

---

EDJJ PROFESSIONAL DEVELOPMENT SERIES

MODULE 5:

**ASSESSMENT OF INDIVIDUALS WITH  
DISABILITIES IN THE JUVENILE  
JUSTICE SYSTEM**

Prepared by: John L. Hosp  
Michelle K. Hosp  
Ken Howell

## **Table of Contents**

- I. Introduction
- II. Objectives
- III. Content Outline
- IV. References
- V. Transparencies

**Introduction**

This module is one in a series of training packages that have been designed for working with adolescents and young adults with disabilities in correctional settings; it focuses on the assessment of individuals with disabilities.

## Objectives

Upon completion of this module, participants should be able to:

1. Describe the characteristics of a direct interventionist approach to instruction.
2. Describe the critical components of the classroom setting that affect instruction and learning.
3. Define and describe the general instructional emphasis of each stage of learning.
4. Describe and illustrate the eight steps in a systematic model of instruction.
5. Describe the similarities and differences between assessment and evaluation.
6. Describe the purpose of collecting assessment data.
7. Describe the characteristics for summative and formative assessment and evaluation.
8. Describe the role of databased decision making in the systematic instruction model.
9. Describe the factors that characterize effective and efficient databased decision making.
10. Describe the four levels of assessment.
11. Explain reasons for conducting assessments.
12. Identify the questions teachers should ask when they are describing the assessment situation.
13. Describe what needs to be done in planning a test exercise.
14. Identify questions that should be asked when describing what student performance will be evaluated.
15. Identify questions to be asked when designing a plan for rating and recording student performance.
16. When evaluating the technical aspects of assessment procedures and instruments, identify what prerequisite information must be considered.
17. Define and describe the following aspects that affect our confidence in an assessment procedure: reliability, validity, standard deviation, norm.
18. Define the purpose of diagnosis.

19. Describe the components of a two-tiered approach to diagnosis.
20. Describe strategies for determining relative standing or peer comparisons.
21. Describe the characteristics and three components of domain-referenced testing.
22. Describe strategies and guidelines for defining item selection procedures.
23. Describe the characteristics and differences of selection and supply type items.
24. Describe the characteristics, guidelines, and considerations associated with the administration, scoring, and reporting of tests and test outcomes.
25. Explain why it is important to assess social skills.
26. Indicate how to examine social behaviors in an objective and positive fashion.
27. Explain the differences among behavioral excesses, deficits, and assets.
28. Identify what social skills can be assessed.
29. Identify what should be assessed when assessing social skills.
30. Describe the six dimensions of a behavior.
31. Describe indirect observation strategies for assessing social skills.
32. Describe direct observation strategies for assessing social skills.
33. Describe what functional relationships are and how they are identified.
34. Describe the difference between testable explanations and explanatory fictions.

## Content Outline

Deleted: -----Page Break-----

### 1 The Direct Interventionist Approach

Formatted: Bullets and Numbering

- 1.1 A direct interventionist approach to educating adolescents and young adults with disabilities in correctional settings will be maintained throughout this module on assessment and evaluation. (For information on juvenile corrections to supplement information here, the reader may refer to the Juvenile Justice Trainers Association (available online at <http://www.jjta.org/>.)
- 1.2 This perspective is characterized as follows. (Refer to *Methods* module for more information; Display T-1.2.)
  - 1.2.1 Active. Teaching and curriculum decisions are actively directed by the teacher. (The reader is referred to Module 6: Curriculum in this series.)
  - 1.2.2 Responsible. The teacher assumes the responsibility for both successful and unsuccessful student learning and curriculum outcomes.
  - 1.2.3 Functional. The curriculum and content of instruction can be directly and immediately applied in the student's academic, social, and vocational environments; and can increase the likelihood of successful functioning in future settings.
  - 1.2.4 Performance based. Curriculum effectiveness is determined by evaluating observable and measurable student performance.
  - 1.2.5 Dynamic. Instructional and curricular adjustments are made continuously and are based on student performance.

### 2 Critical Components of the Instructional Setting

Formatted: Bullets and Numbering

- 2.1 When measuring student skill levels or performance abilities (either academic or social behaviors), it is critical to consider the setting context in which the behavior is occurring. (For instructional methods, the reader is referred to Module 7 in this series.)
  - 2.1.1 Students do not learn or perform in a vacuum. Learning is usually direct and intentional. Teachers present the curriculum, manipulate the instruction, and give rewarding or corrective feedback in a manner that is systematic and direct. These teacher variables must be assessed when evaluating student learning.
  - 2.1.2 Similarly, learning can be unintentional as well as intentional. The teacher can only control and manipulate a limited number of instructional variables.
    - (a) These variables reside outside the instructional procedures and curriculum and may not be influential to student learning or performance.

- (b) These variables are associated with and/or controlled by
  - The student
  - The teacher
  - The social environment which includes
    - Parents/family of the student
    - Peer group
    - School environment

2.1.3 We will talk about each of these briefly. A more detailed and descriptive discussion can be found in the *Methods* module (Display T-2.1.3).

2.2 The student brings factors to the learning environment that must be considered in planning instruction.

2.2.1 The student's disability or behaviors

- (a) Motor/physical limitations
- (b) Behavioral interferences
  - Excesses such as aggression or hyperactivity
  - Deficits such as low rates of verbal interaction or withdrawal
- (c) Learning deficits which generally include unlearned academic prerequisite skills
- (d) The connection between academics and behaviors is noted and is supported in research. For further information on tests and information related to learning disabilities, the reader is referred to Nelson, Rutherford, & Wolford (1987), and the National Council of Juvenile Justice and Family Court Judges (1988).

