

## CHAPTER SIX

**Measures of Training and Experience**

Timothy P. McGonigle and Christina K. Curnow  
*Caliber, an ICF International Company*

**OVERVIEW**

In this chapter we describe the use of measures of training and experience to support personnel management decisions. We begin by briefly summarizing the history of using information about applicants' training and background to screen them in employment decisions. Though popular, these methods produce a number of psychometric concerns that must be addressed if training and experience measures are to be used reliably, validly, and with minimal adverse impact. In subsequent sections of the chapter, we discuss alternative data collection procedures and scoring approaches. The alternatives include task-based questionnaires, KSAO-based questionnaires, and accomplishment records. We next describe procedures for developing each type of measure, including steps to reduce self-report bias. We conclude the chapter by summarizing the major considerations in developing and using measures of applicant training and experience.

**BACKGROUND**

Measures of training and experience (T&E) are among the most commonly used personnel selection methods (Gatewood & Feild, 2001). The main objective of T&E evaluation is to screen applicant backgrounds according to specific job requirements. T&Es differ from biographic inventories in their use of weighting systems based on judgment rather than on empirically derived scoring keys (McDaniel, Schmidt & Hunter, 1988a), and their use of verifiable training, education, and experience indices. In addition, biographic inventories measure

individuals' experiences with a wide range of life events whereas T&Es focus on experience with specific, job-related tasks.

T&Es typically require applicants to report their level of experience performing—or education pertaining to—particular job tasks and are presumed to be valid based on the theory of behavioral consistency (Ash, Johnson, Levine, & McDaniel, 1989; Guion, 1998; Owens, 1976; Schmidt & Hunter, 1998; Wernimont & Campbell, 1968). This theory posits that the best predictor of future performance is past performance in a similar setting. Researchers have studied the relationship between experience and performance in a variety of public-sector occupations, and in other less common occupations such as crisis counselors (Elkins & Cohen, 1982) and surgeons (Sutton, Wayman, & Griffin, 1998). However, many researchers and practitioners traditionally consider T&Es to be poor predictors of job performance. However, recent evidence suggests that it is possible to overcome the psychometric limitations of T&Es.

T&Es have been widely used in public-sector employment (McDaniel & Schmidt, 1985). As the size of the federal government's workforce grew rapidly during the early 1900s, efforts were made to hire civil servants based on merit rather than the patronage of a particular individual or political party. The purpose of merit-based selection was to ensure a stable workforce that operated regardless of politics. In addition, the increasing size of the federal government attracted large volumes of applicants, thereby necessitating a cost-effective method of screening applicants. T&Es were found not only to be cost effective and consistent with merit-based hiring, but they also required little staff training or oversight to implement. As a result, T&Es were used with increasing frequency and are still used to make many staffing decisions. With the advent of automated prescreening systems, T&Es remain as popular as ever.

Despite their administrative ease, relatively little consideration has been given to whether T&Es are valid predictors of job performance. In fact, early research suggested that T&Es were poor predictors of job performance (Hunter & Hunter, 1984). Schmidt et al. (1979) explained experience questionnaires' lack of validity by discussing some assumptions of the method. Specifically, because measures of experience are estimated to correlate .40 with knowledge, skill, and ability (KSA) requirements and KSAs are estimated to correlate .50 with job performance, the authors argued that the validity of experience questionnaires is limited to about .20 (i.e.,  $.40 \times .50$ ). More recent research suggests that modifications in the measurement mode of T&Es can produce much larger validity coefficients (McDaniel, Schmidt, & Hunter, 1988a, 1988b; Quinones, Ford, & Teachout, 1995; Schmidt & Hunter, 1998; Tesluk & Jacobs, 1998). Such modifications are presumed to increase the correlation between the measure of experience and the KSAs.

Only recently have researchers (e.g., Quinones et al., 1995; Tesluk & Jacobs, 1998) begun to develop a theory of experience. Quinones et al. (1995) provided a theoretical model of work experience that categorizes experience in terms of both measurement method and specificity. The model specifies three measures of experience—amount, time, and type—and three levels of experience specificity—experiences with tasks, jobs, and organizations. After fully crossing both dimensions to create nine types of experience measures (e.g., amount of task experience, time

## 6. MEASURES OF TRAINING AND EXPERIENCE

161

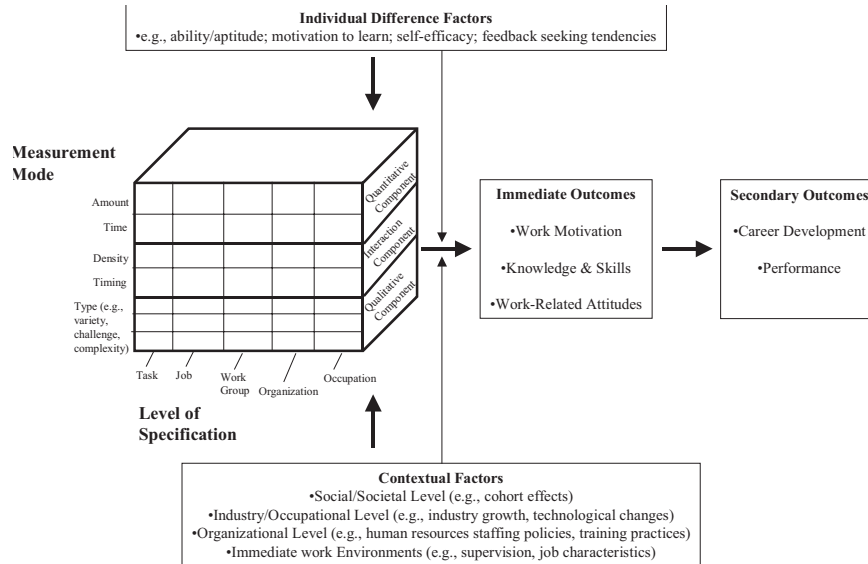


FIGURE 6.1 A model of work experience. Reprinted with permission.

spent in a specific job), the authors conducted a meta-analysis to identify the relationship between each type of experience measure and job performance. They reported corrected correlations up to .43. More recently, Tesluk and Jacobs (1998) expanded the Quinones et al. (1995) model by theorizing multidimensional mediators, criteria, individual differences, and contextual factors that moderate the experience-performance relationship, as well as two additional measurement methods and two additional levels of experience specificity (see Fig. 6.1). These models provide a theoretical foundation for experience measurement that has been lacking for decades.

