CHAPTER 5
Administration

INTRODUCTION

After the development of the performance tasks is completed, the assessment is ready for administering to examinees. Key to a successful administration is the preparation of instructions and the training of staff. This is true whether the performance assessment will be used for licensure of hundreds of examinees, a multisite evaluation of an educational program, or the conduct of a small-scale research study.

Guiding the administration of any assessment is the need for comparability of test scores across testing conditions, such as time, essay items, tasks, equipment, and test administration sites (Cohen & Wollack, 2006; Dwyer, 1993; Haertel & Linn, 1996; Messick, 1993). Those who administer an assessment often want to look at differences in examinees' scores across time. A law school, for example, might want to know whether changes in course offerings are associated with more examinees passing the tasks in the Multistate Performance Test (National Conference of Bar Examiners [NCBE], 2005). However if the difficulty of the tasks used in the bar examination varies across years, then the faculty will not know whether a spike in scores was due to recent changes in course offerings or due to the tasks in the Multistate Performance Test focusing on less demanding topics than before. Only if the context of the test is similar across administrations will scores be comparable and test users able to ascertain that changes in scores are likely due to changes in examinee proficiency. In the case of the law examination, only if the tasks are of equal difficulty will the faculty know that any differences in examinee

scores are not related to the unique demands of the tasks from one test form to another, and thus, any increase (or decrease) in examinee performance is more likely due to the changes in the course offerings.

STANDARDIZATION

Comparability of scores across contexts is strengthened by the standardization of the test administration. Fortune and Cromack (1995) state, "Standardization involves creating the conditions that assure uniformity of the tests with regard to administration, difficulty, clarity in scoring, and establishing psychometric evidence of the quality of the test" (p. 159). Consider, for example, an evaluator who includes a writing assessment to gauge outcomes of a literacy program. If the pre- and posttests are completed using the same administration instructions, then variation in examinee scores is likely due to the program rather than to changes in instructions for test administration. "When directions to examinees, testing conditions, and scoring procedures follow the same detailed procedures, the test is said to be standardized" (AERA, APA, & NCME, 1999, p. 61). It is such standardization that supports the comparability of the pre- and posttest scores.

Standardization also is considered a fairness issue (Fortune & Cromack, 1995). "For tests designed to assess the examinee's knowledge, skills, or abilities, standardization helps to ensure that all examinees have the same opportunity to demonstrate their competencies" (AERA, APA, & NCME, 1999, p. 61). According to Dwyer (1993), the goal of standardization is to provide "identical conditions of test administration, or the 'level playing field' notion that inferences about the comparability of individuals or groups can only be drawn when comparable tasks, under comparable conditions, have been undertaken" (pp. 270–271). In essence, the goal of standardization is to "provide accurate and comparable measurement for everyone, and unfair advantage to no one" (AERA, APA, & NCME, 1999, p. 61).

The United States Medical Licensing Examination (FSMBUS & NBME, 2004) acknowledges this role of standardization in stating to examinees, "… you will have the same opportunity as all other examinees to demonstrate your clinical skills proficiency. The examination is standardized, so that all examinees receive the same information when they ask standardized patients the same or similar questions" (p. 4). This idea of comparability and fairness will be important later when we discuss the topics of test accommodations and test security.

Clemans (1971) states that standardization "implies rigid control over the conditions of administration. It is, in fact, this control that permits the instrument to be termed standardized" (p. 190). However, the degree to which an assessment is standardized varies (AERA, APA, & NCME, 1999, p. 61). For example, some credentialing agencies use equivalent test forms (i.e., tests developed based on the same test specifications) across administrations rather than the same form (Lamb, 2001). Haertel and Linn (1996) provide an example of a continuum of standardization by contrasting a highly structured assessment which meets Clemans's description of standardization with a more fluid assessment that illustrates the statement of AERA/APA/NCME. At one end of the continuum each student in a class

> works alone, silently attending to his or her own paper. If space permits, students may be seated at every other desk. All have received identical instructions, read from a script provided for the test administrator. They work from identical sets of printed questions, recording their responses on identical answer sheets. Rules about what student questions the teacher may answer (and how they are to be answered), whether calculators may be used, and similar matters are clearly specified. The test is accurately timed. (Haertel & Linn, 1996, pp. 60–61)

Haertel and Linn (1996) then offer the contrast of a more fluid administration of a performance assessment:

> Students might be working in groups; might be using nontext equipment or manipulables; might be free to consult whatever reference materials happened to be available in the classroom; might be free to ask the teacher questions the task designers never anticipated. (p. 61)

According to Haertel and Linn (1996), portfolios further complicate standardization. They state:

> The portfolio usually consists of some required and some optional entries representing the student's best work, culled from up to a year or more of classroom instruction. In this context, rules about appropriate versus inappropriate collaboration or coaching are hard to specify and harder to enforce. A major determinant of the quality of portfolios from a given classroom is likely to be the amount of time and effort the teacher devotes to portfolio-relevant assignments. In addition, the conditions under which students create their portfolios may vary substantially from one classroom to another. Research papers written by students with access to well-stocked school libraries versus an incomplete set of encyclopedias are clearly not

> comparable unless the conditions under which they were created can somehow be taken into account—a problem for which there is as yet no solution. (p. 61)

Although it appears that standardization, and thus score comparability, is unachievable for performance assessment, standardization *can* be incorporated into the design of the assessment. For example, contributing to standardized conditions in portfolios is the specification of (1) the conditions of choice for entries and (2) guidelines about the limits of collaboration (AERA, APA, & NCME, 1999, p. 61). An illustration of this point is provided by the process completed by teachers seeking certification by the National Board of Professional Teaching Standards (NBPTS). As part of the certification process, teachers develop portfolios in which they reflect on student work samples (NBPTS, 2004b). For one component of the portfolio used in English language arts certification, teachers select two students' responses to two texts and two writing prompts. The student responses, the texts, and the writing prompts are teacher selected. However, every portfolio must contain these forms of student writing and the teacher's written commentary in which he analyzes the students' work and reflects on his teaching.

Before proceeding we should clarify a common misconception about standardization. By the examples that have been given, it should be clear that standardization is *not* a synonym for multiple-choice, norm-referenced tests (Cohen & Wollack, 2006). In a discussion on performance assessment, Messick (1993a) notes that standardization serves the purpose of supporting the comparability of scores for both norm-referenced score interpretation and criterion-referenced interpretations (p. 69). Given that scores from multiple-choice tests and performance assessments can be used for norm-referenced or criterion-referenced interpretations, the process of standardization supports the comparability of scores for multiple choice tests across test administrations and the comparability of scores from performance assessments across administrations.

The question might be asked whether standardization is only an issue for large-scale testing. In other words, is standardization an issue for an assessment conducted as part of a small research study or program evaluation? Often research studies want to gauge whether examinees' scores from one administration are different in another administration; that is, they want to apply the familiar pre- and posttest research design. Recall the law school faculty who wanted to examine whether differences in scores over time were due to changes in the course offerings. Without standardization of the administration instructions and

task prompts, a positive change might be due to (1) changes in the course offerings, (2) improved test administration instructions, (3) less demanding tasks, or (4) some combination of the preceding.

A similar dilemma could occur in a research study on the benefits of an experimental instructional strategy in the teaching of writing. If the administration of the assessment allows the use of word processing programs for the students associated with the experimental condition, and no such equipment is allowed in the testing of the student instructed using conventional instructional strategies, then any differences might be due to (1) the benefits of the experimental instructional strategy, (2) the availability of word processing programs, or (3) both. Without standardization of test administration, validity is of concern because of the uncertainty of decisions about the reasons for changes in examinee performances. So, uniformity (i.e., standardization) of the test context supports valid comparisons of examinees' scores across time or types of interventions—whether in large-scale assessment programs or single-site program evaluations.

How do you achieve standardization? Standardization requires developing directions for the examinees, preparing administration instructions for the staff (e.g., test administrator, examiners, and proctors), planning accommodations for examinees with disabilities, and establishing measures to secure test materials. In the following sections we present methods for standardizing the administration of an assessment. Much of the literature we draw on addresses the administration of multiple-choice tests or tests that use both multiple-choice items and performance assessments. We blend the relevant ideas from these sources to describe issues in the administration of an assessment that may be focused on performance assessment only or a combination of multiple-choice and performance assessment.

In addition, the literature variously refers to the staff administering an assessment as test administrators, examiners, proctors, and monitors. The roles and responsibilities associated with these descriptors differ depending on the source. In our discussion the test administrator manages the staff, arranges schedules and rooms, and oversees the distribution of materials. The test examiners distribute the tests and read the test instructions to the examinees. Proctors assist the test examiners in distributing materials and monitoring the room for possible inappropriate examinee behaviors. Finally, to illustrate the process of administering a test, throughout the narrative we intersperse the story of Scott, a vice president for operations at a company that develops and delivers examinations for use in certification and licensure.

