Developing Assessments of Trainee Proficiency

By Gordon Welty, Ph.D.

In the world of training the development of assessments is as essential to training administration as keeping score is to any professional sporting event. The author identifies the four key components of the process for developing assessments and how to incorporate them into the training module.

This article is the third in a series on the ADDIE Model. The first article, “Strategy and Tactics of Task Analysis,” appeared in the Journal of GXP Compliance, Volume 11, Number 3, April 2007; the second appeared as “The ‘Design’ Phase of the ADDIE Model,” Volume 11, Number 4, July 2007.

INTRODUCTION

Employees in regulated industries must be trained before they “touch” the product. According to the Food and Drug Administration (FDA), the “person” must have the “education, training, and experience, or any combination thereof, to enable that person to perform the assigned functions.” This requirement has substantial implications for both managers and regulators. It also has substantial implications for the assessment of trainee proficiency.

Education and experience are dealt with differently than training. Education is typically documented in terms of grade level completed by - or diplomas awarded to - students in accredited institutions on the basis of a series of learning experiences over an extended period of time. Experience typically refers to work experience, and is documented by the employee’s length of time in a specific job position; many times that is an extended period of time. None of these measures is entirely adequate; in both work experience and education, the qualitative richness of the process is obscured by the measure of duration. But they provide guidance to corporate policy nonetheless.

The pharmaceutical or other regulated company
seeks to ensure that employees have requisite educational attainment by effectively implementing recruitment policies or policies in support of continuing education. The company can either recruit employees who meet educational levels or can subsidize the employees’ further education until they meet the desired level. Likewise, the company strives to ensure that employees have the requisite work experience. The company can either recruit employees who have specific work experience or can retain employees until they acquire the desired level of on-the-job experience.

Training, by contrast, is typically documented in terms of specific training events in which the employee has participated. Management develops an Individual Training Plan (ITP) for each job position, listing the various training modules that incumbents must complete.

The ITP and the associated training can be contrasted to the employees’ educational attainment and work experience in several ways. First, the ITP tends to foreground short-term objectives rather than long term. Second, it tends to focus on enhancing employees’ task performance on-the-job, rather than their more general knowledge and experience. Third, because of the longer time frames associated with educational attainment and work experience, these factors are taken as givens by management. Training, however, is widely seen as a corrective action for many problems. To the extent employees manifest performance gaps, those gaps will typically be remediated by training. Jack Gordon has pointed out that

“managers send people to [training] courses precisely because they want to see observable behavior changes that will produce observable business results.”

So management’s concern is with behavior change, which necessitates assessment of the effect of training on the trainee’s task performance. Should the training not correct the performance gap, the management will then turn to the more extreme measures of progressive discipline or discharge.

This concern is shared by regulators, who want to ascertain how well the organization has its processes in control. If there are problems in pharmaceutical manufacturing, let us say, the regulator wants to know how the organization responded to the problem. How was management notified? Was an investigation conducted, and if so, how well? How was the root cause identified? How was the corrective action formulated, and how was it executed? Many times the “cause” of the problem is identified as the residual category “human error,” hardly a root cause at all. And then “retraining” is proposed as the “corrective action,” as though the initial training could unexceptionably have been ineffective. Regarding auditors, as James Vesper has put it,

“GMP auditors and regulatory inspectors are becoming more savvy about training and performance. They are asking to see evidence that the training was effective.”

Following the lead of management and regulators, we will focus in this article on the development of training assessments. There are four components to the process of developing assessments: (1) identifying the behavioral objectives of the training module, based on the relevant operational procedure and the Training Outline, (2) determining the kind of assessment to be used, (3) preparing the assessment materials, and (4) incorporating these assessment materials into the training module.

We will now examine the four components of the process of developing training assessments.