**MEASURES OF TRAINING AND EXPERIENCE**

Many forms of T&Es have been developed, including holistic judgments résumé screens, point methods, task methods, and accomplishment records. Most T&E measures have been administered to make both entry-level selection and promotion decisions. When used for entry-level positions, T&Es typically focus on determining if an applicant has the requisite KSAs to perform the job. On the other hand, when used for the purpose of selecting employees for promotions or hiring job applicants into non-entry-level positions, T&Es help to determine whether applicants have experience performing tasks that are similar to those performed in the position or job class to which they are applying. T&Es have been administered as prescreens in a multiple hurdle selection system and as one instrument

in a compensatory test battery. The most common use of the T&E as a prescreening device is as a minimum qualification screen, where applicants must show that they have a certain amount of job-related experience, training, or education to proceed within the selection process.

With the exception of the accomplishment record, virtually all T&E methods can be organized according to two dimensions—data collection procedures and scoring procedures. Data collection procedures generally include application blanks and résumés, checklists, and constructed response forms. Scoring procedures generally include holistic judgment, point methods, and analyst ratings. The creation of specific T&E methods (e.g., improved point method) involves combining a data collection method (e.g., checklist) and a scoring procedure (e.g., scoring algorithm). In the next sections, we describe common data collection and scoring procedures and the accomplishment record.

### Data Collection Procedures

#### *Application Blanks and Résumés*

Perhaps the most ubiquitous T&E data collection method is the application blank. Application blanks are used to collect specific information about an applicant's work history and educational background. Résumés can be substituted for an application blank and may describe the major duties associated with each position in the work history, specific coursework, special skills and credentials, and other qualifications. Typically, an analyst makes a holistic judgment about the applicability of the information provided on the application or résumé. Although these methods of data collection allow applicants to provide information on their background, they provide very little information about the quantity or quality of the applicants' qualifications.

#### *Checklists*

T&E checklists ask applicants to complete a rating form or checklist indicating their experience with a variety of work behaviors or tasks (Gatewood & Feild, 2001). On a typical task-based questionnaire (TBQ), applicants indicate whether they have performed the tasks, how often they have performed the tasks, or how much time they have spent performing the tasks (Lyons, 1984). On some forms, applicants may rate how effectively they have performed the tasks, how closely they were supervised in performing the tasks, whether they have received training directly related to the tasks, or whether they have trained others on the task (Anderson, Warner, & Spencer, 1984; Ash, 1981; Farrell, 1979; Malinowski, 1981; Ocasio, 1983). Checklist methods may also ask applicants to provide information about their work experience, educational background, or specific skills that might have prepared them to perform each task.

An alternate approach to the checklist method is the improved point method (McGonigle & Curnow, 2002; Swander & Shultz, n.d., cited in Ash et al., 1989) or KSA-based questionnaire (KSABQ). To develop a KSABQ, subject matter experts

## 6. MEASURES OF TRAINING AND EXPERIENCE

163

(SMEs) identify activities that applicants could have performed that would indicate their proficiency with each job-relevant KSA. Applicants indicate their level of experience with each activity and receive 1 point for each activity they have undertaken. As a result, the improved point method is a less arbitrary approach to assigning point scores on experience questionnaires and is more likely to comply with the *Uniform Guidelines on Employee Selection Procedures* (1978), because it measures experience with behaviors that are indicative of each KSA.

**Scoring Procedures***Holistic Judgment*

Holistic judgment is the most common method used to evaluate applicant training and work experience. This particular method is not a formally scored T&E evaluation method but rather provides a general evaluation of an applicant's credentials. An example of the use of this method begins when a hiring authority receives a set of résumés from applicants for a particular position. The hiring authority assesses each application as a whole and makes a subjective decision about how to distinguish between qualified and unqualified applicants according to the hiring authority's individual standards. Then, the hiring authority selects several résumés and those individuals are contacted for further evaluation, typically through an in-person interview (e.g., see chap. 8). As might be expected, this method lacks both structure and objectivity. In addition, many judgments based exclusively on an applicant's résumé are confounded with information gained through interaction with the applicant. Consequently, this method can lead to very subjective selection decisions.

*Point Methods*

The point method is the most prevalent formal T&E evaluation technique. It consists of a mechanical formula in which applicants receive a prescribed number of points for each month or year of relevant training, education, and experience. In some cases, the number of points assigned varies by the type and duration of experience. Applicants are either rank ordered or are grouped based on specific education and experience requirements for the target job.

