Date: Feb 14, 2007	Date Range: 01/01/2006 to 02/13/2007 Occupation Code: <All> Sort by: Alpha Order		
Examinee	Examinee ID	Primary Occupation Code	Date Tested	Percentiles			Recommended Category Based on Performance Index
				General Work Attitudes (1-99%ile)	Risk Reduction (1-99%ile)	PERFORMANCE INDEX (1-99%ile)	
Black, Lois	X00008688	11-3061.00	11/14/2006	42	38	40	Moderately desirable
Carter, Linda	X00009112	11-3042.00	11/14/2006	83	88	86	Highly desirable
Castillo, Leo	X00008788	11-3042.00	11/14/2006	78 1	68 1	70 1	Moderately desirable
Clark, Lewis	X00002166	11-3071.02	11/14/2006	38	46	42	Moderately desirable
Duff, Patrick	X00000077	11-3071.02	11/14/2006	74	80	77	Highly desirable
Grover, Arnold	X00005433	11-3042.00	11/14/2006	50	53	51	Moderately desirable
Jones, Fred	X00002907	11-3042.00	11/14/2006	31	34	32	Moderately desirable
King, Billie	X00009922	11-3061.02	11/14/2006	81	86	84	Highly desirable
Lester, Kelsey	X00003768	11-3071.02	11/14/2006	14	13	13	Less desirable
Murphy, Edie	X00003400	11-3071.01	11/14/2006	69	59	66	Moderately desirable
Pak, Ruth	X00008899	11-3061.00	11/14/2006	59	69	66	Moderately desirable
Reynolds, Bart	X00008722	11-3071.01	11/14/2006	50	51	50	Moderately desirable
Smith, Bob	X0000733	11-3061.00	11/14/2006	92	95	93	Highly desirable
Starr, Sam	X00004100	11-3071.02	11/14/2006	35	41	38	Moderately desirable
Tezzini, Frank	X00005644	11-3042.00	11/14/2006	41	44	43	Moderately desirable
Vilanegro, Mari	X00006711	11-3071.02	11/14/2006	60	68	66	Moderately desirable
White, Tashon	X00000944	11-3042.00	11/14/2006	83	88	86	Highly desirable

Note: 1 = Inconsistent responses

**ACT**

# Appendix

## Development of the WorkKeys Performance Assessment and Research Findings

This Appendix is designed to give interested readers a brief background on the development process and psychometric properties of the WorkKeys Performance Assessment. This section lists the steps of the process, offers information on the measure's properties, including reliability and validity, and provides information on other issues of interest, such as analyses of adverse impact. Although a careful review of this section is not necessary to understand the Performance Assessment, it is useful for those who would like a clearer understanding of the technical details of the instrument.

### Development of the Performance Assessment

A three-part process was used to develop the Performance Assessment: (1) preparation of the initial item pool, (2) empirical item selection procedures, including development of supervisor ratings as performance criteria, and (3) examination of scale reliability, validation, and other analyses.

### Preparation of the Initial Item Pool

Preparation of the initial item pool was based on the industrial and organizational psychology literature in which the validity of integrity tests for predicting overall job performance, counterproductive work behaviors, and work safety/risk-taking behaviors is well-documented (e.g., Coyne & Bartram, 2002; Ones, Viswesvaran, & Schmidt, 1993). Additional reviews of the literature on integrity and job performance led to the identification of several key constructs that serve to distinguish individuals in terms of expected job performance and safety/risk-taking behaviors. ACT researchers wrote comprehensive construct definitions and obtained feedback from experts in the fields of industrial/organizational psychology and personality psychology.

Following revisions and face validation, definitions were finalized and shared with item writers. A research team, comprised of four applied psychologists, wrote items representing the constructs. Writers generated items independently and then met to discuss the breadth of coverage and revisions. This procedure yielded an initial item pool of 274 items.

### *Readability Test*

To ensure that the items would be comprehensible to the average worker, items were administered to a sample of employees. Workers were asked to rate the extent to which they understood the meaning of the items using a 5-point, Likert-type scale ranging from *very easy to understand* to *very difficult to understand*. Based on the mean ratings of item clarity, items were deleted or revised. Subsequently, the revised items were presented to a second group of experts in workforce and communication who were asked to comment on item clarity. The items were again revised to reflect this feedback.

The resulting item pool consisted of 216 items, which were randomly ordered and set to a 6-point, Likert-type response scale ranging from *strongly disagree* to *strongly agree*. Administration instructions were developed, along with procedures to maintain the confidentiality of field test participants.

## Empirical Item Selection

To select items for the Performance Assessment, ACT researchers used a multistep procedure that included the following: (1) development of supervisor rating scales as performance criteria, (2) item selection to create two scales, and (3) creation of an overall index of performance based on the scales created in step two.

### *Development of Supervisor Rating Scales*

The first step toward selecting items for the Performance Assessment consisted of developing supervisor rating scales to use as performance criteria. These scales enabled the supervisors of incumbents participating in field studies to complete a set of performance ratings about their employees. To develop the supervisor ratings of employee performance, ACT researchers examined the relevant literature on performance criteria, such as task performance, prosocial/organizational citizenship behaviors, counterproductive behaviors, safety/risk-taking behaviors, as well as normative rating and general performance (Barrick & Mount, 1991; Borman, Penner, Allen, & Motowidlo, 2001; Rotundo & Sackett, 2002; Salgado, 2002). A total of 41 supervisor ratings were developed. A sample item from the prosocial/organizational citizenship behavior scale is featured below.