Need for Directions

Standardized directions, those for examinees, proctors, examiners, and test administrators, support the uniform administration of the test (Fortune & Cromack, 1995). Uniformity of administration reduces extraneous influences on examinees' scores. As Clemans (1971) states, "Score variations should be due to differences in ability, not to different examination conditions" (p. 190). Said another way, the development of test instructions is meant to "minimize the score variance that results from factors external to the examinees" (Clemans, 1971, p. 189). The concern about score variation reflects the likelihood that "items developed under one set of conditions may yield very different results when administered under another set" (Clemans, 1971, p. 189).

For instance, consider the likely effect on writing scores of the following two sets of directions:

> **Directions:** Write your essay on the following two pages. You have one hour to write your essay.

> **Directions:** Write your response on the Final Writing pages in your test booklet.
>
> - You may make a graphic organizer (such as a web, list, or outline) and write a rough draft on scratch paper.
> - After you write the rough draft, you must write the final draft on the Final Writing pages in your test booklet.
> - You may use a dictionary or thesaurus.
> - You must write only one final draft on the pages in your test booklet.
> - You should read over your final draft and make neat changes in your test booklet. (South Carolina Department of Education [SCDE], 2005a, p. 57)

These two disparate sets of directions would produce scores that are not comparable. Why is this so? From *The NAEP 1998 Writing Report Card for the Nation and the States* (Greenwald et al., 1999), we know that students who plan their response to a writing prompt have higher average writing scores than students who do not. Thus if the first set of directions were used for a pretest and the second set of instructions for a posttest, the researcher or program evaluator would not know whether any increase in scores was due to (1) a writing program, (2) examinees being instructed to plan their response for the posttest, or (3) both.

However, in a standardized administration the test examiners would have a script with the same set of directions that were to be read for the pre- and posttest. The uniform directions, then, would serve to produce examinee scores that are comparable across administrations.

Developing Directions

In writing directions for the administration of an assessment, some issues are relevant to all involved: examinees, test administrators, examiners, proctors, and actors (e.g., standardized patients). Some considerations are role specific, for example, the responsibilities of a proctor in an administration. In this section, we first consider general issues in writing administration directions that are relevant to all involved in the test process. Subsequently we consider the relevant details for writing administration directions for examinees and the test staff.

General Considerations in Writing Administration Directions

The development of the administration directions perhaps does not share the creativity and excitement of constructing a performance task, so the temptation is to write the administration directions only when absolutely necessary. However, the development of the test administration directions should not be left until the last minute. Instead, the development of the directions for test administration should parallel the development of the test (Clemans, 1971; Siegel, 1986). Clemans (1971) notes that because the directions for administration are integral to the assessment, the administration directions should be in *final* form before the test is operational. As part of the test development, then, he notes that when tasks are piloted that are novel in format or content, it is desirable also to pilot the directions to make sure they are understood. As part of this development, he suggests that "a critique by administrators or examinees may be helpful in suggesting changes that will improve communications between [test] author, examiner, and examinee and that will serve to eliminate requirements specified by the author that prove unrealistic in an operational setting" (p. 190).

In preparing directions, the target is for them to have the same meaning for the examinee, examiner, and administrator (Clemans, 1971). This requires that directions be clear and simple. In writing the directions, it might help to assume that the examiner and examinee do not know anything about the task. A delicate balance, however, must be achieved so that the directions are not condescending.

In formatting the instructions, Clemans (1971) states, "It is the author's task to find those elements or characteristics of format that will be the most effective in causing examiners and examinees to follow directions accurately" (p. 191). Visual factors to consider include boldface, italics, underlining, enlargement, contrasting colors, circling key phrases or passages, or using bullets to highlight key points. Examples of visual factors are presented in the following sections.

Examinee Directions

In a manner of thinking, directions to examinees begin long before the examination day. Often examinees first engage the examination process through their registration, although many are automatically registered by virtue of their standing (e.g., students in school and an end-of-course examination). Once registered, examinees receive information regarding the date, time, and location of the examination. In addition to the logistical information, examinees might also receive information regarding appropriate dress (e.g., layers to accommodate personal preference if the room is found to be too warm or cool), suggestions for not bringing personal items into the examination room (NBME, 2003), and perhaps even getting a good night's sleep the evening before the examination.

Also to prepare examinees for the assessment, a descriptive handbook might be provided in advance of testing (Clemans, 1971). Such descriptions can be seen in review of the websites for the analytical writing test for the Graduate Record Examination (Education Testing Service [ETS], 2005); the Multistate Essay Exam and Multistate Performance Test (NCBE, 2005); and Step 2 Clinical Skills of the United States Medical License Examination (FSMBUS & NMBE, 2005). These sites provide information about the tests as well as examples of the tasks.

Examinees should also be told which types of materials and equipment that they can use or will be available for their use. In one state, examinees may use a dictionary or thesaurus in one section of the high school exit examination (SCDE, 2005b). In the writing assessment of the National Assessment of Educational Progress (NAEP), students received a brochure with suggestions about planning, editing, and revising their writing (Persky et al., 2003). Additional materials that may be allowed in the assessments include pencils, scratch paper, and calculators for the mathematics section.

In terms of the directions for the assessment, Siegel (1986) indicates that the development of a work sample (i.e., a performance task) should

include the preparation of examiner and examinee directions, as well as an administrator manual. He recommends that examinee directions be developed for each task. In some instances examinee directions might be administered from a CD or tape (Yap et al., 2005). Important to include in the examinees' directions are:

1. A statement of the task
2. Required test materials
3. General directions
4. How to make responses
5. Time limitations
6. How to correct inadvertent responses
7. What type of assistance examinees may receive if they do not understand the task (AERA, APA, & NCME, 1999; Clemans, 1971; Cohen & Wollack, 2006; Siegel, 1986)

Directions should also orient examinees to test materials such as a computer, calculators, or numeric grids that may be used with math performance items. Examinees should have cues where to GO ON and to STOP in a section of a test (Clemans, 1971). In addition, they should be told whether they may return to earlier sections of the test.

Practice exercises can help with the orientation of examinees to the assessment. This would especially be true for practice on equipment used in the assessment, unless the use of the equipment is part of the assessment. For example, some achievement tests include a practice test to familiarize students with test formats.

In preparing examinee directions for performance assessments, we do well to remember the advice of Bond and colleagues (1996), who state, "The objective of assessment should be not so much the standardization of *instructions,* as ensuring that examinees have a common understanding of the tasks involved" (p. 121). The importance of the test staff in achieving this end is reflected in their statement that "Because performance assessment is, or can be, richly interactive, it is vitally important that administrators not only understand the constructs being assessed, but it is essential that they know how to discern when an examinee does not understand what is being asked and what kinds of additional explanation is needed" (Bond et al., pp. 121–122). Thus, we turn to the topic of administration directions for the test staff.

Administration Directions for Staff

Administration manuals should also be developed for each staff member involved in the test. Although each testing organization establishes policy and procedures that test staff follow, a common set of expectations guide the training of examination proctors and test site administrators. In this section, we present a description of relevant topics for inclusion in the administration manuals for staff.

Test Administrator Directions. Table 5.1 presents an outline that can guide the development of a test administration manual. Major topics include a description of the assessment, the roles and responsibilities of staff, the handling of test materials, the establishment of an appropriate test environment, examinee instructions, administration procedures, and preparation of test materials for return to the testing agency. Another method to organize an administration manual is to cover the topics according to the test timeline: Before the Examination, During the Examination, and After the Examination (NBME, 2003; SCDE, 2005a). Whichever method you use to organize the manual, in planning for the test administration, Clemans (1971) indicates it is useful to consider:

1. For whom is the test developed?
2. When will it be used?
3. Where will it be administered?
4. Who will give the examination?
5. Will alternate forms be essential?
6. Will it be administered to individuals or groups and, if to groups, of what size?
7. What response format will be most appropriate?
8. Will any special preparation of the examinee be necessary? (p. 189)

Answers to these questions will guide you in the development of the manuals.

Critical in a test administrator manual is an outline of the responsibilities of the administrator and staff. A review of the list of administrator responsibilities in Table 5.2 shows the myriad duties that the administrator completes in preparation for the testing. Test administrators plan the various schedules required for the test, make arrangements for the facilities required for the test, attend to the equipment needed for the administration, and select and train the test staff.