Although this method of T&E scoring is much more structured than the holistic judgment method, research shows it does little to improve the validity of the resulting scores (Ash et al., 1989; McDaniel et al., 1988b; Schmidt & Hunter, 1998) perhaps due to the dubious job-relatedness and specious precision of the scoring protocol. As discussed by Ash et al. (1989), the weak validity coefficients might also be due to the large amounts of measurement error introduced by a focus on the quantity of applicants' experiences rather than on their quality. Furthermore, because of its primary focus on the quantity of experience, the traditional point method can also result in adverse impact (Ash et al., 1989). There also is evidence to suggest that two individuals with equal amounts of job tenure can differ drastically in the number and types of tasks they have performed (Ford, Quinones, Sego, & Sorra, 1992; Schmitt & Cohen, 1989).

### Accomplishment Records

The accomplishment record (AR) is a means of gathering self-reported and verifiable descriptions of experience on relevant behavioral job dimensions (Hough, 1984). Methodologically, the AR draws on the critical incident technique (Flanagan, 1954; Guion, 1998; also see chap. 3) and written sentence-completion protocols (Loevinger & Wessler, 1970). The goal of this method is to rank order applicants based on the types of achievement behaviors that are necessary for successful performance in the target job. The focus of an AR is on the quality of previous experience rather than on the quantity of experience. In a typical AR, applicants provide written descriptions of accomplishments that demonstrate their level of proficiency with job-related KSAs. Analysts then rate each accomplishment using a behaviorally anchored rating scale and the ratings are used to compute a total score.

### PSYCHOMETRIC CHARACTERISTICS OF MEASURES OF TRAINING AND EXPERIENCE

In this section, we describe the reliability, validity, and subgroup differences in measures of training and experience. In addition, we discuss the issue of response distortion and describe methods for reducing this problem.

#### Reliability

The reliability of T&Es varies depending on the particular data collection and scoring procedures. Due to their unstructured nature, data collected through application blanks or résumés and subjected to holistic judgments are likely to have little to no reliability. Task-based methods have reliabilities in the .80s (Ash & Levine, 1985; Schmidt et al., 1979) and above (Sneed, Vivian, & D'Ocasta, 1987), primarily due to the straightforward nature of the scoring. Although only a few studies report on the reliability of the AR, there is evidence that it demonstrates reliabilities by dimension ranging from .75 to .85 (Hough, 1984; Hough, Keyes, & Dunnette, 1983). Sadowski and Hess (1994) reported a reliability of .84 when using an AR to evaluate teaching performance. Finally, past research showed inter-rater reliability estimates of T&E ratings to be around .80 (Gatewood & Feild, 2001).

#### Validity

Several meta-analyses reported validity coefficients for T&Es (McDaniel et al., 1988b; Quinones et al., 1995; Schmidt & Hunter, 1998). These meta-analyses show that behavioral consistency (i.e., AR) measures are the most valid ( $r = .45$ ), followed by self-ratings of KSA proficiency ( $r = .20$ ), self-ratings of task proficiency ( $r = .15$ ), and the point method ( $r = .11$ ). Quinones et al. (1995) report a correlation of .43 between amount of task experience (i.e., number of times performing a

## 6. MEASURES OF TRAINING AND EXPERIENCE

165

task) and job performance, although most traditional T&Es do not measure experience in this way.

There is a limited amount of evidence about the validity of ARs, although available research indicates that they are valid measures of performance with validity as high as .45. Hough (1984) and Hough et al. (1983) obtained significant correlations between ARs and supervisory ratings of performance for attorneys. Hough (1984) and Hough et al. (1983) also found AR scores to correlate significantly with pay grade, but not with other traditional predictors such as grades, or scores on tests of law aptitude, knowledge, and achievement. Therefore, the use of ARs could increase the validity of a selection system.

At least four factors contribute to the typically low validity coefficients associated with T&Es (McDaniel, Curnow, & McGonigle, 2003). First, there are relatively few studies of T&E validity, so the meta-analytic validity coefficients cited earlier are likely to be unstable. Second, most T&Es do not account for individual differences in what is gained from training and experience. At least three factors influence whether an individual gains skill from experience (McGonigle & Curnow, 2003):

- Individual differences such as cognitive ability (Kanfer & Ackerman, 1989), conscientiousness, openness to experience, and extraversion (Barrick & Mount, 1991).
- Situational characteristics such as opportunities to perform tasks (Alexander, 1997).
- Motivation to pursue available opportunities (Kanfer & Ackerman, 1989).

Third, due to the fact that all applicants who complete and return a T&E are likely to have some relevant experience, T&Es suffer from range restriction. Finally, in many cases the relationship between experience and performance is nonlinear. Ackerman (1988) demonstrated that task consistency and complexity moderate the validity of the experience-skill acquisition relationship, with early career experience typically yielding the greatest improvements in job knowledge. To the extent that these factors can be controlled through data collection and scoring procedures, the validity of T&Es can be expected to improve.

### Subgroup Differences

There is no evidence of subgroup differences on T&Es. In fact, one author suggested that an explanation for the popularity of T&Es is their perceived lack of subgroup differences (Aramburu-Zbala-Higuera, 2001). However, education and experience may reflect subgroup differences. Specifically, T&Es may reflect the adverse impact of different rates of graduation between African Americans and Whites. Greene and Winters (2005) noted that in 2002, about 78% of White students graduated from high school with a regular diploma, compared to 56% of African American students and 52% of Hispanic students. The *Journal of Blacks in Higher Education* (2002) noted that the nationwide college graduation for

African American students is 42% whereas the graduation rate for White students is 62%. Consequently, T&Es will likely reflect these different rates in their measures of education. In terms of experience, the extent to which a group is under- or overrepresented in the workforce also will foster adverse impact. For example, according to the Office of Personnel Management's *Federal Civilian Workforce Statistics Fact Book* (2004), in 2003, 79.5% of clerical employees were women. Thus, for experience in clerical jobs, women will likely obtain higher T&E scores than men when applying for such jobs.