#### **Compliance towards organization/supervisor**

- Shows respect for people in positions of authority
- Is responsive to supervisory requests
- Has a good working relationship with supervisor
- Consistently follows policies and procedures
- Speaks to supervisors with respect

Never    Not Very Often    Sometimes    Often    Very Often    Always

Ratings from 1,082 supervisors who participated in this and other field tests were used to derive the final performance criteria scales. ACT researchers conducted both exploratory and confirmatory factor analyses. The supervisor sample was randomly split into two groups, with 70% of the sample in the “exploratory” group ( $n = 757$ ) and the remaining 30% in the “confirmatory” group ( $n = 325$ ). A factor analysis on the exploratory group resulted in four factors. Subsequently, a confirmatory factor analysis specifying four latent factors was run on the data from the “confirmatory” group using the maximum likelihood estimation method. The extent to which the model fit the data was examined by using the combination of several fit indexes (i.e., Comparative Fit Index, Root Mean Square Error of Approximation, and Standardized Root Mean Square Residuals). After completing the aforementioned factor analyses, 29 items were selected to comprise the four performance factors (scales). Table A1 features scale intercorrelations of supervisor rating scales. (Note: subsequent tables featuring supervisor criteria include the four supervisor scales as well as combinations of these scales). The associations illustrated in Table A1 are consistent with research on the structure of job performance ratings and

work behaviors (Rotundo & Sackett, 2002; Sackett, 2002), in which task and general job performance ratings are more strongly associated with prosocial/organizational citizenship behaviors than with counterproductive or safety/risk-taking behaviors.

**Table A1**  
*Intercorrelations of the Supervisor Performance Rating Scales*

Scale (# of items)	1	2	3	4
1. Task and Job Performance (12)	–			
2. Prosocial/Organizational Citizenship (6)	.74	–		
3. Counterproductive Behaviors (6)	.46	.54	–	
4. Safety/Risk-taking Behaviors (5)	.33	.35	.43	–

*Note.*  $N = 757$ .

### *Characteristics of Sample*

Participants represented nine organizations spanning different industries, including manufacturing, healthcare, education, information services, as well as testing and publishing. The size of participating organizations ranged from small businesses to branches of multinational companies. The average participating supervisor had been in his/her position for over two years and had been supervising his/her employee for an average of one to two years. The most common O\*NET occupation areas in the incumbent sample were: Production and Manufacturing (49.5%), Computer and Mathematics (14.8%), and Transportation and Material Moving (13.9%). Other occupation areas included: Education, Training, and Library; Food Preparation and Serving; Healthcare Support; Installation, Maintenance and Repair; and Office and Administrative Support. The average participating incumbent had occupied the same position for over two years.

Incumbents' responses were matched to their respective supervisors' ratings. Out of 743 matches, 51 records were not used because of outliers, inconsistent responding (random responding), or lack of variability in responses (e.g., answering *strongly agree* to every item). All subsequent analyses presented in this Appendix are based on the remaining records ( $N = 692$ ). Typical incumbents were approximately 38.9 years of age ( $SD = 10.9$  years; range 18 to 78 years), and a majority were male, Caucasian, and had completed a high school diploma. A more detailed breakdown of participants' demographic characteristics is provided in Table A2.

**Table A2**  
***Demographic Characteristics of the Normative Sample***

Characteristic	%
<b>Age</b>	
18–30	25.9
31–45	43.8
46–60	28.9
61+	1.4
<b>Gender</b>	
Female	38.6
Male	61.4
<b>Race/Ethnicity</b>	
African American/Black	20.3
Native American & Alaskan Native	0.9
Caucasian American/White	57.4
Hispanic/Latino	8.8
Asian American & Pacific Islander	5.8
Multiracial	1.3
Other	2.6
No Response	2.9
<b>Education</b>	
No formal education	0.6
Elementary/Middle School	1.5
High School Diploma	47.1
GED	9.5
Trade School Certification	11.3
Associate’s Degree	11.5
Bachelor’s Degree	13.7
Master’s Degree	4.5
Doctorate Degree	0.5

*Note.*  $N = 692$ .

### ***Item Selection for the Performance Assessment***

We randomly split the sample of matched responses ( $N = 692$ ) into two subsamples. The first consisted of 70% of participants ( $n = 484$ ), which was used as the development sample, and the second consisted of the remaining 30% of participants ( $n = 208$ ), which was used as the cross-validation sample. This is a process commonly used in test construction to assess the effects of sampling error and provide replication of the findings.

Item selection for the development sample was based on three criteria: (1) magnitude of correlation with performance criteria (i.e., supervisor ratings) to establish validity, (2) item-total correlation (e.g., internal consistency) to maximize reliability, and (3) close examination of the appropriateness of the content of each item. Incumbents' responses were correlated with the supervisor rating scales. Items that correlated above a specified threshold with any of the performance criteria (or combinations of criteria) were flagged as candidates for inclusion into each scale, namely General Work Attitudes and Risk Reduction. Items with the highest correlations and most appropriate content were selected to form "seed scales." Subsequently, researchers began to add other flagged items to the seed scales and examine how the addition of such items affected the following: (a) the observed validities with performance criteria, (b) the internal consistency reliabilities of the scales, and (c) the correlation between the scales. Objectives for this process were to (a) maximize observed validities with performance criteria, (b) maximize internal consistency reliability, and (c) maintain each scale as relatively unique—that is, keep the scales from being too highly correlated with each other and therefore less likely to provide redundant information and more likely to maximize the predictive power of the scales.