TABLE 5.1. Outline of Potential Topics for a Test Administration Manual

I. Description of the assessment
  A. Purpose of the assessment
  B. Assessment structure and time allotments
  C. Schedule for test administration
  D. Discussion of the importance of a uniform administration and comparability of scores
  E. Test security
  F. Test accommodations
  G. Ethical test practices
II. Staff (administrator, examiners, proctors, and actors)
  A. Qualifications for selection
  B. Responsibilities
  C. Training
III. Test materials
  A. Receipt and secure storage
  B. Supplementary materials required (e.g., clock, calculators, dictionaries)
IV. Test Environment
  A. Facility availability
  B. Room arrangement
  C. Seating arrangement
  D. Equipment for testing rooms and centers
V. Examinees
  A. Eligibility
  B. Notification about test
  C. Materials allowed/not allowed to bring to the test site
  D. Admissions procedures for examinees
VI. Administration
  A. Procedures for distribution of materials
  B. Directions for completion of examinee identification information
  C. Instructions (i.e., script) to be read to examinees
  D. Timing the examination sessions
  E. Instructions for the examiner and proctor about assistance that they may provide
  F. Review of equipment used in the test
  G. Completion of practice items
  H. Instructions for completing each item type (e.g., multiple choice, constructed response, essay)
  I. Guidelines for examinees who finish prior to the allotted time
  J. Collection of materials at end of administration
VII. Preparation of test materials for return to test agency
  A. Irregularity reports
  B. Count of test materials
  C. Packaging of materials

Note. Summarized from Clemans (1971); Fitzpatrick and Morrison (1971); Massachusetts Department of Education (2005); National Board of Medical Examiners (2003); and South Carolina Department of Education (2005a).

TABLE 5.2. Typical Responsibilities of Test Administrators

Prior to the test

- Review the administration directions, test booklets, and answer sheets.
- Develop and distribute testing schedules: weeks to administer the test, daily schedule, length of testing sessions, breaks.
- Provide examinees with the time and location for the test, the materials needed (pencil), the name of the examiner, and the make-up date.
- Develop checklists for the steps completed in administration of tests, packaging of completed tests.
- Establish testing stations and rotation schedules.
- Plan for the distribution of materials on the examination day.
- Select staff (e.g., examiners, proctors, and actors).
- Tell examiners and proctors their duties during each stage of the test administration.
- Provide each examiner with a copy of the manual and the test (if security permits).
- Train actors in simulations in their roles and to follow the script closely.
- Train the observers of performances/simulations to understand the criteria and scoring rules.
- Check the functioning of equipment used in completion of performance tasks.
- Rehearse administration with new examiners.
- Determine that the correct number of tests are available.
- Review the examination rooms to make the physical environment as optimal as possible.
- Ensure that those who are not being tested do not disturb the testing environment.

During the test

- Make sure that all needed materials are available.
- Sign out only the secure materials needed for that day of testing.
- Ensure that all personnel involved in the test administration adhere to test security guidelines. Report any breach of test security.
- Monitor the test administration by briefly visiting each testing room.
- Be available to answer questions that may arise.
- Return and sign in all secure test materials at the end of testing each day.

After the test

- Ensure that all examination materials are collected after testing.
- Label tests on which an examinee lost considerable time because of sickness or emergency.
- Record any testing irregularities, such as individual examinee sudden illness, writing careless answers, or group being interrupted or distracted.
- Return all test materials and equipment to a secure location.

Note. Summarized from Clemans (1971); Fitzpatrick and Morrison (1971); Massachusetts Department of Education (2005); Roeber (1996); and South Carolina Department of Education (2005a).

Another important role of the test administrator is ensuring the security of the test materials (AERA, APA, & NCME, 1999). Generally, a test site administrator is expected to confirm the contents of the packages arriving from the testing agency (NBME, 2003; SCDE, 2005a). If the contents of the shipment do not match the inventory, the test site administrator will alert the testing agency of the discrepancy. In addition, the test site administrator will also ensure that sufficient materials are available for the number of examinees scheduled for the examination. Some testing agencies permit "walk-on" examinees and ship more materials than needed for the registered examinees. Testing agencies generally expect that test site administrators will keep test materials in a secure location until the time of the examination (NBME, 2003; SCDE, 2005a). At the end of the examination, the test administrator collects the materials for return to the appropriate department or agency. Test materials should be stored in a secure place to prevent theft and protect the confidentiality of examinees.

An administrator manual includes information required for the training of examiners and proctors and descriptions of the test situations. Topics for the manual include:

1. The logic for the testing program
2. The organizational structure
3. Roles and responsibilities of staff members
4. Training materials, directions for administering each task
5. A list of test materials and equipment
6. Procedures for ensuring standardization
7. The handling of special problems, test security
8. Forms for keeping records (SCDE, 2001; Siegel, 1986)

The checklist of administrative activities in Figure 5.1 shows one type of form included in an administrators' manual for a state testing program (SCDE, 2005b).

Examiner Directions. Duties of the examiners include distributing the test, reading the script during the examination, monitoring examinees, and collecting the completed tasks (see Table 5.3). Examiners generally receive examination materials from the test site administrator shortly before the examination begins, and these materials are kept in a

- ____ District Test Coordinator notifies Test Administrators of the testing window.
- ____ Test Administrators publicize the testing window.
- ____ District Test Coordinator and Test Administrators ensure that schools have secure storage space, adequate testing rooms, and sufficient staff.
- ____ District Test Coordinator receives materials for district and schools.
- ____ District Test Coordinator signs test security agreement.
- ____ District Test Coordinator verifies receipt of shipment and faxes Receipt Acknowledgement Form to test company within 24 hours.
- ____ District Test Coordinator verifies only district overage materials and notifies the testing contractor if any materials are missing.
- ____ District Test Coordinator instructs Test Administrators to sign test security agreements.
- ____ District Test Coordinator trains Test Administrators.
- ____ District Test Coordinator distributes school boxes to Test Administrators.
- ____ Test Administrators receive and verify materials.
- ____ Test Administrators notify District Test Coordinator of any missing materials from school shipments.
- ____ District Test Coordinator notifies the testing contractor of any missing school materials.
- ____ Test Administrators store materials in secure location when not in use.
- ____ Test Administrators instruct Examiners to sign test security agreements.
- ____ Test Administrators train Examiners and Proctors and distribute manuals.
- ____ Test Administrators compile class packets.
- ____ Test Administrators distribute class packets to Examiners on testing day.
- ____ Examiners initial security checklists and receive materials.
- ____ Examiners administer tests to students on testing day.
- ____ District Test Coordinator and Test Administrators oversee test administration, maintain test security, complete Testing Irregularity Forms (if needed), and answer questions.
- ____ Examiners collect and count materials after testing and before dismissing students.

(continued)

FIGURE 5.1. Example of a checklist of test administrative activities in a state assessment administered in schools. Adapted with permission from the South Carolina Department of Education (2005b).

- ____ Examiners return materials and make-up rosters to Test Administrators immediately after testing.
- ____ Examiners initial security checklists for return of materials.
- ____ Test Administrators collect and account for all materials.
- ____ Examiners give make-up tests, receive materials, and initial the security checklists on testing day.
- ____ Examiners administer make-up tests.
- ____ Examiners collect and count all make-up materials after testing and before dismissing students.
- ____ Test Administrators return scorable materials to District Test Coordinator.
- ____ District Test Coordinator schedules return of scorable materials to the testing contractor.
- ____ Test Administrators return nonscorable materials to District Test Coordinator.
- ____ District Test Coordinator takes inventory of all school shipments and returns nonscorable materials to the testing contractor.

FIGURE 5.1. *(continued)*

safe location under the supervision of the examiner until all examinees are admitted to the testing room. With the admission process complete, the test examiner will distribute the examination materials as protocols require, recording which examinee received which materials.

The examiners' manual should include a list of materials and equipment. Types of materials listed include examinee information sheets; a do-not-disturb testing sign; test books; readers, dictionaries, and thesauruses for language arts tests; protractors, rulers, and calculators for mathematics tests; the examiner's manual; pencils; and a clock or stopwatch (Clemans, 1971; SCDE, 2005a). A checklist in the examiner's manual can summarize the materials.

A key component of the examiners' manual is a script that guides the process of the testing. The instructional script read by the examiner to the examinees is designed to ensure that all examinees, regardless of administration or examiner, receive the same, or at least a highly similar, examination experience. Personal differences in the examiner and physical differences in examination sites should be minor and contribute nothing substantial to the performance of one examinee over another. Deviations from the script without compelling reason (e.g.,

TABLE 5.3. Typical Responsibilities of Examiners

Prior to the test

- Review the testing schedule, administration directions, test booklet, and answer sheet.
- Tell proctors their duties.
- Learn permissible answers to examinees' inquiries.
- Learn her role if she serves as an actor in a simulation.
- Learn the scoring criteria if he is an observer in a simulation.
- Rehearse administering the test.
- Collect and count the materials required for the testing.
- Make arrangements so there will be no disturbances.

During the test

- Administer pretest if appropriate.
- Read instructions slowly, clearly, and loud enough to be heard by all examinees.
- Use exact wording of the directions to standardize testing conditions for all examinees.
- Time the examination.
- Monitor whether examinees' responses are written in correct section.
- Prevent talking or sharing answers.
- Provide assistance only in the mechanics of the test.
- Avoid coaching examinees.
- Monitor that examinees use only the allowable supplemental materials specified in the administrative manual.
- Announce when a half-hour remains for tests of more than an hour.
- Inform the administrator if any problem occurs.
- Help maintain test security by moving about the room and monitoring unusual examinee behaviors.