In addition, T&Es may reflect the subgroup differences of other selection procedures when an applicant's acquisition of experience is considered. When using TBQs to select employees, the probability of being selected will necessarily increase with previous experience. If the opportunity to perform tasks (i.e., gain experience) is earned as a result of being selected using methods that were not free of subgroup differences in prediction, as described earlier, T&Es could perpetuate these differences (Roth, Bobko, Switzer, & Dean, 2001). The same effect would be seen if managers used inherently biased procedures to assign individuals to tasks.

### Response Distortion

A final concern about T&Es is that they rely on the integrity of the applicant. Applicants applying for jobs may be tempted to overestimate the amount of education or experience they have had to obtain a higher score and increase their chances of getting a job. There are methods that can help reduce fallacious reporting of education and experience. For example, warnings to applicants suggesting that responses will be verified tend to reduce falsification of information (Lautenschlager, 1994). Other procedures can be used to identify inaccurate responses and minimize their occurrence. For example, Pine (1995) reported that relative frequency scales produced more incidents of false reporting on task inventory statements than absolute scales. The inclusion of bogus tasks as part of a lie scale can be used to identify applicants with a propensity to provide inaccurate information. For example, Green and Stutzman (1986) reported that 57% of the respondents in a sample indicated that they spent time performing bogus tasks and 72% indicated that these bogus tasks were at least somewhat important aspects of their job. Green and Veres (1990) used a similar method in three different samples using a variety of response scales and found the percentage of respondents endorsing bogus items ranged from 12.6% to 70.3%. A potential concern with using bogus tasks is the similarity of such tasks to real tasks. People may claim to perform tasks using nonexistent pieces of equipment, not because they are lying, but because they do not know the technical name of the equipment and believe that the bogus piece of equipment sounds similar to the real one.

By taking appropriate steps during the development of T&E measures we believe that the types of psychometric difficulties summarized in this section can be overcome. It is possible to develop measures that possess adequate reliability and validity, and that minimize adverse impact.

## 6. MEASURES OF TRAINING AND EXPERIENCE

167

### HOW TO DEVELOP MEASURES OF TRAINING AND EXPERIENCE

In the previous sections, we discussed the many forms of T&Es and their associated psychometric characteristics. In this section, we discuss the two major steps for developing T&Es: select the type of T&E and develop T&E forms and scoring system.

#### Select the Type of T&E

After conducting a thorough job analysis and determining which KSAs will be measured using a T&E (e.g., see chaps. 2, 3, and 4), it is necessary to determine which type of T&E measure to use: task based, KSA based, AR, or some combination of these possibilities. Questions that can help guide this decision are shown in Table 6.1.

In answering these questions, consider the typical applicants' experience: Is it an entry-level job? Are applicants likely to have experience performing the tasks on the job? Are applicants likely to have varied experiences? For example, applicants for a first-line supervisory position may have performed some of the technical tasks in a previous job, but may not have experience with the supervisory tasks. For entry-level jobs and jobs in which previous experience is unlikely, a KSA-based T&E is most appropriate. For promotions and for jobs in which applicants are expected to have previous experience with the specific job tasks, a task-based T&E or an AR is appropriate. If applicants are expected to vary in their experiences (e.g., if there are numerous feeder jobs), it may be difficult to develop items that are applicable across the applicant pool.

Also consider some of the characteristics of the job: Does the job require writing? How many skilled incumbents are there? If the job does not require writing, it is advisable to avoid the AR due to its reliance on constructed (i.e., written) responses. If there are not a large number of job incumbents (i.e., at least 30) available to serve as SMEs and help with test development, then it is advisable to use a task or KSA checklist rather than an AR. Questions to consider when selecting a T&E instrument are provided in Fig. 6.2.

#### Develop T&E Forms and Scoring System

##### *Task-Based Questionnaire*

Developing TBQs involves four steps:

- Determine whether each task is needed at entry.
- Create a response form.
- Create a scoring algorithm.
- Develop a methodology to encourage truthful responses.

SMEs play a role in each of these activities. We describe each step next.

***Determine Whether Each Task Is Needed at Entry.*** To ensure that a TBQ only includes those tasks that applicants need to be able to perform upon entry into

Question	TBQ	KSABQ	AR
Is it an entry-level job?	No	Yes	No
Are applicants likely to have experience performing the tasks required on the job?	Yes	No	Yes
Are applicants likely to have varied achievements?	No	No	Yes
Does the job require writing?	No	No	Yes
Is there a sufficient incumbent pool to assist with test development?	No	No	Yes

FIGURE 6.2 Selecting a type of T&E instrument. TBQ = task-based questionnaire; KSABQ = knowledge-, skills-, and abilities-based questionnaire; AR = accomplishment record.

the job, it is important to collect job analysis data on whether the ability to perform the task is needed on the first day of the job. The majority of SMEs should rate a task as needed at entry to the job for it to be included in the TBQ. In addition, the T&E developer should remove or revise any tasks that require knowledge that can only be gained on the job in question (e.g., knowledge of a specific form or procedure). Next, to ensure that experience on each task in the TBQ demonstrates one or more qualifying KSAs, it is important to have SMEs indicate which qualifying KSAs are needed to perform each of the selected tasks. Only tasks that are linked to at least one KSA by the majority of SMEs should be included in the TBQ.