**BEHAVIORAL OBJECTIVES IN THE ADDIE MODEL**

The ADDIE model is a generic instructional design model. It provides guidance at a fairly high level for instructional designers, software engineers, etc., as they author and revise learning products. The phases of the ADDIE model are Analyze, Design, Develop, Implement, and Evaluate. These phases are sequential - each depends upon the successful completion of the preceding phase.

Behavioral objectives have been specified in the Design phase of the ADDIE model. These objectives have several important functions. First, they permit
intended training outcomes to be aligned with organizational objectives. Second, they permit trainees to have clear expectations of intended training outcomes. Third, they provide a sound basis for the design and development of training materials.\(^7\)

Behavioral objectives have a number of dimensions.\(^8\) Each must specify:
- The training audience
- The trainee performance
- Conditions for the performance
- The measure of the performance
- Criteria for successful performance

Let us consider each of these dimensions in turn.

**Training Audience**
The training audience is the set of job positions whose incumbents must receive training before they "touch" the product, or, in the case of supervisors, before they supervise employees who will be touching the product.\(^9\) All job positions that have responsibilities in a given Standard Operating Procedure (SOP) are part of the training audience for that procedure. For example, by regulation, sanitization SOPs apply to contractors and temporary employees as well as to full-time employees.\(^10\) As Roelof Kuipers has pointed out, it is important to define who needs what kind of training in a given organization.

"With a large pool of electrical, mechanical and maintenance engineers, electricians, machine operators, contractors, and many more, not everyone needs the same skill set."\(^11\)

The Boones have summarized this nicely: “Your behavioral objective should identify the specific audience you plan to target.”\(^12\)

**Trainee Performance**
Trainee performance is the set of behaviors that the trainee will demonstrate upon completing the training. In sanitization processes, regulations stipulate that employees must follow written SOPs.\(^13\) As an example of a behavioral objective that highlights observable performances, “At the end of this training session, the trainee will be able to demonstrate the correct mopping technique for sanitizing controlled areas.” This demonstration - this performance - will be observable to the trainer who will certify the performance took place; in principle, the performance would be observable to the trainee's manager or an auditor as well.

**Conditions of Performance**
The conditions of the performance is a list of conditions under which the trainee is expected to perform the behavior. For example, a behavioral objective might state "At the end of this training session, the trainee will be able to demonstrate the correct mopping technique for sanitizing controlled areas, using the double-bucket process." The bolded text in the objective would be (part of) the conditions of the performance.

**Measure of Performance**
The measure of the performance provides the categories or scale that represents the performance in qualitative or quantitative terms.\(^14\) A measure of performance on a paper-and-pencil quiz could be the number of correct answers (perhaps compared to the total number of questions). The measure of sanitization of an area could be provided by environmental monitoring data. In James Popham's terms, it is a major shortcoming when

"behavioral objectives were not accompanied by assessment instruments specifically linked to those objectives."

He continues that learning objectives

"will have little effect on day-to-day instruction if not accompanied by relevant assessment devices."

FDA regulations stipulate that

"A protocol is required to contain …A description of the observations and measurements to be made to fulfill the objectives of the study."\(^15\)
**Performance Criteria**

Finally, criteria for performance specify the limits of successful performance. For instance, many times, the performance on a paper-and-pencil quiz is considered successful when 80% of the responses are correct. Again, the sanitization of an area has been successful when environmental monitoring data for each room after sanitizing is within acceptable limits. The inclusion of criteria is important because it highlights that the behavioral objectives are built into the assessment measures. As Jack Gordon (ibid) has put it,

“When you know what targets you’re shooting at, keeping score becomes much easier.”

**Section Summary**

In this section we have considered the five dimensions of behavioral objectives - training audience, trainee performance, conditions for the performance, the measure of the performance, and criteria for successful performance - and the role they play in providing focus to the training module. Given the behavioral objectives and their associated measures and criteria, the particular kind of assessment can be stipulated.