After the test

- Label tests for which examinees lost considerable time due to sickness or emergency.
- Record any testing irregularities, such as an individual examinee's sudden illness, an individual writing careless answers, or the group of examinees being interrupted or distracted.
- Have proctors collect test materials, count the materials, arrange the stack in serial order number or alphabetically, and check against a list.

Note. Summarized from Clemans (1971); Fitzpatrick and Morrison (1971); and South Carolina Department of Education (2005a).

ADA compliance) are discouraged. Adherence to the script limits the potential that one examinee will experience testing conditions that could have an unintended and unanticipated influence on the examination outcomes.

In some scripts the directions that are to be read to the examinees are highlighted by using a different font format, such as bold (SCDE, 2005a, 2005b). Also, as seen in Figure 5.2, examiners' manuals use graphics and shading to highlight the information that should be read to examinees (SCDE, 2005a, 2005b). Directions meant only for the examiner should use a plain font. Included in these directions would be instructions about where to pause and the speed with which the examiner should read (Clemans, 1971).

In testing examinees, at the beginning of the test session the examiner should announce the test, its purpose, and, in the case of young students, the importance of examinees applying themselves (Clemans, 1971). To ensure that everyone has an understanding of the directions, the test examiner should read the directions aloud as the examinees read

DIRECTIONS
This test is divided into two sessions: 1 and 2. You will take Session 1 today and Session 2 tomorrow. Today's session contains one extended-response question asking you to write a composition, followed by reading selections with multiple-choice questions.

Write your composition and mark your answers to the multiple-choice questions directly in your test booklet. Use only a number two pencil to write your composition and mark your answer choices.

WRITING PROMPT

- **Read the prompt carefully before you begin to write.**
- **Use a dictionary and thesaurus to write your composition.**
- **Be sure your composition addresses all parts of the prompt.**
- **Refer to the checklist below the prompt with the features of good writing.**
  Do your prewriting on the separate, lined scratch paper provided—your prewriting WILL NOT be scored.
- **Allow enough time to write your FINAL composition in the test booklet on the three lined pages marked "Writing" at the top.**

FIGURE 5.2. An example of the type of formatting used to highlight instructions that are to be read to the examinees (South Carolina Department of Education, 2005b).

silently. When a change in item type occurs, a new set of directions should occur and the directions should be read aloud. The inclusion of sample questions can help clarify instructions.

Directions to examiners should also specify what they can and cannot do. For example, are examiners allowed to help student examinees pronounce words in a set of science task directions? Instructions should also clarify for examinees the focus of the task. An example would be for an examiner to instruct examinees who are completing a writing assessment that it is important to make sure their essay has a clear message and that mistakes in spelling and punctuation do not interfere with their message.

If the test is timed, a place to record the beginning and ending time should be included in the examiner's manual. The examiner's directions should be to tell the examinee the amount of time for a particular section of a test. For sections of an examination that last more than an hour, the directions to the examiner should indicate that she announce when a half-hour remains (Clemans, 1971). Other intervals may be used as long as the timing of the announcements is part of the examiner directions.

Test examiners are trained to begin an examination on time. Deviations from the starting time occasionally arise when problems occur with admitting candidates into the examination room. In addition, late-arriving examinees can disturb the concentration of the other examinees. For this reason some testing agencies do not permit examinees to enter the examination room after the official beginning of the test session, although some agencies do permit a grace period (e.g., 30 minutes) to accommodate examinees who have experienced minor trouble (e.g., traffic) reaching the examination site. Generally, late-arriving examinees who are permitted to enter the examination room are not permitted additional time to complete the examination (NBME, 2003).

It is the responsibility of the test examiner to make sure that all materials are returned, counted, and filed (NBME, 2003). When the examination is finished the examinees are usually permitted to turn in all materials and leave the testing room. In some instances examinees leave the room as a group, and it is the test examiner's responsibility to ensure that all materials are collected and secure before releasing the examinees. When the inventory of test materials indicates missing components, some testing agencies do not permit examiners to release the examinees until the testing agency itself grants the examiners permission to release the examinees (NBME, 2003). In such an instance, examinees leaving the room without the permission of the examiners expose

themselves to consequences ranging from nullified examination results to suspicion of theft.

Test examiners generally return examination materials to the test administrator immediately after the examination. Both examiner and test site administrator complete an inventory of the materials and agree on the counts to ensure that all materials are still secure (NBME, 2003). Shortly after the examination, the test administrator returns the materials to the testing agency. Representatives at the testing agency review and inventory the materials returned, contacting the test administrator in the event of a discrepancy.

Proctor/Monitor Directions. Table 5.4 lists the responsibilities of proctors or monitors. Proctors assist the examiner in distributing test materials and equipment, monitoring examinees during the test, and collecting materials after the test. A key role for the proctor is to monitor examinee behavior to forestall any attempts at cheating or copying the content of the test (Dows, 2005; Newton, 2005). One proctor can monitor 15–25 examinees (e.g., SCDE, 2005a).

Other than the instructions in the script, during the examination, proctors' responses to examinees are generally limited to requests to be excused from the examination room to visit the bathroom, although other personal circumstances can arise. Testing agencies rarely authorize the proctors to answer direct questions from examinees regarding the examination. Examinees with questions regarding examination content are usually encouraged to write such questions on a comment form that is returned to the testing agency with the other examination materials. To permit proctors to discuss examination content with individual examinees could compromise the standardization of the examination administration.

Actor and Observer Directions. In some simulations and performance assessments, the examinee interacts with actors who play a role in the task. For instance, the United States Medical License Examination includes a performance task in which the examinees interact with a standardized patient (FSMBUS & NCBE, 2008). The actor who plays the patient must understand her role and follow the instructions closely (Fitzpatrick & Morrison, 1971). Also, in some performances, an observer scores the performance as it occurs. Just as it is crucial for the actor to follow her script, the observer must understand the scoring criteria and rules. The scoring of performances will be addressed in subsequent chapters.

TABLE 5.4. Typical Responsibilities of Proctors/Monitors

Prior to the test

- Learn the testing schedule.
- Read the administration directions.
- Help prepare the room.
- Fill out student information sheets.
- Retrieve additional test materials from the test administrator if needed prior to or during testing.
- Make sure each examinee has the test booklet with his/her identification information.

During the test

- Monitor whether responses are written in correct place.
- Discourage talking or sharing answers.
- Provide assistance only in the mechanics of the test.
- Avoid coaching examinees.
- Monitor that examinees use only the allowable supplemental materials specified in the administrative manual.
- Inform the administrator if any problem occurs.
- Help maintain test security by monitoring unusual behaviors, such as an examinee being unusually nervous, sitting in unusual positions, finishing the examination very quickly, or spending an inordinate amount of time on a few questions.
- Check restrooms before allowing examinees to enter to make sure the area is free of reference materials (e.g., books, notes).
- Serve as a messenger during testing if an emergency occurs.

After the test

- Label tests for which examinees lost considerable time due to sickness or emergency.
- Record any testing irregularities, such as individual examinee's sudden illness, an individual writing careless answers, or the group of examinees being interrupted or distracted.
- Help collect tests at the end of testing.
- Collect equipment (e.g., rulers, protractors) used in the assessment.

Note. Summarized from Cohen and Wollack (2006) and South Carolina Department of Education (2005a).

Training of Staff

The development of staff administration manuals and examinee directions is for naught if staff is not trained in the use of the materials. Handing a staff member a manual and expecting him or her to self-train is unreasonable. The importance of trained staff is reflected in the statement of Bond and colleagues (1996) that "the selection and training of administrators and scorers are critical features of the overall validity of the assessment" and "attention must be paid to administrators and scor-

ers, and their biases" (p. 121). Staff training should provide explanations of all procedures and practice in the administration of the training and testing situations (Siegel, 1986).

Test administrators must conduct training sessions for all examiners, including possible substitute examiners and proctors (SCDE, 2005b). Training should also include examiners responsible for customized administrations, such as accommodations that require individual administration of the test. Topics for the training of examiners include the logistics of administering the test, the directions for completing any examinee information, the script for administration, security of the test material, distribution and collection of test material, irregular behavior, questions from examinees, restroom break policy, emergency situations, specific duties for each proctor, and staff biases (Bond et al., 1996; NBME, 2003; SCDE, 2005b). Training sessions also should stress the need to account for all materials before, during, and after testing.

Training should also address the uses of any manipulatives and equipment required for any performance task. In some instances the examiner may also serve as the actor who plays a role in the simulation and the observer who records the performance (Fitzpatrick & Morrison, 1971). If the examiners are also observers or actors in the task, then their training will address the recording and scoring of responses (Roeber, 1996). Those with acceptable scoring levels of accuracy will be certified as examiners, whereas others must receive additional training or be dismissed from the assessment project. We present additional scoring guidelines in the following chapters.