**Create a Response Form.** Two examples of response forms that could be used for a variety of electrician tasks are shown in Fig. 6.3. Both forms require applicants to estimate the number of times they have performed each task rather than simply the duration of experience (Quinones et al., 1995). On the first form, applicants indicate the duration of their experience with each task in years and months as well as the frequency of their experience (i.e., daily, weekly, monthly, quarterly, yearly) during that time period. An analyst can use that information to calculate the number of times each applicant has performed each task. For example, if applicants indicate that they performed a task quarterly for 3 years and 6 months, the applicants would have performed the task 14 times. The advantage of this form is that it provides applicants with a structure within which to estimate their experience. It collects memorable elements of their experience that can be systematically combined by an analyst.

One disadvantage of the first form in Fig. 6.3 is that it does not account for variation in experience over time. For example, it would be difficult for applicants

## 6. MEASURES OF TRAINING AND EXPERIENCE

169

who performed a task monthly for 2 years and then weekly for 1 year to accurately represent their level of experience. The second form is designed to accommodate varying frequency of experience. On this form, applicants select the option that is closest to the number of times they have performed each task and provide the name of someone who can verify their experience. To help applicants make more accurate judgments of their experience, the form also provides examples of how long it would take to amass each amount of experience assuming that applicants performed the task daily, weekly, or monthly.

**Create a Scoring Algorithm.** The scoring algorithm is designed to take into account applicants' experience on each task. As workers perform a task, they generally become more effective at it, although improvement in performance associated with more experience is likely to diminish at greater levels of experience (McDaniel et al., 1988a; Schmidt, Hunter, & Outerbridge, 1986). To ensure that scores on the TBQ reflect this relationship between experience and performance, the scoring algorithm should award points for increasing amounts of experience performing each task. However, when experience with each activity reaches the point at which additional experience is not expected to improve performance (i.e., the performance asymptote), the algorithm should award no additional points.

Identification of the performance asymptote is a judgment-based process similar to setting a cutoff score. For each task, SMEs are asked to judge the number of times that an individual would need to perform the task to reach the performance asymptote. The SMEs should be instructed to judge the performance asymptotes individually and then come to consensus about the appropriate performance asymptote for each activity. Alternatively, SMEs could be asked to estimate the probability that an applicant with specific amounts of experience performing each task would perform well on the job.

In the two examples shown in Fig. 6.3, the resulting frequency of task performance scores can range from 0 to 5. A score of 0 is assigned when applicants indicate that they have no experience with a task. A score of 5 is assigned when applicants have performed a task at least the number of times indicated by the performance asymptote. Scores from 1 through 4 are assigned for linear increases in experience. In Fig. 6.3, the asymptote for the task "interpreting residential electrical plans" is 50. This means that after interpreting residential electrical plans 50 times, one is not likely to become more proficient at this task.

**Develop a Methodology for Encouraging Truthful Responses.** There are three techniques that can be used in concert to reduce false or exaggerated responses. First, the form can require applicants to sign a statement certifying the accuracy of their responses and describing the consequences of inaccurate responses. Second, the form can require applicants to identify one or more individuals who can attest that the applicants performed the task. Finally, the form can include a series of bogus tasks (i.e., tasks that would be impossible for applicants

Skill in performing residential electrical work				
Please rate your experience with each of the activities below:	How much work experience do you have performing this activity?			Who can verify your work experience with this activity?
	Duration		Frequency*	
	Years	Months		
1. Installing a ceiling mount fixture.			① ② ③ ④ ⑤	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
2. Installing a three-way switch.			① ② ③ ④ ⑤	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
3. Interpreting residential electrical plans.			① ② ③ ④ ⑤	① ② ③ ④ ⑤ ⑥ ⑦ ⑧ ⑨ ⑩
*0 – No experience; 1 – Yearly; 2 – Quarterly; 3 – Monthly; 4 – Weekly; 5 – Daily				

Skill in performing residential electrical work					
Activity	Frequency	Examples			Verifier(s)
	Which option is closest to the number of times you have performed this activity?	To perform this activity the number of times listed in the option to the left, you'd have to perform the task...			Who can verify your work experience with this activity?
		every day for about...	or every week for about...	or every month for about...	
Interpreting residential electrical plan.	a. 0 times	--	--	--	
	b. 10 times	2 weeks	2 months	1 year	① ⑥
	c. 20 times	4 weeks	5 months	1½ years	② ⑦
	d. 30 times	6 weeks	7 months	2½ years	③ ⑧
	e. 40 times	8 weeks	9 months	3½ years	④ ⑨
	f. 50 times or more	10 weeks	1 year	4 years	⑤ ⑩

FIGURE 6.3 Examples of T&E rating scales; on a different part of this instrument, respondents indicated people who can verify their experience with the activities listed and those individuals were labeled “1” through “10”; thus, response options 1 through 10 refer to individuals who can verify experience.

## 6. MEASURES OF TRAINING AND EXPERIENCE

171

to have experience performing). Applicants should be informed that these bogus tasks are randomly placed throughout the form to help alert scorers when applicants falsify or exaggerate their level of experience. Taken together, these features discourage false or exaggerated responses and provide ways to detect such instances when they occur. To develop bogus tasks, one can create statements that sound plausible but are not performed on any job. One method for developing these tasks is to create a statement using words that might be relevant to the job in question when taken alone, but not when put together. It is important to have SMEs review the bogus tasks to make sure they could not be performed and are not obviously bogus. For example, the real tasks “install batteries in parallel to provide backup power source” and “splice high-voltage cable to establish service to a residential building” could be used to produce the following bogus task: “Splice high-voltage cable to provide backup power source.” It is unlikely that any qualified applicant would indicate experience performing the bogus task. As mentioned previously, candidates may endorse a bogus task because they perceive it to be a legitimate task, not necessarily to falsify their application. Including several bogus tasks, some of which do not sound similar to real tasks, may serve to reduce this error.