**WHICH KIND OF ASSESSMENT**

At a very general level, training involves two components - a Training Event, followed by a Performance that demonstrates whether the training had (or did not have) the desired impact on the job, in the workplace.17

The Training Event might be a structured on-the-job training (SoJT) event; it might be a classroom event; it might be an e-learning event. The Performance might be the SoJT trainee’s independent execution of the relevant tasks; it might be the trainee’s accurate logbook entry following a classroom session; it might be the trainee’s completion of a quiz at the end of an on-line session with 80% of the responses correct. Of course, the performance might be unsuccessful - the trainee might fail to execute the relevant tasks, might make an inaccurate logbook entry, might score less than 80% on the quiz.

The Training Event is a set of independent variables (the predictors); the associated Performance is a set of dependent variables (the criteria). Both components - the Training Event and the Performance - are multidimensional.18

The Training Event includes trainer(s), trainee(s) with skill-set(s) and disposition(s), training organization (supervisory factors, business case), training facilities (allocated space, allotted time, utilities), and training materials (instruments and equipment, raw and in-process materials).19 Training materials also include the training script - for a typical SoJT event, for instance, the script would spell out in some detail the steps in the Prepare, Tell, Show, Do, and Follow-up cycle to be followed in this event.20

The Performance component (continuing with the SoJT illustration) includes the trainee's independent and satisfactory performance of the relevant tasks in a real work setting, as judged by a supervisor or as indicated on some business-process metric, and usually has both individual level and group level (work team) elements. There is a feedback loop between the performance and the training event. As we observed before, it is possible that the task performance by the trainee was unsuccessful. In that case, the adequacy of the trainer’s ability or preparation, the suitability of the training materials, the capability or motivation of the trainee, as well as the timing or situation of the training event (or a combination of these) can be called to account.

The core concepts of Performance are as follows:

- a real work setting wherein
- a task is completed21

This concept of performance is not always logistically feasible. Which tasks in a specific process must be completed? How can a real work setting - with all the demands of production and output - be accessed for training purposes? These are difficult questions to answer, difficult enough that it is frequently necessary to use proxies for purposes of training assessment.

Whether core concepts or their proxies are utilized in assessment of training, they must be documented in procedures, protocols, SOPs. An SOP stipulates the standards for the core concepts of Performance or for the proxies.
Turning first to the core concept “real work setting,” if that setting is unavailable for task samples, a surrogate measure must suffice. Brinkerhoff gives the example of training on cardiopulmonary resuscitation (CPR) techniques:

“Barring a workplace heart attack, we would find no on-the-job application of the skill learned.”

The surrogate in such a case is a Skill Demonstration Assessment (SDA), where the trainee independently performs the task(s) on relevant equipment outside of the real work setting - off-hours, on placebo batches, during production shutdowns, etc.

Turning next to the core concept “task completion,” there are situations where the process cannot be broken into discrete tasks, or is for some reason inaccessible. Consider, for example, equipment that has a biennial preventive maintenance schedule. That equipment may not be available for the training of mechanics for more than a year. In such a case, another kind of proxy must suffice. That is a Knowledge Transfer Assessment (KTA). A KTA is a paper-and-pencil test that predicts performance on-the-job. If task completion or non-completion can be correlated with a test score, so that high scores correlate with task completion and low scores correlate with non-completion, then the KTA is validated, and performance on-the-job can be predicted from trainee performance on the KTA.

If the KTA has not been validated, it can still prove useful as an interactive element within the courseware itself. It can take the form of “study questions,” providing guidance to trainers as they interact with trainees in facilitating the course. Perhaps needless to say, in this form the questions are not part of any assessment.