Testing Environment

Standardization requires a testing environment that provides reasonable comfort and that avoids noise, disruption during testing, inadequate lighting, limited work space, and poorly reproduced materials (AERA, APA, & NCME, 1999). Also, to keep the environment conducive for testing, some testing agencies prohibit intrusive equipment, such as cell phones, pagers, beepers, calculators that "talk," radios, and food (NBME, 2003; SCDE, 2005b).

In terms of performance assessment in elementary or secondary schools, Haertel and Linn (1996) note that challenges to comparability include the accuracy of timing due to variation in equipment setup and cleanup; the number of students in a class; the size of the room and configuration of facilities (e.g., desks, sink, electrical outlets); and the demands made on the test administrator to maintain order and provide

BOX 5.1. ADMINISTERING A LICENSURE EXAMINATION

Scott is the vice president for operations at a company that develops and delivers examinations for use in certification and licensure. He manages the delivery for very small and moderately large testing organizations, where "very small" might mean as few as a dozen examinations per year, whereas "moderately large" might mean more than 10,000 examinations per year. We present his story in chronological order, beginning with the signed contract that specifies a date and location (and probably more than one location) of an examination, ending with the return of testing materials and scoring of examinee results.

Several Months before the Examination

When Scott receives notice that an examination is scheduled, his first action is to contract with an organization (e.g., hotel, college) in the city where the test is to be given, after determining that the available space meets the requirements of the examination. Many exams occur in multiple cities on the same day. Some exams require little more than tables and chairs with sufficient space (e.g., 3 to 4 feet) between to ensure privacy of the examinee. Other exams require an extensive setup for materials used during the testing. (Scott tells the story of one organization that rented an entire hotel in downtown Chicago to provide sufficient space for the examination while maintaining adequate security of test materials and separation of the examinees.)

In addition to securing a site suitable for the administration of the examination, Scott identifies test site administrators from a list he maintains of people trained and available for this work. The test site administrators, in turn, identify test examiners. Because Scott manages the administration of exams around the world, his list of supervisors is extensive. In all cases, the people on the contact list have passed training in the protocols of test site administration as defined by Scott's company.

Ten Weeks before the Exam

Because Scott manages the delivery of nearly 150 exam titles each year, the preparation for a given administration is heavily scripted, and project management is an important skill for Scott to have and use. Even if the exam preparation involves little more than the copying of printed materials, that printing begins 10 weeks before the delivery of the examination. If the examination involves the use of other materials, Scott must identify the vendors for those materials and make the appropriate arrangements. Such arrangements could be as mundane as the purchase of 1,000 four-function calculators (solar powered, not battery). However, those arrangements could be more complex, such as the instance where an examination required the procurement of 43 severed horse heads.

In addition to ordering the materials for the examination, Scott also wraps up the loose ends regarding the test sites and the test site administrators at the "10 week prior" window. Occasionally, letters are lost in the mail and the contracts for examination sites need to be re-sent. Other times, test site administrators need a nudge to return agreements for the examination and confirmations of adequate numbers of test examiners. Regardless, these matters must be resolved well in advance of the examination.

Six Weeks before the Examination

Six weeks before the examination, Scott reviews the information regarding each site. This information includes the space and facilities available, the shipping information for each site, and the recent performance of the test site administrators and examiners. If he finds that a site has become unavailable, he still has time, but not much, to find a new one. He also has time to adjust space rentals and administrator contracts to reflect changes in the scale (e.g., number of examinees) of the examination.

In addition, he knows at this point what types of materials he will be shipping to the sites, and although the convergence of shipping and sites can seem somewhat pragmatic, there are times when the consideration of what is going where is important. For example in one instance he received a surprise call the day before an examination from a test site administrator indicating that no materials had arrived, and the follow-up investigation determined that the shipment was being held in Chinese customs.

Because there is very little that Scott can do to expedite the passage of several boxes of printed materials across the Chinese border, in this instance he decided to break with security protocols and fax a copy of the exam and answer sheets to the site administrator, who in turn made the necessary copies. (Materials such as the severed horse heads would most likely be procured locally for a foreign administration.)

logistical support. To the degree possible, such factors require standardization. As an example, the United States Medical License Examination includes a standardized patient performance assessment referred to as Step 2 Clinical Skills. In order to standardize the assessment, the examinee must take the examination at one of five Clinical Skills Evaluation Centers scattered throughout the United States (FSMBUS & NBME, 2004). In this assessment the examinees ask the standardized patient questions and perform a physical examination to gather information to develop a preliminary diagnosis and a diagnostic work-up plan. The developers of this complex assessment have introduced uniformity to the assessment by training actors to portray a patient with a clinical problem. The actors (i.e., standardized patients) are trained so that all examinees receive the same information when they ask the standard-

ized patient the same or similar questions. The environment is standardized by providing each examinee with a simulated medical examination room in which the same equipment is available in all rooms (FSMBUS & NBME, 2005). In addition, proctors are responsible for attending to malfunctioning equipment.

Test Sites for Computer-Based and Internet-Delivered Examinations

As pencil-and-paper examinations change to a degree when moved to electronic form, so do the test sites. Test environments now have computers on the desks, and those computers have some form of network access appropriate for the application. In addition, those computers have software available for use with the examination. If the examination includes the use of a printer, then a printer will be appropriately available. Earphones must be available with each computer if the examination makes use of audio and microphones for recording voice (SCAAP, 2006). Test staff must have instructions about the use of the equipment and software. This special equipment requires staff to understand the appropriate procedures to follow. As shown in Figure 5.3, SCAAP (2006) uses the screen capture utility in Windows® to write directions that guide staff through the administration and recording of the music performance tasks.

Other aspects of the test site are unchanged. The examinees must be sufficiently separated to preclude interference and to maintain exami-

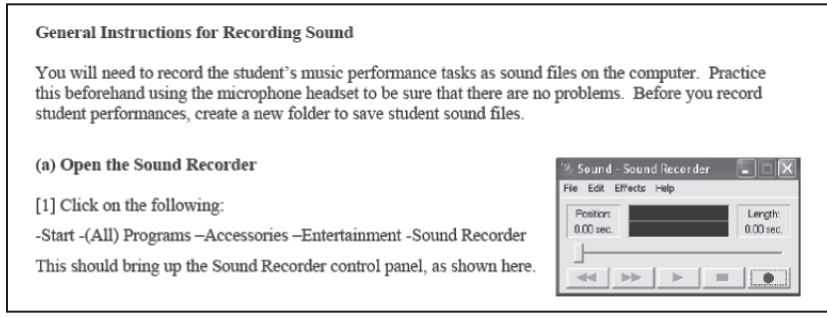

FIGURE 5.3. An example of the use of graphics to guide test examiners through the steps in the recording of an examinees' music performance in a computer-based assessment.

nation security (SCAAP, 2006). The room must have adequate lighting and ventilation. The room must be quiet with reasonable accessibility to water and restrooms. At least one proctor must be in the room, although the number of proctors required depends on factors such as room size, examination complexity, and the age and number of examinees.

Accommodations

Testing accommodations are required when standardized forms "of presenting information or of responding may not be suitable for specific individuals, such as persons with some kinds of disability, or persons with limited proficiency in the language of the test, so that accommodations may be needed" (AERA, APA, & NCME, 1999, p. 61). Accommodations in testing is a result of the Americans with Disabilities Act (ADA) passed by Congress in 1990 to prohibit discrimination against persons with disabilities (Duhl & Duhl, 2004; Sireci, 2005). The ADA protects examinees with physical and mental disabilities, who are otherwise qualified (i.e., the examinee "meets the essential eligibility requirements for the receipt of services or the participation in programs or activities provided by a public entity" [ADA, 1990, Section 201]) by extending reasonable accommodations to level the playing field relative to nondisabled examinees, without giving an unfair advantage to those with disabilities.

Recall that earlier we said the goal of standardizing test administration is to "provide accurate and comparable measurement for everyone, and unfair advantage to no one" (AERA, APA, & NCME, 1999, p. 61). Standardization requires that everyone receives the same examination (or a reasonably parallel form). Everyone receives the same time, materials, instruction, and environment. The idea is that the only factor contributing to the performance of an examinee on the examination is his or her standing on the knowledge and skills (i.e., construct) that the examination measures; all other factors are equivalent for all examinees.

It may appear that accommodation by its very nature compromises, if not negates, the standardization of the examination. Why? Every instance of accommodation represents an exception permitted to one examinee that was not made available to all examinees. However, legal opinion holds that accommodations made to standardized conditions only create the conditions under which examinees with disabilities, but who are otherwise qualified, can compete fairly with examinees without disabilities (Duhl & Duhl, 2004).