### **KSA-Based Questionnaire**

Developing KSABQs is largely similar to developing TBQs. The largest difference involves developing the item content. The process of determining whether each task is needed at entry to the job, creating a response form, creating a scoring algorithm, and developing a methodology to encourage truthful responses is identical to that for a TBQ.

KSABQs are designed to measure applicants' experience with job-relevant KSAs without assuming that applicants have direct job-related task experience. Therefore, it is important that applicants can describe their experiences with behaviorally oriented, observable activities measuring job-related KSAs. Activities are behaviors that represent qualitatively different levels of proficiency with each skill or ability. In addition, the activities are designed to measure experience related to each skill or ability even if the applicant has no direct experience performing the job tasks. For example, although applicants may not have experience with a specific task (e.g., “measure length of wire needed to pull through conduit and attach to the tugging machine”), they may have experience performing other activities that require the same underlying skills or abilities (e.g., “measure the length of wire needed to install receptacle”).

To generate activities, SMEs should identify the types of non-job activities they themselves might have performed to develop proficiency with each KSA. For example, if people are applying for the job of an entry-level electrician, it is possible that they have wired or rewired household appliances. The resulting activities

should be retained only if the SMEs agree that they are related to the KSA in question and applicants could be expected to have performed them. Much like developing any high-quality test items, the process of activity generation is likely to involve iterative review and revision of item content.

### **Accomplishment Record**

The process for developing an AR consists of four basic steps:

- Conduct job analysis.
- Develop the inventory.
- Administer the inventory.
- Develop rating scales and rating principles.

Step 1 involves conducting an accurate job analysis, which is a crucial step in developing a reliable and valid AR. To support an effective AR, the job analysis data should specify job dimensions or competencies. Once the job analysis is complete, the AR inventory can be developed (Step 2). An AR inventory provides targeted performance dimension definitions and typically requests specific information for each competency or job dimension listed, including: a general statement of the accomplishment, a precise description of exactly what was done, a time period, a description of any formal recognition that resulted from the accomplishment, and contact information of one person who can verify the accomplishment. Applicants are instructed to write about their most meritorious accomplishment for each work performance dimension. An example of a completed accomplishment statement is shown in Fig. 6.4.

Once the AR has been developed, Step 3 involves administering it to current job incumbents to collect examples of accomplishments and to test the usability of the AR instructions. Incumbents are instructed to complete the AR based on their experiences prior to their current jobs. However, collecting accomplishments from job incumbents may result in a set of accomplishments that are all highly rated. Alternatively, non-incumbents with similar experience to potential applicants could complete the AR inventory. However, non-incumbents participating in the development process would be precluded from subsequently applying for that particular job.

Step 4 is to develop the AR rating scales and rating benchmarks. This involves a four-part process of evaluating dimensionality of the accomplishments, gathering expert ratings, inducing the principles underlying the rating process, and selecting benchmarks. The discussion of the critical incident technique in chapter 3 explains key parts of the process in more detail.

**Part 1.** The first step in developing the AR scoring key is to evaluate the accomplishments to ensure that they comprise clearly differentiable dimensions. To do this, raters classify accomplishments generated in Step 3 (with no identifying competency information) into competency categories. The accomplishments

**WRITTEN COMMUNICATION**

***Definition: Communicates technical and nontechnical information in writing, using correct English grammar and sentence structure that can be understood by the intended audience.***

***Time Period:*** 1995–96

***General Statement of what you accomplished:***

In conjunction with another I/O psychologist in my office, I co-authored a training manual on how to conduct job analysis.

***Description of exactly what you did:***

There is a wide range of methods for how to conduct job analysis. Within my organization we decided that it would be best to use one consistent method. So, a co-worker and I wrote a training manual on how to conduct job analysis. I wrote two sections of the manual: conducting job observations and collecting job analysis ratings. This manual is still being distributed to all of our new staff on an almost daily basis. We have requested several printings—we are currently on about our fourth printing of the manual. I have received a great deal of positive feedback about the helpfulness of the manual to workers. Many supervisors have commented on how quickly they are able to get new hires up to speed on our job analysis process now that they have the training manual.

***Award or Recognition:***

None

***The information can be verified by:***

Joan Q. Supervisor, (123) 456-7890

FIGURE 6.4 Example of an accomplishment record response for written communication.

are presented to raters without information about the dimensions for which they were originally written. Accomplishments that were written for one dimension and classified by raters into another dimension are considered misclassifications. Any misclassifications can be assessed and the job dimensions can be reviewed to determine if they should be combined or eliminated. Raters can be experts in job analysis and test development or SMEs.

Part 2. The next step is to gather expert ratings on the effectiveness of accomplishments for each competency. SMEs rate the accomplishments within each dimension on a 5- or 6-point scale. Parts 1 and 2 can be combined and both classification and ratings can be done at the same time. For parts 1 and 2, a minimum of three raters has been suggested (Hough, 1984). In operational scoring, responses to AR questions can be very diverse and can tax the ability of scorers to achieve rater agreement. By using multiple independent raters and averaging their ratings for each candidate, the developer can increase the reliability of the ratings.

Part 3. Next, the accomplishments are ordered from high to low within each dimension based on mean SME ratings. A research team should analyze the content of the accomplishments in the high range to identify the themes or principles that SMEs used to judge the incidents. Repeat the content analysis at the middle and low levels of the scale and for each performance dimension. These themes specify the elements that raters will look for when rating achievement level.

Part 4. Means and standard deviations of expert ratings are then calculated. Accomplishments with standard deviations lower than 1.0 can be chosen as benchmarks at various points on the scale (Hough, 1984). An example of the benchmarks appears in Fig. 6.5.