Once researchers were satisfied with the results of the item selection process using the development sample, the properties of the General Work Attitudes and Risk Reduction scales were examined using the cross-validation sample. After the results were replicated with the cross-validation sample, researchers created the Performance Index by first standardizing General Work Attitudes and Risk Reduction scales and then summing them.

### **Properties of the Performance Assessment**

This section features the properties of the Performance Assessment scales and the Performance Index based on the normative sample, including descriptive statistics, reliability, and validity estimates, as well as norms and suggested levels of desirability. Details about the development of a response inconsistency index to identify examinees with inconsistent responding, as well as examination of adverse impact issues, are also provided.

#### ***Descriptive Statistics***

Descriptive statistics (i.e., mean, standard deviation, and range) for the Performance scales based on the full normative sample ( $N = 692$ ) are presented in Tables A3 and A4. Table A3 features descriptive statistics averaged across items for the General Work Attitudes and Risk Reduction scales, both of which have a range from 1 to 6. Table A4 features descriptive statistics based on scale totals for all scales, including the Performance Index. Further, Figures A1 through A3 feature distributions of the scale totals. As illustrated, both the General Work Attitudes and Risk Reduction scales, as well as the Performance Index, approximate a normal distribution. A line representing an ideal normal distribution accompanies each plot for comparison.

**Table A3**  
*Descriptive Statistics at the Item Level*

Variable	Mean	Std Dev	Minimum	Maximum
General Work Attitudes	4.73	.53	2.94	5.94
Risk Reduction	4.33	.62	2.50	5.86

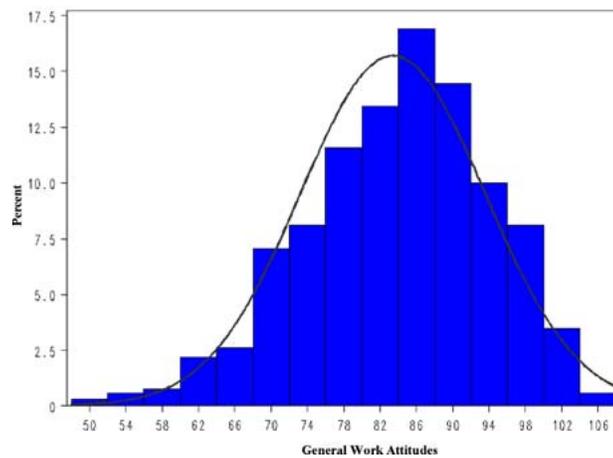
Note. N= 692. The scales' potential range is from 1 to 6.

**Table A4**  
*Descriptive Statistics at the Scale Level*

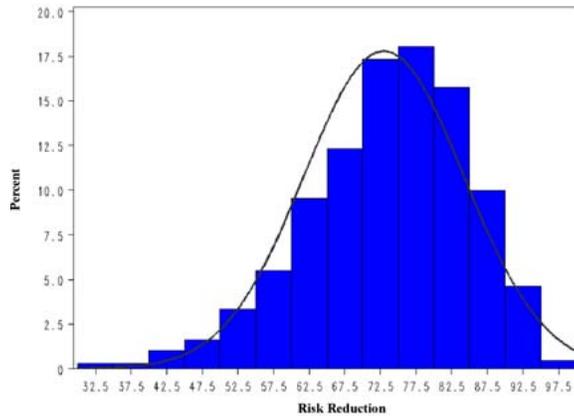
Variable	Mean	Std Dev	Minimum	Maximum
General Work Attitudes	122.9	13.8	76.0	155.0
Risk Reduction	99.5	14.3	58.0	135.0
Performance Index*	100.0	18.1	43.9	142.0

Note. N= 692. \*Calculated by standardizing and summing the other two scales.

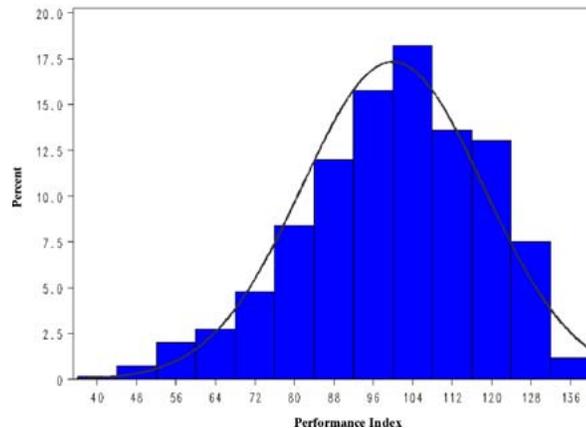
**Figure A1**  
*Frequency Distribution of General Work Attitudes Raw Scores*



**Figure A2**  
*Frequency Distribution of Risk Reduction Raw Scores*



**Figure A3**  
*Frequency Distribution of Performance Index Scores*



### ***Reliability and Validity Estimates***

Table A5 features internal consistency reliabilities (Cronbach's alpha) and observed validities for the development and cross-validation samples. In terms of reliabilities, both the General Work Attitudes and Risk Reduction scales demonstrated good to excellent internal consistency reliabilities (range of alphas across samples = .79 to .84). Further, the Performance Index, which is a combination of both scales, demonstrated excellent internal consistency reliability (range of alphas across samples = .89 to .90).