AERA, APA, and NCME (1999) recommend that examinees should be informed prior to testing about procedures for requesting and receiv-

ing accommodations. Such a procedure is followed for the sponsoring organizations for the Graduate Record Examination (ETS, 2005) and the United States Medical Licensing Examination (FSMBUS & NBME, 2007), which have information about test accommodations available on their websites.

In education, the appropriateness of providing test accommodations is determined by state policy (e.g., SCDE, 2005a). Accommodations are appropriate when specified in a student's individualized education plan (IEP). The IEP is a document that stipulates the instructional goals and objectives for a student with special needs, in addition to the appropriate forms of testing. A state policy might also stipulate accommodations for examinees with limited English proficiency.

In licensure, testing agencies generally evaluate particular requests for accommodations on a case-by-case basis (Duhl & Duhl, 2004). Any request for accommodation should be supported by medical opinion (Duhl & Duhl, 2004). If an agency plans to decline an accommodation request, then it should have a medical expert review the case. Legal challenges to requests denied will require the agency to show the request is unreasonable based on efficiency, cost, feasibility, and test validity. Duhl and Duhl (2004) indicate that to this point courts have neither required agencies to provide accommodations that threaten test security for licensure examinations nor required the lowering of pass scores or waiving of the licensure examination—in this instance, in the bar examination.

What are common accommodations? They include large print, magnifying devices, Braille, audiotapes, word processing devices, readers, transcribers, voice recognition devices, increased rest periods between examination segments, and private rooms to take the examination while removed from the distractions produced by a room full of other examinees (Cohen & Wollack, 2006; Duhl & Duhl, 2004; Massachusetts Department of Education, 2005; Sireci, 2005). Accommodations also include increased time limits for testing—generally allowing 1.5 or 2 times the standard time (Cohen & Wollack, 2006). Additional accommodations used for students in the National Assessment of Educational Progress have included one-on-one testing, bilingual dictionaries, and small-group testing (O'Sullivan et al., 2003; Persky et al., 2003).

Appropriate accommodations should be provided for computer-based and Internet-delivered examinations (International Test Commission, 2006). The International Guidelines on Computer-Based and Internet Delivered Testing provide guidance for addressing accommodations in the design and development of these tests.

In the past, many testing agencies have flagged examination results achieved under accommodations, most often by placing a mark (e.g., an asterisk) by the examinee's score on the test report. Although such practice is congruent with the standards described in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), it is considered discriminatory by many people. To avoid the legal consequences of reporting examination results in a manner that many perceive as discriminatory, many testing agencies have stopped flagging such scores (Sireci, 2005).

BOX 5.2. ADMINISTERING A LICENSURE EXAMINATION
(*continued*)

Four Weeks before the Exam

One month before administration of the examination, Scott calls each test site administrator to confirm the room reservations and the recruitment of proctors. If changes must be made, there is still time to do so, depending on the magnitude of the change. In addition, Scott confirms the accuracy of the addresses he has for the shipments of materials to the test site administrators and the site directions he will send to the examinees.

It is now, 1 month out from the examination, when Scott receives official notification of ADA requirements that he must accommodate. These accommodations can take many forms. Sometimes Scott must arrange for printed materials to be larger. Most often, he has to arrange for a longer examination period, usually time and a half or double time. Occasionally he will need to arrange a separate testing environment for examinees needing to take the examination without having other candidates in the vicinity.

Three Weeks before the Exam

At 3 weeks before the examination, Scott has made the necessary arrangements for the requests for ADA accommodations that he received. Examinations with larger type are being produced. A reader has been approved in response to an examinee's request for such an accommodation. A special room and separate proctor will be available for another examinee. The reservation for a test site and contract with a proctor has been extended to permit a longer administration.

Scott has informed the examinees of the exam location and provided suggestions about preparations for taking the examination. These suggestions include dressing in layers to accommodate personal taste in room temperature, bringing earplugs if ambient noises present a distraction, and other suggestions designed to make the test-taking experience a little less difficult.

He also sends final registration information to the test site administrators, who in turn might have to arrange additional space at some locations

because of ADA accommodations or increased attendance. Finally he sends notice to his shipping department that they can expect to ship materials for the examination the following week. (He might send this notice earlier if international shipments were included.)

Two Weeks before the Exam

Two weeks before the examination administration, Scott checks that all materials have been shipped to the correct places. He checks with each test site administrator to confirm reservations, proctors, and ADA accommodations. Each site is rechecked for conditions such as adequate space, tables, chairs, and clocks. He also confirms that the facility will open early enough for the exam and that the examination space is sufficiently removed from other large activities (e.g., wedding receptions, cheerleader practice) that might interfere with the examination.

One Week before the Exam

Seven days before the examination, Scott sends the final updates to the test rosters to the test site administrators. The administrators confirm that the conditions of the examination sites remain adequate. Occasionally, local construction, weather, and unexpected events necessitate rapid communication with the examinees and sometimes the selection of a replacement examination site. (More than once, Scott has had to arrange alternative examination sites because of inclement weather on very short notice. Once, "short notice" involved 12 hours and a Boston snowstorm.)

In the communication with the test site administrators, Scott reminds each that all examination materials must be returned within 24 hours of the exam. He goes over special considerations appropriate for both the examination and the examinees. Some examinations permit "walk-on" examinees; others do not. Some examinations require the distribution of materials at particular times. One of the examinations that Scott delivers arrives at the examination site in an armored truck under armed guard, and Scott makes certain that the test site administrators know how to receive and return examination materials from those guards.

Security

Test security is a final factor in test administration that supports comparability of scores (Haertel & Linn, 1996). Test security helps to assure that no examinees have an unfair advantage (AERA, APA, & NCME, 1999). Historically, the issue of unfairness did not stop examinees from attempting to have an advantage in the civil service examinations in China more than 1,000 years ago (Cizek, 2001a). To prepare for the civil

service examination, some examinees wrote on the fabric of an undergarment some 722 responses to potential examination questions. Today, as we shall see, methods for cheating have proliferated and become more sophisticated.

Examination security during administration, much as other secure activities, involves many details that are generally not visible to the public. In addition, the security of the examination begins long before the examinees reach the room and continues long after they depart. Trusted personnel reproduce test booklets in secure areas and gather ancillary materials (e.g., calculators) that are in turn numbered and stored in a secure area. Generally, information regarding the numbers and types of materials that are gathered for an examination are considered as secure as the examination itself because a clever examinee with some information regarding the equipment and materials likely to appear in the examination could use that information to build an unfair advantage over the other examinees.

The loss of an examination can involve a great deal more than the recovery of the exam or regeneration of analogous content (Lamb, 2001). Many examinations are linked from one version to the next by the use of a subset of items common to both versions. These common items are used to equate statistically the scores across examination versions to ensure that all examinees are held to the same passing standard (i.e., the minimum score required to pass an examination), regardless of the difficulty of a particular examination version.

If the content of an examination that contains common items is made public, then the testing agency responsible for replacing the lost content is also responsible for creating a new passing standard. Although care can be taken when creating a new passing standard, the new standard will at best only be comparable to the old; it will not be equivalent. To this end, subsequent examinees will be held to a new passing standard that only approximates the old standard. Granted, all passing standards undergo review from time to time, but that activity is better undertaken as a part of the formal assessment process rather than as a response to a compromised examination.

One challenge in the use of performance assessments is the examinees' ability to remember the items and discuss them with others. Contributing to this is the novelty of the tasks, the use of a smaller number of such tasks, requirements for special equipment, and greater examinee interest and involvement (Haertel & Linn, 1996). Haertel and Linn (1996) suggest limiting the exposure of tasks by administering a task to

different, randomly equivalent groups of examinees in successive years. Such a method would be feasible if the focus is on overall trends as opposed to measurement of individuals. However, in licensure and certification decisions it is the examinee who is of interest.

Preexamination Security

During the development of the test, breaches of security may occur. The Pennsylvania Board of Law Examiners implemented procedures to maintain security of test materials for their bar examination (Dows, 2005). The agency prints, packages, and stores essay materials in-house and stores the Multistate Bar Examination (MBE; NCBE, 2006) at the site. To protect these materials, the Board office in Pennsylvania uses key-card access and digital cameras to electronically monitor entrances and record visitors (Dows, 2005). In addition, visitors must be accompanied by staff.

Another threat to security occurs in the drafting and review of essays (Dows, 2005). Drafts of essay questions for the Pennsylvania bar examination are developed by Board members and stored on out-of-house computer networks. Although the files were password protected, they were vulnerable to hackers. Also, possible exposure of files stored on disks, paper, or memory sticks posed a problem. The remedy was to buy a laptop for each examiner and train them to use encryption and passwords when electronically transferring files for review.

To protect test security, some agencies have agreement forms that must be signed by those who come in contact with items during development or the test during administration. The form shown in Figure 5.4 is signed by anyone who has access to secure testing materials. The form describes their responsibility to follow the security procedures outlined in the agreement and take the necessary precautions to ensure test security at all times (SCDE, 2005b).