We have not included Donald Kirkpatrick’s Level 1, the “trainee reaction” measure, in our list of assessments for several reasons. First, there is no evidence that a trainee’s appreciation of - or affective response to - a training event correlates with the trainee’s task performance. Thus the trainee reaction is not a surrogate for performance. Second, if an assessment of the utility of the training content or materials is needed, a review of the module during the pilot implementation, by the training and development peers, will likely provide a more focused and accurate assessment than the reactions of the trainees. Third, the use of trainee reactions raises the possibility of documented negative judgments. For instance, suppose the trainee reaction form uses a question such as “What can be done to improve this training module (or training event).” What shall be the corrective action of the trainer to negative judgments? A regulator may come across these documents during an audit, and can rightly ask about the remediation that followed from them. Better these judg-

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 1
Illustrative SDA items from GMP Train-the-Trainer Program
ments were not documented in the first place, if there were no intention to take corrective action.

**SECTION SUMMARY**

In this section we reviewed several kinds of assessments that can be considered for incorporating in a particular training module. These range from assessments that approximate the core concept of a Performance, through a series of increasingly distant proxies, to an assessment based on the trainee’s knowledge of the job, as reflected in a validated KTA.

**PREPARING THE ASSESSMENT MATERIALS**

Once the kind of assessment has been selected, the assessment materials can be prepared. The first step in preparing assessment materials is to complete a task analysis. Once the task analysis has been completed, the specific tasks and sub-tasks will be listed, groups of tasks will be aggregated or “chunked,” the flow of the process will be indicated, and concepts providing the “science” for task performance will have been associated with each chunk.

This completed task analysis will include an extensive set of tasks. The second step is to winnow through the particular tasks whereby the trainee’s performance will be assessed. One way would be to take a representative or random sample of the set of tasks. Another would be to take a purposive sample of those tasks that are judged critical to the whole process.

Once the list of tasks is a manageable length, this becomes the work sample for assessment. The third step is to prepare a protocol for the assessment, indicating that the trainee is expected to perform the listed tasks under specified conditions, and meeting certain criteria for success. As Vivian Bringslimark has expressed it,

*"using an approved [operational] SOP, a qualified observer or trainer should observe the employee performing the [operational] procedure, compare the performance to the [operational] SOP, and record the results on a qualification or competency assessment sheet. The results should be communicated to the employee, his or her supervisor, and to the trainer responsible for the original training, indicating whether the prescribed level of competency has been attained.*

As we have noted above, there are circumstances where task sampling is not practicable, and a surrogate is necessary for assessment. That surrogate is the SDA. A training procedure stipulates how, when, and where the trainee can independently perform the task on relevant equipment outside of the real work setting. As Bringslimark (ibid) has put it, the process of how the assessment sheets are approved, distributed, and evaluated also should be defined in that [training] SOP as part of the overall training system. We will briefly describe that process.

The originator uses the number and version of the relevant operational SOP as the course number and version for the SDA form. The form includes a number of yes/no statements. These describe the identified critical or representative tasks to be assessed on the SDA. These are the items assessing the trainee’s performance. (See Figure 1)

The trainee performs, and the trainer or some other subject matter expert (SME) monitors the performance and checks each item in turn: “yes” if the performance was successful, “no” if not. When the performance is complete (whether successful or not), the trainee and the trainer sign and date the SDA. Area management may sign as well. The completed form is submitted to the data entry personnel of the validated training tracking system, or, in case of manual data processing, to the staff of the document repository.

If SDAs are not available, situational judgment testing can be a proxy. In a typical situational judgment test, trainees are presented with a variety of situations they might encounter in the workplace. Most situational judgment tests take a paper-and-pencil form, although they could take an on-line form. These situations (or scenarios) are usually established on the basis of a job or task analysis. The trainee selects the best way to handle each situation. The trainee’s choice is com-
pared to a response called “correct.” The “correct” response is established either empirically or by the collective judgment of a panel of SMEs. Should situational judgment testing not be a feasible alternative, a job knowledge test can be a surrogate. A KTA is a paper-and-pencil test that predicts performance on-the-job. The items in the KTA can be constructed either (a) out of material contained in training courses, user manuals, technical handbooks, etc. or (b) from material provided by a panel of SMEs; in either case the material reflects the content of the job. The items that should be selected are the best discriminators between employees who are judged more proficient and less proficient performing the task. Thus high scores correlate with proficiency and low scores correlate with less proficiency; the KTA is validated, and performance on-the-job can be predicted from trainee performance on the KTA.