Regarding the observed validities, although there were some minor differences in the magnitude of correlations from one sample to the other, the pattern of relations with the performance criteria (i.e., supervisor ratings) is quite consistent. Indeed, the differences in observed validities between the two samples ranged from insignificant to small (range of difference = .01 to .14; median difference = .03). This pattern suggests that the Performance Index and its scales are robust across samples. Another pattern worth noting involves the convergent/discriminant relations of the General Work Attitudes and Risk Reduction scales. For example, in terms of convergent relations, both scales appear to tap prosocial/organizational citizenship

behaviors and counterproductive work behaviors. In contrast, in terms of discriminant relations, it appears that the General Work Attitudes scale is more sensitive to task performance behaviors whereas the Risk Reduction scale is more sensitive to safety/risk-taking behaviors. When the two scales are combined into the Performance Index, the index captures the full range of performance criteria.

**Table A5**  
*Observed Validity Correlations and Internal Consistency Reliability for Development and Cross-validation Samples*

	Performance Index		General Work Attitudes		Risk Reduction	
	A	B	A	B	A	B
<b>Task</b>	.16	.14	.22	.19	.08	.06
<b>Prosocial</b>	.29	.22	.31	.23	.22	.16
<b>Counter</b>	.27	.25	.22	.20	.26	.24
<b>Safety</b>	.24	.15	.18	.15	.26	.12
<b>Task, Prosocial, &amp; Counter*</b>	.26	.24	.31	.29	.16	.13
<b>Counter &amp; Safety*</b>	.30	.23	.24	.19	.30	.21
<b>All Supervisor Ratings*</b>	.27	.25	.31	.30	.18	.14
<b>Reliability (coefficient alpha)</b>	.90	.89	.82	.79	.84	.82

*Note.* A = Development Sample [ $n = 484$ ; correlations  $\geq .10$  are significant ( $p \leq .05$ )].

B = Cross-validation Sample [ $n = 208$ ; correlations  $\geq .13$  are significant ( $p \leq .05$ )].

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

#### *Validity Corrections*

We generally do not rely on observed validity as a final estimate of the criterion-validity of a test. This is because observed validity estimates tend to be attenuated or reduced by a variety of biasing effects, such as measurement error (i.e., unreliability in supervisor ratings) and range restriction. For instance, although one is mainly interested in the ability of a test to predict the performance of applicant samples, most test validation research is performed on incumbent samples since it is more feasible to conduct such research with incumbents. Further, incumbent performance cannot be measured without measurement error, as supervisors tend to be idiosyncratic in their ratings. To obtain the “true” (a.k.a. operational) validity of a test, one must use psychometric techniques to correct for such biasing effects (Callender & Osburn, 1980; Raju & Burke, 1983; Sackett & Yang; 2000).

To account for these biases, researchers corrected the observed validities of the Performance Assessment scales shown in Table A5. First, they corrected for measurement error (i.e., supervisor unreliability) using meta-analytically derived inter-rater reliability (ranging from .52 – .60, depending on the content of performance measure) as reported by Viswesvaran et al. (1996). After this correction, the observed validity of the Performance Index using the development sample ( $r = .27$ ), increased to  $r = .38$  (see Table A6).

However, this validity estimate is still not accurate for the desired application of the Performance Assessment, as the correction is limited to the incumbent sample—the basis for supervisors’ ratings—and thus influenced by range restriction. Direct range restriction (DRR) applies when individuals are selected only on the basis of one predictor (i.e., the test of interest). However, in reality, a single predictor is rarely used as the final selection criterion; instead, a variety of sources of information are commonly used (e.g., test scores, structured interviews, letters of recommendation). Thus, indirect range restriction (IRR) applies to cases where a variety of information is used to make selection decisions, which is typical in almost all selection decisions. Validity corrected for IRR is seen as more accurate and more appropriate for use in utility analyses (Schmidt, Oh, & Le, 2006). Researchers further corrected the validity estimates of the Performance Index and its component scales for both DRR and IRR using meta-analytically derived mean range restriction ratio ( $u_x$ ) of .82 as reported in Salgado (2003) and the local reliability estimates (range of .79 to .84) (for a detailed explanation of DRR and IRR, see Hunter & Schmidt, 2004; and Hunter, Schmidt, & Le, 2006). After the additional corrections, the observed validity of the Performance Index using the development sample ( $r = .27$ ), increased to  $r = .44$  when corrected for DRR and to  $r = .46$  when corrected for IRR (see Table 6). Corrections for supervisor unreliability and range restriction using the cross-validation sample are presented in Table A7. As was the case with the observed validities, both sets of corrected validities feature similar patterns.

**Table A6**  
*Validity Correlations for the Development Sample*

Job Performance Criteria	Performance Index				General Work Attitudes				Risk Reduction				
	Operational Validity				Operational Validity				Operational Validity				
	Obs	r	cME	IRR	Obs	r	cME	IRR	Obs	r	cME	IRR	
Task	.16		.21	.26	.27	.22	.28	.34	.36	.08	.10	.12	.13
Prosocial	.29		.40	.47	.49	.31	.43	.50	.53	.22	.30	.36	.38
Counter	.27		.37	.43	.46	.22	.31	.37	.39	.26	.36	.43	.45
Safety	.24		.33	.39	.41	.18	.25	.30	.32	.26	.35	.41	.43
Task, Prosocial, & Counter*	.26		.36	.42	.44	.31	.43	.50	.52	.16	.22	.27	.28
Counter & Safety*	.30		.41	.48	.50	.24	.33	.39	.41	.30	.41	.48	.51
All Supervisor Ratings*	.27		.38	.44	.47	.31	.44	.51	.53	.18	.25	.30	.31

*Note.*  $N = 484$ , Obs  $r$  = observed correlation, cME = corrected only for measurement error in the criterion measure (i.e., supervisor unreliability), DRR = cME further corrected for direct range restriction, IRR = cME further corrected for indirect range restriction. Correlations  $\geq .10$  are significant ( $p \leq .05$ ).