#### SUMMARY

Despite their dubious ability to predict performance, measures of training and experience have been among the most commonly used selection tools for decades due to their conceptual appeal, low cost, and ease of administration. However, more recent theoretical work has provided a foundation for improving the measurement of training and experience. By using theory-based measurement techniques, the accuracy and quality of the resulting data are likely to be improved, as is the validity of the resulting decisions. Specifically, T&Es that collect amounts of experience (i.e., number of times having performed a task) or quality of experience (i.e., ARs) show fairly strong levels of validity, whereas holistic judgment-based procedures such as resume screens and application blanks show almost no validity. There is little evidence of subgroup differences on T&Es; however, T&Es could perpetuate differences in opportunity to perform tasks (i.e., gain experience) as well as differences in educational degrees received.

There are two steps to developing T&Es: (a) selecting the type of T&E and (b) developing T&E forms and the scoring system. In selecting a type of T&E, it is important to consider the typical applicants' level and type of experience as well as characteristics of the job in question. TBQs are most appropriate when applicants are expected to have some opportunity to gain experience with the job tasks; otherwise KSABQs are more appropriate. ARs are most appropriate for professional jobs that require significant writing ability (e.g., attorney). Developing TBQs involves identifying the most appropriate tasks from the job analysis, whereas developing KSABQs involves developing activity statements exemplifying the types of experiences applicants might be expected to have. Developing ARs

## 6. MEASURES OF TRAINING AND EXPERIENCE

175

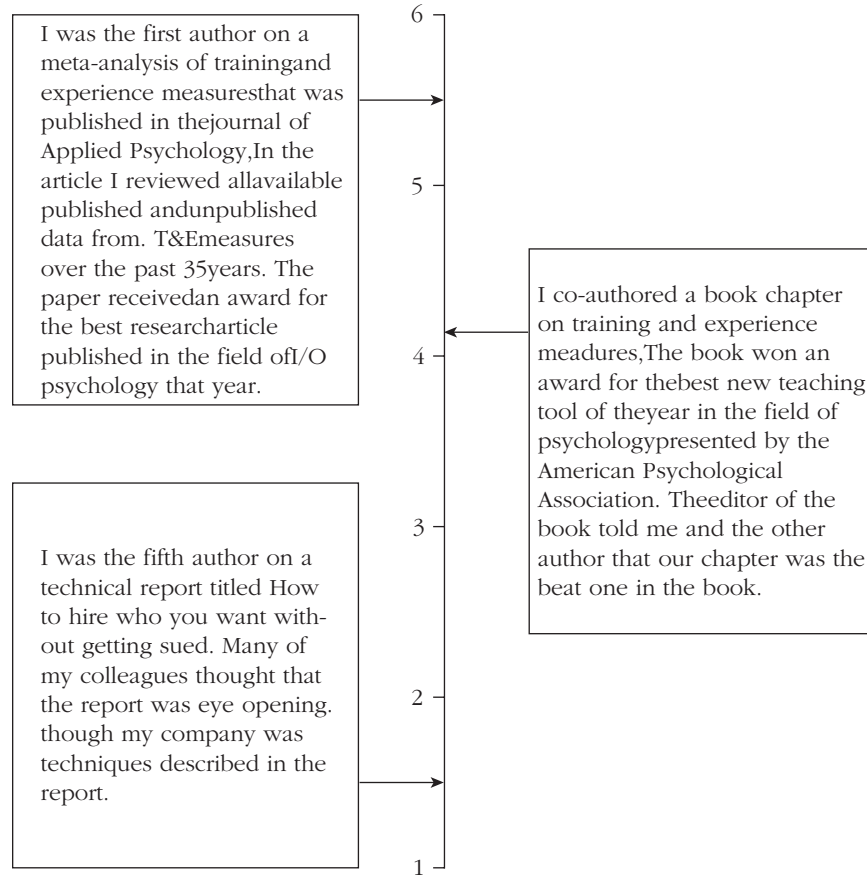


FIGURE 6.5 Accomplishment record rating scale for written communication.

involves identifying the critical performance dimensions for which applicants must describe their accomplishments. Scoring TBQs and KSABQs should take into account the fact that the improvement in performance associated with more experience is likely to diminish at greater levels of experience. Scoring ARs involves developing behaviorally anchored rating scales for each performance dimension.

**REFERENCES**

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*, 288–313.