Section Summary
In this section we have sketched out the preparation of several forms for assessing training, ranging across the continuum from work sampling, through SDAs, situational judgment tests, and finally to KTAs. Once the assessment forms have been prepared, they can be incorporated into the training module.

INCORPORATING ASSESSMENTS INTO THE TRAINING MODULE
Assessments can be incorporated into a training module in several ways: as a pre-training assessment, as a post-training assessment, and interspersed throughout the training material.

Pre-training assessments (pre-tests) are administered before the training begins. These assessments can take the form of task samples, SDAs, or KTAs. If they have been administered before the trainees congregate at the training site, the trainer can compile the scores, which may allow the trainer to adapt the training materials to the specific levels of trainee preparedness.

Post-training assessments (post-tests) are administered after the training has been completed. Again, they can take many forms. They can be administered before the trainees leave the training site, or they can be administered at a later date, or both. If the post-tests are administered while the trainees are still on-site, and then at one or more later times, they can serve as measures of the sustainability of the training as well as the effects of the training. Tennant, et al, (ibid) suggest three kinds of post-training assessments: an “immediate test,” to be carried out when the training has been completed, an “intermediate test” when the trainee has returned to the job, and an “ultimate test” to be employed “after an appropriate time has elapsed in order to measure the improvement of the skills, and behavioral changes.”

Post-test scores can also be compared to pre-test scores. Given equivalent forms, differences in scores can be taken as some evidence of training effects.

Finally, depending on how the work process has been chunked and conceptualized, assessments can be incorporated throughout the training material, in addition to any other assessments that are used as pre- or post-tests. Assessments throughout the material serve to reinforce training at a more fine-grained level, to break up training material into lengths closer to adult attention span, etc.

Test Security
Assessment of training can place trainees under considerable personal and organizational pressure to succeed. In addition, valid assessment forms can be quite costly to develop. Therefore, attention must be paid to ensuring test security - that is ensuring that the training event and associated performance comply in terms of the five dimensions of the behavioral objectives listed above. The performance must be identifiably that of the individual trainee, under the stipulated conditions, and demonstrably successful (or not). These compliance issues have been highlighted by the increasingly widespread use of on-line testing and assessment. The compliance issues have a much longer history, of course, as a review of test security problems involving home-schooled children makes clear.

There are several approaches to test security for
assessment of training. These include verifying the identity of the trainees, and monitoring the assessments. These approaches are quite familiar to those of us who work in regulated industry, and should be stipulated in an appropriate training SOP. Verifying the identities of users (trainees) means being Part 11 compliant. Monitoring the assessment of training means having the task defined in the Action column, and the responsibility for that task listed in the Responsibilities column of the relevant operational SOP.

Section Summary

In this section we have commented on several issues that arise as we incorporate assessments into the training module. These include the timing of assessments - whether to conduct assessments before, during, or after the training event - as well as how to ensure the integrity of the assessment process.

CONCLUSION

In this paper, we considered four components to the process of developing assessments. First, we reviewed five dimensions of behavioral objectives - training audience, trainee performance, conditions for the performance, the measure of the performance, and criteria for successful performance - and the role they play in providing focus to the training module. Next, we examined the kinds of assessments that can be incorporated in a particular training module, ranging across the continuum from work sampling, through SDAs, situational judgment tests, and finally to KTAs. Third, we outlined the preparation of several kinds of forms for assessing training, ranging across the continuum from work sampling, through SDAs, situational judgment tests, and finally to KTAs. Fourth, we commented on several issues that arise as we incorporate assessments into the training module, including the timing of assessments as well as how to ensure the integrity of the assessment process.