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

**Table A7**  
**Validity Correlations for the Cross-validation Sample**

Job Performance Criteria	Performance Index				General Work Attitudes				Risk Reduction			
	Operational Validity				Operational Validity				Operational Validity			
	Obs	r	cME	IRR	Obs	r	cME	IRR	Obs	r	cME	IRR
<b>Task</b>	.14	.18	.22	.24	.19	.25	.30	.32	.06	.08	.10	.10
<b>Prosocial</b>	.22	.31	.37	.39	.23	.32	.38	.41	.16	.23	.27	.29
<b>Counter</b>	.25	.34	.41	.43	.20	.28	.33	.36	.24	.34	.40	.42
<b>Safety</b>	.15	.20	.25	.26	.15	.20	.24	.25	.12	.17	.20	.22
<b>Task, Prosocial, &amp; Counter*</b>	.24	.33	.39	.42	.29	.41	.48	.51	.13	.18	.22	.23
<b>Counter &amp; Safety*</b>	.23	.31	.37	.39	.19	.27	.32	.34	.21	.28	.34	.36
<b>All Supervisor Ratings*</b>	.25	.34	.41	.43	.30	.42	.49	.52	.14	.19	.23	.25

*Note.* N= 208, Obs r = observed correlation, cME = corrected only for measurement error in the criterion measure (i.e., supervisor unreliability), DRR = cME further corrected for direct range restriction, IRR = cME further corrected for indirect range restriction. Correlations  $\geq .13$  are significant ( $p \leq .05$ ).

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

It is worth noting that the operational validities of the Performance Index for predicting supervisory ratings of job performance are comparable to those reported in the literature. For example, the mean operational predictive validity of integrity tests reported by Ones et al. (1993) in a comprehensive meta-analysis of integrity tests was .41. In our samples, the operational validities ranged from .44 to .47 for the development sample, and from .41 to .43 for the cross-validation sample (see Tables A6 and A7).

Finally, as a way to address sampling error issues, we constructed the 95% confidence intervals (CI) for observed and operational validities corrected for DRR and IRR for both the development and cross-validation samples (see Tables A8 and A9). These intervals represent a 95% level of confidence that the population validity lies between the lower and the upper estimates, as long as the CI does not include zero. As can be seen, very few cases featured in Tables A8 and A9 include zero, and most are well above it.

**Table A8**  
*Confidence Intervals of Validity Estimates for the Development Sample*

Job Performance Criteria	Performance Index								
	95% Obs $r$			95% DRR			95% IRR		
	Obs $r$	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
Task	.16	.08	.25	.26	.12	.39	.27	.13	.42
Prosocial	.29	.21	.37	.47	.34	.60	.49	.35	.64
Counter	.27	.18	.35	.43	.30	.57	.46	.31	.60
Safety	.24	.16	.33	.39	.25	.52	.41	.27	.55
Task, Prosocial, & Counter*	.26	.17	.34	.42	.28	.56	.44	.30	.59
Counter & Safety*	.30	.22	.38	.48	.35	.61	.50	.36	.64
All Supervisor Ratings*	.27	.19	.35	.44	.31	.58	.47	.32	.61

*(Table A8 continues on next page)*

(Table A8, continued)

	General Work Attitudes								
	95% Obs <i>r</i>			95% DRR			95% IRR		
	Obs <i>r</i>	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
<b>Task</b>	.22	.13	.31	.34	.21	.47	.36	.22	.50
<b>Prosocial</b>	.31	.23	.39	.50	.37	.63	.53	.39	.67
<b>Counter</b>	.22	.14	.31	.37	.23	.51	.39	.24	.54
<b>Safety</b>	.18	.10	.27	.30	.16	.44	.32	.17	.46
<b>Task, Prosocial, &amp; Counter*</b>	.31	.23	.39	.50	.37	.63	.52	.39	.66
<b>Counter &amp; Safety*</b>	.24	.15	.32	.39	.25	.52	.41	.26	.56
<b>All Supervisor Ratings*</b>	.31	.23	.39	.51	.38	.64	.53	.40	.67

	Risk Reduction								
	95% Obs <i>r</i>			95% DRR			95% IRR		
	Obs <i>r</i>	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
<b>Task</b>	.08	-.01	.17	.12	-.02	.26	.13	-.02	.28
<b>Prosocial</b>	.22	.13	.30	.36	.22	.50	.38	.23	.53
<b>Counter</b>	.26	.18	.34	.43	.29	.56	.45	.30	.59
<b>Safety</b>	.26	.18	.34	.41	.28	.54	.43	.29	.57
<b>Task, Prosocial, &amp; Counter*</b>	.16	.07	.25	.27	.12	.41	.28	.13	.44
<b>Counter &amp; Safety*</b>	.30	.22	.39	.48	.35	.61	.51	.37	.64
<b>All Supervisor Ratings*</b>	.18	.09	.27	.30	.15	.44	.31	.16	.47

Note. *N* = 484, Obs *r* = observed correlation, DRR  $\rho$  = direct range restriction correction, IRR  $\rho$  = indirect range restriction correction, LL = lower limit, UL = upper limit. Correlations  $\geq .10$  are significant ( $p \leq .05$ ).