The loss of a test poses security threats. Although materials are counted and shipped with tracking, they are recounted upon receipt to ensure (1) that an examinee's materials have not been lost, and (2) that no secure materials turn up in public. Nonetheless, testing is a human process, and materials are occasionally lost despite best practices.

When the loss is a matter of a tracked package lost in transit, it will usually reappear at some point. Occasionally, the package is gone forever, and at that point the testing officials are faced with the decision of declaring the examination "compromised." If the package was lost in transit to the examination site and there is no replacement examination,

STATE OF SOUTH CAROLINA
DEPARTMENT OF EDUCATION
Office of Assessment
Agreement to Maintain Test Security and Confidentiality
For District Test Coordinators, School Test Coordinators, and Test Administrators

Test security is essential to obtain reliable and valid scores for accountability purposes. Accordingly, the Department of Education must take every step to assure the security and confidentiality of the state test materials. It is the responsibility of those individuals who serve as test coordinators, test administrators, and monitors; those who handle test materials; and/or those who use the results to follow test security laws, regulations, and procedures. The test administration manual for each test provides detailed test security information and administration guidelines. District Test Coordinators (DTCs), School Test Coordinators (STCs), and Test Administrators (TAs) are expected to read and follow the instructions provided in these manuals.

To help all personnel involved in testing have a common understanding of test security and appropriate testing practices, District and School Test Coordinators must provide appropriate training for these individuals. Test administrators, monitors, and/or other individuals who assist in the distribution and packing of test materials must be familiar with test security laws, regulations, and procedures, as well as with their responsibilities for each test. Test administration manuals and/or other appropriate materials should be distributed to these personnel at least one week prior to the testing window. DTCs and STCs must review test security policies and procedures with them and must encourage them to read all appropriate materials.

I acknowledge that I will have access to one or more of the following tests that are a part of the South Carolina statewide testing program: BSAP, End-of-Course Examination Program (EOCEP), High School Assessment Program (HSAP), HSAP-Alt, PACT, or PACT-Alt. I acknowledge that I have or will provide appropriate training for all individuals involved in administering or monitoring the tests and/or handling test materials.

(continued)

FIGURE 5.4. Example of a test security agreement for staff who handle test materials. Reprinted with permission from the South Carolina Department of Education.

I understand the tests are secure, confidential, and proprietary documents owned by the Department of Education. I hereby agree that I will not discuss, disseminate, describe, or otherwise reveal the contents of the tests to anyone. I will not keep, copy, reproduce, or use in any manner inconsistent with the instructions provided by or through the State Department of Education any test, test question, or specific test content. I will not keep, copy, or reproduce in any manner inconsistent with the instructions provided by or through the State Department of Education any portion of examinee responses to any item or any section of a secure test, secure administration manual, oral administration script, or any other secure test materials. I will return all test materials on time to the appropriate person or place (i.e., to the District Test Coordinator for School Test Coordinators; to the testing contractor for District Test Coordinators). I will follow all of the state laws and regulations regarding testing ethics and test security.

I understand that failure to follow these laws, regulations, or procedures could result in action being taken against my certificate and/or criminal prosecution.

Signature Print name

District and School Date

1429 SENATE STREET COLUMBIA, SOUTH CAROLINA 29201
(803) 734-8492 FAX (803) 734-8624
HTTP://ED.SC.GOV/

FIGURE 5.4. *(continued)*

the testing officials find themselves in the position of having to cancel the administration, or at least postponing it until a replacement examination is available.

It is of interest to note here that the insurance provided by most common carriers applies only to the physical contents of the package. That is to say, the insurance will pay to replace the paper in the box, not the writing on the paper. Replacement of the content of the examination falls to the testing organization or its assigned vendor.

To ensure the appropriate person takes the examination, security should require identification (AERA, APA, & NCME, 1999). An admission letter and proper photo identification are generally required to enter the examination room. Examples of acceptable photo identification are usually a current driver's license, a current passport, or another current government-issued identification that includes a picture (e.g., military I.D.) (FSMBUS & NBME, 2008). Examples of unacceptable identifica-

tion often seen at examination sites are expired forms of acceptable identification, gym membership cards, club membership cards, and other cards with a signature only (i.e., no identifying photograph). The examiner or proctor will compare the card picture to the face of the person in addition to comparing the card signature to the check-in signature of the person before assigning a seat or position to the examinee. Additional security measures used by test agencies include an admissions ticket, fingerprinting, identification badges, video cameras, and metal detectors (Dows, 2005).

Examination Security

Security during the examination includes limiting the items that are permitted into the examination room. Test booklets, answer sheets, simulation materials, pencils, erasers, and admission letters are generally allowed. The personal effects of the examinees are not permitted, although there is some discretion available to proctors and test site administrators. Calculators, especially those without alphabetic keyboards, might be permitted with some examinations. Generally not permitted are cell phones, pagers, cameras, and reference materials (e.g., books, notes, papers). Food and beverage are typically not permitted, although some exceptions are permitted if an examinee has sufficient reason (e.g., diabetic) to request an exception.

Other types of equipment not allowed include electronic imaging or photographic devices, personal digital assistants (e.g. Palm Pilots), headphones, some forms of calculators, digital watches, paging devices, recording/filming devices, electronic translators, briefcases, coats, or brimmed hats (Dows, 2005; NBME, 2003; SCDE, 2005b). Threats to security also include the visual and audio capture of test material through digital cameras disguised as lighters and working pens (Newton, 2005).

To reduce cheating and protect the integrity of test scores, seating charts and space between seats are recommended (AERA, APA, & NCME, 1999; NBME, 2003). Also, monitoring by the test examiner and proctor during the examination reduces the likelihood of cheating (AERA, APA, & NCME, 1999; Newton, 2005). The examiner and proctor should monitor examinees for being unusually nervous, sitting in unusual positions, finishing the examination very quickly, spending an inordinate amount of time on a few questions, or leaving the seat frequently to sharpen a pencil or to go to the bathroom (Cohen & Wollack, 2006).

Materials are generally tracked during the examination to ensure that they remain in the possession of the examinee or in the custody of the test examiner. Test materials are generally not shared between examinees because the use of such materials by one examinee can leave hints for the next (e.g., pencil marks left on tables of trigonometric values). In addition, examinees are generally not allowed to take materials from the view of the test examiner. In the event that an examinee must leave the examination room, the test examiner is usually responsible for gathering and retaining such materials until the examinee returns, say, from the bathroom (NBME, 2003). At the end of the examination the examinee surrenders all materials to the test examiner.

High-tech devices for cheating possibly provide information for the examinee during testing (Newton, 2005). During testing, wireless-fidelity equipment can be used to communicate with outside sources. Possible sources of information would be websites or text messaging with someone outside the room. Access to sources also occurs through cell phones, personal digital assistants, and wristwatches. Also, information stored on MP3 players and accessed through headphones or ear buds could be used during the examination. Devices can be bought either to detect or disrupt electronic devices; however, security can best be supported by proctors attending to examinee behavior and use of the equipment.

Best practice dictates that at least one test examiner remain in the examination room at all times (NBME, 2003), although the number of test examiners required is driven by the number of examinees and the policy of the testing organization (e.g., one test examiner for every 20 or 30 examinees). The policy of the testing organization will also dictate the degree of security found in the areas surrounding the testing room. Some organizations will post security personnel in the halls leading to the testing rooms and at the doors to the bathrooms; others will post security personnel in the bathrooms. Some organizations also place security at the building exits and the surrounding grounds to preclude the unauthorized transfer of information both to and from examinees.

ADA accommodations also present challenges to test security during administration in that test materials are forgotten and left in rooms (Lamb, 2001). Also, staggered start and stop times allow possible exchange of information about test content. In the bar examination for Pennsylvania, examinees with accommodations complete an affidavit indicating they did not discuss the exam with other examinees (Dows, 2005). Extra staff monitor the activities and assist with any specialized equipment.

Postexamination Security

After the test is over the usual protocol is for the examinee to give all materials to the test examiner, sign out on the test registry, and then leave the room quietly. Most testing organizations provide notice of copyright on the test materials and then remind examinees not to discuss the assessment with others. However, it is unlikely that any reasonable testing organization is going to engage secret police to monitor the postexamination discussions of examinees; the intent is to prevent examinees from making examination content available in public forums such as Internet posting and other publications. In fact, some organizations make use of Web crawlers for the very purpose of identifying sources of secure examination content on the Internet. Other organizations use statistical procedures to monitor item and examination performance in different populations to identify examination content that might no longer be secure.

Another potential problem is the loss of confidentiality if an examinee's completed examination is viewed by anyone other than test staff or scoring agency members with the appropriate authorization. In addition, protection of examination materials is required during scoring (Dows, 2005).