Returning then to the ADDIE model, the Analysis phase identifies a performance gap, a discrepancy between a standard stipulated in a procedure (SOP) and some employee performance. A performance gap can be addressed by a training module, that is, a set of training and assessment materials. This is followed by the Design phase, where a planned approach - documented in a Training Outline - is prepared to address the performance gap. Also, behavioral objectives are specified in the Design phase.

If management approves the design, a Development phase comes next, where the training module is created and approved. This phase is where training materials - trainee activities, instructional content, delivery methods, and instructional technologies - and assessment materials are created to close the performance gap. We focused our attention in this article on the development of training assessments.

After the training materials and assessment materials have been developed, the training module can be rolled out in the Implementation phase, and finally studied in the Evaluation phase.

ENDNOTES and REFERENCES

1. Thus for pharmaceuticals - see 21 CFR 211.25; for non-clinical lab personnel, 21 CFR 58.29; for biopharm personnel, 21 CFR 600.10; for medical device personnel, 21 CFR 820.25; for human tissue recovery personnel, 21 CFR 1271.170.
15. Popham, W. James, “Whittling Wish-list Standards Down to
12. Boone and Boone, ibid.
7. Boone, Harry and Deborah Boone “ABC’s of Behavioral
16. See 21 CFR 312.23 (a)(6)(iii)(f) This is true of the study of
18. The Training Event/Performance model is overly general.
19. The organization and its environment - within which the
Training Event, training organization and training facilities
are located - are also important for situating employees and
their tasks. These two categories can have a profound
impact on the conduct and effectiveness of training.
20. See also Paul Lyons, “Scripts Creation Training;” Industrial
and Commercial Training, Vol. 37, No. 6 (2005), pp. 309-
313.
21. Callinan, Militza and Ivan Robertson “Work Sample Testing,”
International Journal of Selection and Assessment, Vol. 8,
No. 4 (2000), pp. 248-260. See also C. L. Brisley, "How You
Can Put Work Sampling to Work;" Factory Management
and Maintenance, Vol. 110, No. 7 (July 1952), pp. 84-89.
(Feb 1988), p. 66.
23. There are substantial legal implications to the use of non-
validated tests in employment-related situations; see U.S.
Department of Labor, “Testing and Assessment: An
Employer’s Guide to Good Practices,” Washington, DC:
Employment and Training Administration, (2000), Chap. 2,
p. 5. See also Elizabeth Shoenfelt and L. Pedigo, “A Review
of Court Decisions on Cognitive Ability Testing, 1992 -
2004,” Review of Public Personnel Administration, Vol. 25,
No. 3, (Sept 2005), pp. 271-287.
24. Kirkpatrick, Donald, Evaluating Training Programs, San
25. Alliger, George, S. Tannenbaum, W. Bennett Jr, H. Traver,
and A. Shotland, “A Meta-Analysis of the Relations Among
Training Criteria,” Personnel Psychology, Vol. 50, No. 2
30-31.
27. Vivian Bringslimark, “If Training Is So Easy, Why Isn’t Every-
one in Compliance?” Biopharm International, Vol. 17, No. 1
(Jan 2004), pp. 49.
28. See Michael McDaniel and N. T. Nguyen, “Situational Judg-
ment Tests,” International Journal of Selection and Assess-
ment, Vol. 9, No. 1 (March 2001), pp. 103-113; also Michael
McDaniel, F.P. Morgeson, E. B. Finnegan, M.A. Campion,
and E.P. Braverman, “Use of Situational Judgment Tests to
Predict Job Performance,” Journal of Applied Psychology,
29. Charles Tennant, M. Boonkrong, and P. Roberts have
stressed the significance of the timing of assessments in

30. This is a broader social issue; see Carolyn Kleiner and M. Lord, “The Cheating Game,” *US News & World Report* (22 Nov 1999).


### ABOUT THE AUTHOR

Gordon Welty, Ph.D., has been designing, developing, facilitating, evaluating, and managing technical and regulatory training programs in the healthcare and pharmaceutical industries for more than twenty years. Contact Gordon at gwelty@wright.edu.