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

**Table A9**  
*Confidence Intervals of Validity Estimates for the Cross-validation Sample*

Job Performance Criteria	Performance Index								
	95% Obs <i>r</i>			95% DRR			95% IRR		
	Obs <i>r</i>	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
<b>Task</b>	.14	.01	.28	.22	.01	.43	.24	.01	.47
<b>Prosocial</b>	.22	.09	.35	.37	.15	.58	.39	.16	.62
<b>Counter</b>	.25	.12	.38	.41	.20	.62	.43	.21	.66
<b>Safety</b>	.15	.02	.29	.25	.03	.46	.26	.03	.50
<b>Task, Prosocial, &amp; Counter*</b>	.24	.11	.37	.39	.18	.61	.42	.19	.65
<b>Counter &amp; Safety*</b>	.23	.10	.36	.37	.16	.58	.39	.17	.62
<b>All Supervisor Ratings*</b>	.25	.12	.38	.41	.19	.62	.43	.21	.66

	General Work Attitudes								
	95% Obs <i>r</i>			95% DRR			95% IRR		
	Obs <i>r</i>	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
<b>Task</b>	.19	.06	.32	.30	.09	.50	.32	.10	.54
<b>Prosocial</b>	.23	.10	.36	.38	.16	.59	.41	.17	.64
<b>Counter</b>	.20	.07	.33	.33	.11	.55	.36	.12	.59
<b>Safety</b>	.15	.01	.28	.24	.02	.45	.25	.02	.49
<b>Task, Prosocial, &amp; Counter*</b>	.29	.17	.42	.48	.27	.68	.51	.29	.73
<b>Counter &amp; Safety*</b>	.19	.06	.33	.32	.10	.53	.34	.11	.58
<b>All Supervisor Ratings*</b>	.30	.18	.43	.49	.28	.69	.52	.30	.74

*(Table A9 continues on next page)*

(Table A9, continued)

	Risk Reduction								
	95% Obs <i>r</i>			95% DRR			95% IRR		
	Obs <i>r</i>	LL	UL	DRR $\rho$	LL	UL	IRR $\rho$	LL	UL
<b>Task</b>	.06	-.08	.20	.10	-.12	.31	.10	-.13	.34
<b>Prosocial</b>	.16	.03	.30	.27	.05	.50	.29	.05	.53
<b>Counter</b>	.24	.11	.37	.40	.19	.61	.42	.20	.65
<b>Safety</b>	.12	-.01	.26	.20	-.02	.42	.22	-.02	.45
<b>Task, Prosocial, &amp; Counter*</b>	.13	-.01	.27	.22	-.01	.45	.23	-.01	.48
<b>Counter &amp; Safety*</b>	.21	.08	.34	.34	.12	.55	.36	.13	.59
<b>All Supervisor Ratings*</b>	.14	.00	.28	.23	.01	.46	.25	.01	.49

Note.  $N = 208$ , Obs  $r$  = observed correlation, DRR  $\rho$  = direct range restriction correction, IRR  $\rho$  = indirect range restriction correction, LL = lower limit, UL = upper limit. Correlations  $\geq .13$  are significant ( $p \leq .05$ ).

\*These scales are based on combinations of the Task, Prosocial, Counter, and Safety dimensions.

### ***Norms and Desirability Levels***

Table A10 features a crosswalk between the scales' scores and the corresponding percentiles using the normative sample (i.e., the combination of the development and cross-validation samples). As noted in Chapter 5, percentiles scores are presented in all Performance Assessment reports along with the following desirability levels: low (1–15th percentile), moderate (16–75th percentile), and high (76–99th percentile). We provide these levels as a frame of reference for making inclusion/exclusion decisions. Although we recommend a “top-down” approach to selection, it is helpful to provide a visual based on the normal distribution.

**Table A10**  
***Percentiles for the Performance Assessment Scales***

Percentiles				Percentiles				Percentiles			
Score	GWA	RR	PI	Score	GWA	RR	PI	Score	GWA	RR	PI
23		1		69	1	3	8	115	28	86	77
24		1		70	1	3	8	116	32	87	79
25		1		71	1	3	9	117	33	89	80
26	1	1		72	1	4	9	118	36	91	81
27	1	1		73	1	5	10	119	39	91	83
28	1	1		74	1	5	11	120	40	93	84
29	1	1		75	1	6	11	121	43	93	85
30	1	1		76	1	7	12	122	45	95	87
31	1	1		77	1	7	14	123	47	96	88
32	1	1		78	1	8	15	124	50	96	90
33	1	1		79	1	9	16	125	55	97	90
34	1	1		80	1	9	17	126	57	97	91
35	1	1		81	1	10	19	127	61	98	92
36	1	1		82	1	12	20	128	64	99	92
37	1	1		83	1	14	21	129	66	99	93
38	1	1		84	1	15	22	130	70	99	95
39	1	1		85	1	17	23	131	71	99	95
40	1	1	1	86	1	19	25	132	74	99	95
41	1	1	1	87	1	20	25	133	77	99	96
42	1	1	1	88	1	21	27	134	78	99	97
43	1	1	1	89	1	24	28	135	81	99	97
44	1	1	1	90	2	24	30	136	84	99	97
45	1	1	1	91	2	27	31	137	85	99	98
46	1	1	1	92	2	31	33	138	86	99	98
47	1	1	1	93	3	33	36	139	89		98
48	1	1	1	94	3	34	38	140	91		98
49	1	1	1	95	3	38	39	141	91		99
50	1	1	1	96	4	41	41	142	93		99
51	1	1	1	97	4	42	43	143	94		99
52	1	1	1	98	5	46	45	144	96		99
53	1	1	1	99	6	50	48	145	97		99
54	1	1	1	100	7	51	50	146	98		99
55	1	1	2	101	8	54	52	147	99		99
56	1	1	2	102	10	59	54	148	99		99
57	1	1	2	103	10	59	56	149	99		99
58	1	1	2	104	11	62	58	150	99		99
59	1	1	3	105	11	65	59	151	99		99
60	1	1	3	106	13	68	61	152	99		99
61	1	1	3	107	15	71	63	153	99		99
62	1	1	4	108	15	73	65	154	99		99
63	1	1	5	109	16	75	67	155	99		99
64	1	1	5	110	18	77	68	156	99		99
65	1	1	5	111	19	79	70	157			99
66	1	2	6	112	22	82	72	158			99
67	1	2	7	113	24	85	73	159			99
68	1	2	7	114	25	85	75	160			99