- Alexander, P. A. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic sources. *Advances in Motivation and Achievement, 10*, 213–250.
- Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology, 69*, 574–580.
- Aramburu-Zbala-Higuera, L. (2001). Adverse impact in personnel selection: The legal framework and test bias. *European Psychologist, 6*(2), 103–111.
- Ash, R. A. (1981). Comparison of four approaches to the evaluation of job applicant training and work experience. *Dissertation Abstracts International, 42*, 4606B.
- Ash, R. A., Johnson, J. C., Levine, E. L., & McDaniel, M. A. (1989). Job applicant training and work experience evaluation in personnel selection. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 7, pp. 183–226). Greenwich, CT: JAI Press.
- Ash, R. A., & Levine, E. (1985). Job applicant training and work experience evaluation: An empirical comparison of four methods. *Journal of Applied Psychology, 70*, 572–576.
- Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.
- Black student college graduation rates remain low, but modest progress begins to show. (2002). *Journal of Blacks in Higher Education, 50*. Retrieved April 14, 2006, from <http://www.jhbe.com>
- Elkins, R. L., & Cohen, C. R. (1982). A comparison of the effects of prejob training and job experience on nonprofessional telephone crisis counselors. *Suicide and Life-Threatening Behaviors, 12*(2), 84–89.
- Farrell, B. (1979). *Task performance self-evaluation: An alternative selection procedure to traditional experience and training ratings*. Minneapolis: Selection Research Unit, Minnesota Department of Personnel.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327–358.
- Ford, J. K., Quinones, M., Segó, D., & Sorra, J. (1992). Factors affecting the opportunity to perform trained tasks on the job. *Personnel Psychology, 45*, 511–527.
- Gatewood, R., & Feild, H. (2001). *Human resource selection* (3rd ed.). Fort Worth, TX: Harcourt Brace.
- Green, S. B., & Stutzman, T. (1986). An evaluation of methods to select respondents to structured job analysis questionnaires. *Personnel Psychology, 39*, 543–564.
- Green, S. B., & Veres, J. (1990). Evaluation of an index to detect inaccurate respondents to a task analysis inventory. *Journal of Business and Psychology, 5*, 47–61.
- Greene, J. P., & Winters, M. A. (2005). *Public high school graduation and college-readiness rates: 1991–2002* (Education Working Paper No. 8). New York: Center for Civic Innovation at the Manhattan Institute.
- Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hough, L. M. (1984). Development and evaluation of the “accomplishment record” method of selecting and promoting professionals. *Journal of Applied Psychology, 69*, 135–146.
- Hough, L. M., Keyes, M. A., & Dunnette, M. D. (1983). An evaluation of three “alternative” selection procedures. *Personnel Psychology, 36*, 261–275.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin, 96*, 72–98.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/attribute-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*, 657–690.

## 6. MEASURES OF TRAINING AND EXPERIENCE

177

- Lautenschlager, G. J. (1994). Accuracy and faking of background data. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 391–419). Palo Alto, CA: CPP Books.
- Loevinger, J. S., & Wessler, R. (1970). *Measuring ego development I: Construction and use of a sentence completion test*. San Francisco: Jossey-Bass.
- Lyons, T. J. (1984). *Development and validation of the unassembled clerical exams for clerk typist (GS-4) and clerk stenographer (GS-5): A status report*. Washington, DC: U.S. Office of Personnel Management.
- Malinowski, F. (1981). Job selection using task analysis. *Personnel Journal*, *60*, 288–291.
- McDaniel, M. A., Curnow, C. K., & McGonigle, T. P. (2003, March). *Experience measures: Current theory and practice*. Paper presented at the Personnel Testing Council of Metropolitan Washington, Washington, DC.
- McDaniel, M. A., & Schmidt, F. L. (1985). *A meta-analysis of the validity of training and experience ratings in personnel selection (OSP-85-1)*. Washington, DC: U.S. Office of Personnel Management.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988a). Job experience correlates of job performance. *Journal of Applied Psychology*, *73*, 327–330.
- McDaniel, M. A., Schmidt, F. L., & Hunter, J. E. (1988b). A meta-analysis of the validity of methods for rating training and experience in personnel selection. *Personnel Psychology*, *41*, 283–314.
- McGonigle, T. P., & Curnow, C. K. (2002). Development of a modified improved point method experience questionnaire. *Applied H.R.M. Research*, *7*, 15–20.
- McGonigle, T. P., & Curnow, C. K. (2003, May). *Current theory and practice on the measurement of experience*. Paper presented at the Mid-Atlantic Personnel Assessment Consortium, Rehoboth Beach, DE.
- Ocasio, B. P. (1983). *Validation of the custodial unassembled examination procedure*. Washington, DC: U.S. Postal Service.
- Office of Personnel Management. (2004). *Federal civilian workforce statistics fact book*. Retrieved April 14, 2006, from [www.opm.gov/feddata/factbook](http://www.opm.gov/feddata/factbook)
- Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology* (pp. 609–643). Palo Alto, CA: Consulting Psychologists Press.
- Pine, D. (1995). Assessing the validity of job ratings: An empirical study of false reporting in task inventories. *Public Personnel Management*, *24*, 451.
- Quinones, M. A., Ford, K. J., Teachout, M. S. (1995). The relationship between work experience and job performance: A conceptual and meta-analytical review. *Personnel Psychology*, *48*, 887–905.
- Roth, P. L., Bobko, P., Switzer, F. S., & Dean, M. A. (2001). Prior selection causes biased estimates of standardized ethnic group differences: Simulation and analysis. *Personnel Psychology*, *54*, 591–616.
- Sadowski, C. J., & Hess, A. K. (1994). A modified accomplishment record approach to evaluating teaching effectiveness under the talent development model. *Journal of Personnel Evaluation in Education*, *8*, 41–46.
- Schmidt, F. L., Caplan, J. R., Bemis, S. E., Decuir, D., Dunn, L., & Antone, L. (1979). *The behavioral consistency method of unassembled examining (TM-79-01)*. Washington, DC: U.S. Office of Personnel Management.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). The impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.
- Schmitt, N., & Cohen, S. (1989). Internal analyses of task ratings by job incumbents. *Journal of Applied Psychology, 74*, 96-104.
- Sneed, J., Vivian, V., & D'Ocasta, A. (1987). Work experience as a predictor of performance: A validation study. *Evaluation and the Health Professions, 10*, 42-57.
- Sutton, D. N., Wayman, J., & Griffin, S. M. (1998). Learning curve for oesophageal cancer surgery. *British Journal of Surgery, 85*, 1399-1402.
- Tesluk, P. E., & Jacobs, R. R. (1998). Toward an integrated model of work experience. *Personnel Psychology, 51*, 321-355.
- Uniform guidelines on employee selection procedures. (1978). *Federal Register, 43*, 38290-38315.
- Wernimont, P. E., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology, 52*, 372-376.