Prior to a cheating incident, a policy that defines cheating should be established in order for an agency to pursue and prosecute an alleged cheater (Dows, 2005). An agency should have due-process procedures established. Penalties also should be established, such as sanctions or criminal prosecution. On the report of a cheating incident an immediate investigation should be conducted. In gathering evidence, potential sources about cheating include examinees, as well as staff members and proctors who observed the behavior. Members of licensure boards can be interviewed if the breach is in the development or scoring stage. In the instance of the Multistate Bar Examination, a security director flies to the location of the security breach and interviews staff and examinees (Lamb, 2001).

Test Security during Computer Administration

The general elements of security do not change when the form changes from "classical" delivery to delivery by computer, assuming the examination is appropriate for electronic delivery. However, special considerations for computerized delivery exist that do not arise in classical delivery, or at least they arise in a different form.

Forms of Computerized Delivery. Two forms of networked computerized delivery are often used in testing. The difference between the two involves the manner in which the examination reaches the computer. The up-and-coming manner of delivery makes use of the Internet to move an examination from a server to a computer and an examinee. An example is the South Carolina Arts Assessment Program (SCAAP), in which students complete a Web-based assessment with both multiple-choice items and performance tasks. In the performance assessment component of SCAAP, a software program is used to record students' voices as they perform a familiar song (Yap et al., 2005). After testing is completed, each student's performance is downloaded and is ready for electronic scoring. Because the Internet is publicly accessible, steps must be taken to ensure that the content of the examination is not compromised as it travels between the server and the computer.

An alternative to the Internet is a private intranet that moves information between the examinee's computer and the testing agency's server. Intranets can be made as secure as necessary, but they must be built and maintained independently, which means that intranets are typically more expensive than Internet connections. For this reason few, if any, testing agencies are now building remote test sites that connect to an intranet. Instead, those organizations are making use of secure Internet connections.

The previous administration guidelines apply to on-demand assessments that occur on a specific date, at a specific time, and at a specific location. In the case of some performance assessments, such as portfolios, some of the administrative procedures will not be relevant. For example, the portfolio used in the certification of teachers by the NBPTS is constructed by the teacher in her school setting. The process typically takes about a year to complete. Thus security issues as they were described in this chapter play less of a role in a portfolio assessment. However, the NBPTS does provide teachers with directions about completing the portfolio, so the guidelines provided in this chapter about developing test directions will inform the development of the portfolio directions.

With the administration of an assessment completed, the scoring of examinees' performance begins. However, first the scoring tools must be developed and the raters trained. Chapter 6 describes the process of developing the tools required to train raters in the scoring of performance tasks.

BOX 5.3. SCOTT'S WORST NIGHTMARE

If you ask Scott about his worst nightmare, he will describe something from his reality, as he has already lived his worst nightmare: the lost exam. After the safety of examinees during the examination, Scott's greatest concern is the safety of the examination materials. Some examinations can cost as much as $1,000,000 to create. This is why Scott maintains close control over the chain of custody for examination materials. This is also why he insists on tight security from the printer to the loading dock, to the examination room, and back to his warehouse. He tracks the location of every examination by a serial number from the time of its creation until the test materials are placed in secure storage or destroyed.

Scott describes the proctor as the single most important component of examination security. Without a vigilant proctor, the unscrupulous examinee could take secure content from the examination room, either by directly stealing materials or perhaps by photographing materials with high-tech cameras.

During the Exam

Scott has little direct control over what happens during the administration of the examination; he must rely of the efficiency and effectiveness of his test site administrators and the proctors. Fortunately, experience has taught him on whom he can rely and on whom he cannot.

After the Examination

Within 10 days after the examination, Scott has received all the materials associated with the examination. These materials are counted and processed to be sure that everything that left was returned. Scott reviews the irregularity reports written by the proctors and decides which reports require follow-up action. One such report, which required no follow-up action, indicated that the proctor fell and rolled down the steps of the lecture hall after having distributed the examinations. The proctor was uninjured, and he reported that no examinee had been injured in the "drop and roll" demonstration. He also reported extending the examination administration time by exactly two minutes to "make up" for the unintentional distraction.

In addition to auditing the returned materials, Scott collects the comments written by the examinees, and then he passes those comments to the project manager after copying those that pertain to the administration. When all materials are returned and counted, when all answer sheets are graded, and when all irregularities are documented, Scott files the paperwork describing the administration and begins a new cycle (if he does not have one already underway).

FURTHER READING

Cohen, A., & Wollack, J. (2006). Test administration, security, scoring, and reporting. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 355–386). Westport, CT: American Council on Education and Praeger.

Synthesizes the literature related to test administration.

Chapter Questions

1. If the developers of the NAEP visual and performing arts assessment wanted to include a performance task to assess student acting proficiency, what challenges to comparability must the developers address? In other words, what aspects of the performance task would you require to be standardized?

2. Consider a state that allows use of calculators on an examination, but does not provide calculators or specify the types of calculators allowed.

- **a.** How does this lack of standardization threaten comparability of scores and affect the interpretation of scores?
- **b.** How does this lack of standardization threaten the security of the examination?
- **c.** How does this lack of standardization influence the validity of the assessment?
- **d.** How could this lack of standardization be used to undermine the examination?

3. Review the *Outline of Potential Topics for a Test Administration Manual* (Table 5.1)and consider the administration of a set of reading and writing tasks for the evaluation of an adult literacy project. The project is only at one site and has 25 program participants.

- **a.** Which administration topics are most relevant in this case?
- **b.** What test administration staff will be required?

4. Consider a science achievement test and a test of reading comprehension for which a student requests an ADA accommodation for a reader. Why would a reader be considered an appropriate accommodation for the science test but not for the reading comprehension test?

CHECKLIST 5.1

Completed	To-Do List for Planning the Administration of Performance Tasks
✓	Parallel the development of test directions with the construction of the performance tasks. p. 121
	Draft directions to be simple and clear and have the same meaning for examinees, examiners, and administrators. p. 121
	Review a draft of the directions with test administrators, examiners, examinees, etc. p. 121
	Use text and page formats (e.g., boldface, enlargement, colors, spacing) to guide examinees and test staff to accurately follow directions. p. 122
	Pilot the test directions and finalize changes prior to operational testing. p. 121
	Examinee Directions
	Develop examinee handbook (or website page) with registration directions (e.g., date, time, location of examination; identification requirements for check-in; appropriate dress; personal effects and equipment allowed in examination room; requests for ADA accommodations; policies about cheating) and descriptive information about the test (e.g., content, KSAs, sample items). p. 122
	Write directions for examinees that state (1) the task, (2) required test materials, (3) general directions, (4) how to make responses (5) time limitations, (6) how to correct inadvertent responses, (7) what type of assistance they may receive if they do not understand the task, and (8) whether they may return to earlier sections of the test. p. 123
	Include in the directions an orientation to the test materials, such as computer software and equipment, numeric grids, or headphones. p. 123
	Use graphics in the test booklet to indicate whether examinees should GO ON or STOP. p. 123
	Administration Directions
	Develop manual that describes (1) the assessment, (2) the responsibilities of staff, (3) handling of test materials, (4) test environment (e.g., facilities, room arrangement), (5) examinees (e.g., eligibility, notification about test, admission to the examination site), (6) administration (e.g., distribution of materials, scripts, collection of materials at end of session), and (7) preparation of test materials for return to test agency. pp. 124–125

(continued)

CHECKLIST 3.1 *(page 2 of 2)*

	Delineate in the manual the responsibilities of the staff (e.g., test administrator, examiner, and proctor) and accompany the information with a checklist. pp. 124, 126–134
	Conduct staff training to explain the procedures and to provide practice in the administration of the test. pp. 134–135
	Testing Environment
	Arrange for a testing environment that provides reasonable comfort and that avoids noise, disruption in the testing, inadequate lighting, limited work space, and poorly reproduced materials. p. 135
	Standardize the performance environment to the extent possible by providing comparable equipment and logistics (e.g., number of examinees, size of room, availability of assistance). pp. 135, 137–138
	Standardize the software, equipment, and assistance available for computers. Provide instructions to staff about the use of the equipment and software. pp. 138–139
	Accommodations
	Establish timeline for requesting accommodations and publicize with examinee registration materials. pp. 139–140
	Review requests for accommodations, obtain expert opinion for accommodations requests that will be declined, and notify examinees of decisions. p. 140
	Security
	Formulate and publish policies about cheating and penalties. p. 149
	Establish procedures (e.g., limited access to facilities, encryption of files, test security, and confidentiality agreement) for the secure development of items and prompts. pp. 144–146
	Count materials to confirm numbers in tracking documents and store in location only accessible by test staff. p. 144
	Require photo identification and admissions letter for entry to the test site. p. 146
	Limit personal effects (e.g., briefcases, brimmed hats, palm pilots, digital watches, cell phones, pagers) and other equipment (e.g., MP3 players, cameras) allowed in the room. p. 147
	Arrange seating with space between examinees and prepare seating chart. p. 147
	Have at least one test staff member in the examination room at all times. p. 148