*Note.* GWA = General Work Attitudes, RR = Risk Reduction, PI = Performance Index.

### ***Classification Accuracy***

Table A11 features the cross-tabulation of individuals' scores on the Performance Assessment (across rows) and supervisor ratings (by columns), with each set divided into three levels: low, medium, and high. The desirability levels noted in Chapter 5 and the section above were selected with two goals in mind: (1) to maximize correct classification rates (e.g., individuals who scored in a range consistent with their supervisor's ratings) and (2) minimize misclassification rates (e.g., individuals whose scores were highly discrepant with their supervisor's ratings, such as someone who scored in the low range but was evaluated as being a high performer by a supervisor or vice versa).

**Table A11**  
***Classification Rates for the Normative Sample Using the Performance Index***

---

		Supervisor Rating			Totals
		Low	Med	High	
Examinee Score	Low	8.3	4.8	2.5	15.6
	Medium	17.8	21.6	18.0	57.4
	High	5.8	10.2	10.9	27.0
	Totals	32.0	36.6	31.4	100.0

---

*Note.*  $N = 692$ .

### ***Development of Response Inconsistency Indicator***

As noted in Chapter 5 of this guide, when the Performance Assessment is scored, individuals with inconsistent response patterns are flagged. Response inconsistency is detected when individuals respond to items randomly and without regard to the item's content. Scores for individuals who are flagged for inconsistent response patterns should be interpreted with considerable caution. In the field study conducted by ACT, we found that approximately 7% of examinees were flagged for inconsistent responding.

### *Examination of Adverse Impact*

As noted in Chapter 3, research on integrity testing has established that appropriate use of integrity tests does not result in adverse impact. ACT researchers conducted analyses of the Performance Assessment to ensure that there are no indications of adverse impact on the basis of age, gender, race/ethnicity, or other demographic characteristics. Table A12 features correlations between Performance Assessment scores (Performance Index, General Work Attitudes, and Risk Reduction) and several demographic characteristics using the normative sample. As can be seen, most of these correlations did not reach statistical significance. Of those correlations that reached statistical significance, most reflected education level, suggesting that more educated individuals tend to obtain slightly higher scores on the Performance Assessment. Further, the correlations that reached statistical significance were small in terms of magnitude ( $r < .15$ ), suggesting that these differences are very unlikely to result in adverse impact when used in applied settings. This is consistent with the research literature, which has found that low magnitude correlations with demographic variables do not result in adverse impact (Sackett & Wanek, 1996; Schmidt et al, 2001).

**Table A12**  
***Performance Correlations with Demographics***

	Performance Index	General Work Attitudes	Risk Reduction
Age (continuous)	-.02	-.03	.00
Gender (M = 0 F = 1)	.03	.08	-.02
Ethnicity (Cauc = 0 Min = 1)	-.05	-.10	.02
Education (continuous)	.14	.14	.12

*Note.*  $N = 692$ . Correlations  $> .07$  are significant ( $p < .05$ )

To further examine the issue of adverse impact, ACT researchers compared the classification rates for Caucasians and individuals from racial/ethnic minority groups when using Performance Index scores. Tables A13 and A14 feature classification percentages for Caucasians and racial/ethnic minorities (collapsed into one group), respectively, using the normative sample. Similar to the classification information presented in Table A11, Tables A13 and A14 feature the cross-tabulation of individuals' scores on the Performance Assessment (across rows) and supervisor ratings (by columns), with each set divided into three levels: low, medium, and high. As can be seen, Caucasians and minority individuals were classified at similar rates. Indeed, when the two tables were compared using chi-square statistical tests, the tests *did not* reveal any differences between the two groups.

**Table A13**  
*Classification Rates for Caucasians Using the Performance Index*

		Supervisor Rating			Totals
		Low	Med	High	
Examinee Score	Low	6.4	4.8	3.3	14.5
	Medium	17.9	19.9	18.4	56.1
	High	6.6	9.4	13.3	29.3
	Totals	30.9	34.2	34.9	100.0

Note. N = 392.

**Table A14**  
*Classification Rates for Racial/Ethnic Minorities Using the Performance Index*

		Supervisor Rating			Totals
		Low	Med	High	
Examinee Score	Low	10.6	4.7	1.1	16.4
	Medium	18.2	23.7	17.2	59.1
	High	4.7	11.7	8.0	24.5
	Totals	33.6	40.1	26.3	100.0

Note. N = 274.

## Summary

This Appendix presented a brief background on the development process and psychometric properties of the Performance Assessment. As described, the Performance Assessment was developed using a multistep, rigorous scientific process in which the development results were replicated using a cross-validation sample. As a result, the test produces reliable findings, is predictive of a variety of performance criteria (e.g., task and job performance, prosocial/organizational citizenship behaviors, counterproductive behaviors, safety/risk-taking behaviors), and has evidenced robust validity estimates consistent with those in the meta-analytic literature. Also consistent with the literature, the research findings of the Performance Assessment show that the assessment does not result in adverse impact on the basis of demographic characteristics such as age, gender, race/ethnicity, or education.

# References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education–The Joint Committee on Standards for Education and Psychological Testing. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Association of Test Publishers. (2002). *Guidelines for computer-based testing*. Washington, DC: Association of Test Publishers.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.
- Borman, W. C., Penner, L. A., Allen, T. D., & Motowidlo, S. J. (2001). Personality predictors of citizenship performance. *International Journal of Selection and Assessment, 9*, 52-69.
- Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65*, 543-558.
- Coyne, I., & Bartram, D. (2002). Assessing the effectiveness of integrity tests: A review. *International Journal of Testing, 2*, 15-34.
- Equal Employment Opportunity Commission. (1978). Civil Service Commission, Department of Labor and Department of Justice. Uniform guidelines on employee selection procedures, *Federal Register, 43*(166) 38290-38315.
- Furnham, A. (2001). Personality and individual differences in the workplace: Person-organization-outcome fit. In B. W. Roberts, & R. Hogan (Eds.), *Personality Psychology in the Workplace* (pp. 223-251), Washington, DC: American Psychological Association.
- Hiredesk. (2006). Personality and job fit assessment in candidate screening. Pulse Report (3). Retrieved September 12 from [http://www.hiredesk.com/knowledge/best\\_practices.htm](http://www.hiredesk.com/knowledge/best_practices.htm)
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.
- Hunter, J. E., & Schmidt, F. L., & Le, H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594-612.
- International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing, 6*, 143-171.
- Le, H., & Schmidt, F. L. (2006). Correcting for indirect range restriction in meta-analysis: Testing a new analytic procedure. *Psychological Methods, 11*, 416-438.
- Murphy, K. R. (2000). What constructs underlie measures of honesty or integrity? In R. Goffin and E. Helmes (Eds.), *Problems and Solutions in Human Assessment: A Festschrift to Douglas N. Jackson at Seventy* (pp. 265-284). London: Kluwer.
- Ones, D. S. & Viswesvaran, C. (1998a). The effects of social desirability and faking on personality and integrity assessment for personal selection. *Human Performance, 11*, 245-269.

- Ones, D. S., & Viswesvaran, C. (1998b). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83*, 35-42.
- Ones, D. S., Viswesvaran, C. & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679-703.
- Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity generalization. *Journal of Applied Psychology, 68*, 382-395.
- Rotundo, M. & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology, 87*, 66-80.
- Sackett, P. R. (1994). Integrity testing for personnel selection. *Current Directions in Psychological Science, 3*, 73-76.
- Sackett, P. R. (2002). The structure of counterproductive work behaviors: Dimensionality and relationships with facets of job performance. *International Journal of Selection and Assessment, 10*, 5-11.
- Sackett, P. R., & Wanek, J. E. (1996). New development in the use of measures of honesty, integrity, conscientiousness, dependability, trustworthiness, and reliability for personnel selection. *Personnel Psychology, 49*, 787-829.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112-118.
- Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment, 10*, 117-125.
- Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology, 76*, 323-346.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.
- Schmidt, F. L., Thoresen, C., Le, H., Ilies, R. & Holland, E. (2001). *Report on development and validation of the index of work attitudes (IOWA)*. Unpublished manuscript, State of Iowa.
- Schmidt, F. L., Oh, I., & Le, H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. *Personnel Psychology, 59*, 281-305.
- Society for Industrial and Organizational Psychology. (2003). Principles for the validation and use of personnel selection procedures. (4th ed.) College Park, MD: SIOP
- Spherion. (2006). The Spherion Emerging Workforce Study. Retrieved November 2, 2006, from [http://spherion.com/about\\_us/emerging\\_workforce/emerging\\_workforce.jsp](http://spherion.com/about_us/emerging_workforce/emerging_workforce.jsp)

- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology, 90*, 108-135.
- Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81*, 557-560.
- Wanek, J. E. (1999). Integrity and honesty testing: What do we know? How do we use it? *International Journal of Selection and Assessment, 7*, 183-195.
- Wanek, J. E., Sackett, P. R., & Ones, D. S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology, 56*, 873-894.

# Support and Customer Service

## Telephone

Client assistance will also be provided over the phone at 1-800/WorkKey (1-800/967-5539) or via e-mail at [workkeys@act.org](mailto:workkeys@act.org)

## Web

Electronic customer support can be found at [www.act.org/workkeys/assess/performance/index.html](http://www.act.org/workkeys/assess/performance/index.html)

The WorkKeys Internet Version User Guide can be downloaded at [www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf](http://www.act.org/workkeys/pdf/WorkKeysInternetUserGuide.pdf)