2.2.2 The student's biology or inherited characteristics

- (a) Physical impairments
- (b) Metabolic impairments
- (c) Sensory impairments (vision, hearing, etc.)

2.2.3 The student's parental/family dynamics

- (a) Racial, cultural, religious characteristics
- (b) Vocational/financial/economic status/living arrangements
- (c) Parenting style, both affect and behavior management style
  - Authoritarian = dictatorial, undemocratic, uncompromising
  - Authoritative = democratic, compromising
  - Permissive = nondirective, unquestioning

2.2.4 The student's learning history

- (a) Academic history
- (b) Social skill history

2.2.5 The student's educational placement history

- (a) Regular education

- (b) Special education
- (c) Corrections

2.3 Not only must teachers assess and consider the student's characteristics and learning performance, they must also assess and evaluate their own influences or biasing characteristics.

2.3.1 The teacher's biology or inherited characteristics

- (a) Physical impairments
- (b) Metabolic impairments
- (c) Sensory impairments

2.3.2 The teacher's background history

- (a) Racial, cultural, religious characteristics
- (b) Vocational/financial/economic status
- (c) Familial history, both affective and management style

2.3.3 The teacher's learning history

- (a) Academic history
- (b) Social skill history

2.3.4 The teacher's educational history

- (a) Teacher training experiences
- (b) Previous employment history
- (c) Current employment perspectives, requirements, limitations, etc.

2.4 The social environment of the adolescent has a significant influence on instructional programming and teaching effectiveness, especially with respect to academic and social behavior management. Thus, it also must be carefully considered during the assessment of academic and social behaviors.

2.4.1 As mentioned above, the parents/family of the adolescent shape the student's values, attitudes, overt behaviors, etc.

2.4.2 Peers and the role of the adolescent in the peer group must be considered during assessment procedures.

- (a) The role of the peer group becomes more important in shaping social behaviors during adolescence than during early childhood experiences.
- (b) The adolescent must deal with rules and guidelines associated with popularity, leadership, and followership. Each adolescent in each peer group displays behaviors that establish a variety of roles (i.e., leader, follower, instigator, scapegoat, fall guy, etc.).

2.4.3 The school or educational placement also affects the social and academic development of the adolescent.

- (a) Insensitivity to student's individuality

- Failure to accommodate individual learning and behavioral differences
- (b) Inappropriate expectations
  - Teacher expectancies affect how teachers behave toward individual students.
- (c) Inconsistent management
  - A condition under which teacher behavior cannot be predicted by students
- (d) Instruction in nonfunctional and irrelevant skills
  - Students must assess and be convinced that a particular learning activity is important and that it will be important in the future
  - Teachers must assess for and select instructional skills that are functional and produce critical effects (i.e., those behaviors which will be associated with rewarding outcomes).
- (e) Nefarious contingencies of reinforcement
  - The student receives attention which strengthens and maintains inappropriate behaviors
  - There is also a failure to provide regular attention and feedback for appropriate behaviors
- (f) Undesirable models
  - Both desirable and undesirable behaviors may be modeled by teachers and peers.

2.4.4 The social development of the adolescent is affected by what is presented and emphasized in the mass media. Knowing what components of the media influences learning and behavior can be a powerful piece of evaluation knowledge.

← - - - - Formatted: Bullets and Numbering

- (a) Television
  - Generalization of TV aggression is affected by the student's ability to discriminate between what is fantasy and reality. The more reality-based the student's ability to interpret what is viewed, the less likely the aggressive behavior will be modeled or learned.
- (b) Music trends
- (c) Clothing and dress habits
- (d) Movies

### 3 Stages of Learning

← - - - - Formatted: Bullets and Numbering

3.1 The assessment and evaluation of the academic and social behavior of the adolescent or young adult with a disability should be conducted within the context of the kind of learning in which the student is engaged.

← - - - - Formatted: Bullets and Numbering

3.1.1 By assessing at what stage of learning a student is functioning, specific remedial interventions and strategies may be implemented.

- (a) Instructional programming would be different for a skill that is being taught and learned for the first time (acquisition) than for a skill that has been mastered but needs to be practiced for improved fluency or proficiency.

(b) For example, the student who fails to calculate the correct answer for a long division problem because s/he is learning the computation sequence for the first time would receive different instruction than the student who has mastered the skill, but fails to show consistent performance over time.

- In the former case, a series of computational rules would be taught (i.e., antecedent instructional manipulations would be conducted).
- In the latter case, the rules have been learned, so the instructional emphasis would be placed on motivating the student to show more consistent performance (i.e., consequence manipulations).

Formatted: Bullets and Numbering

3.2 The different kinds of learning stages may be characterized in one of five different ways. Each of these might require a different instructional emphasis (Display T-3.2).

Formatted: Bullets and Numbering

### 3.2.1 Acquisition

- (a) Definition—a condition in which a student has never displayed the skill or has shown it at extremely low correct rates
- For example, a student who has never owned a checking or savings account probably would not be able to balance his/her account without specific instructional assistance.
- (b) General instructional emphasis
- Focus on strategies which emphasize the presentation of instruction and the shaping of correct responses.
  - Focus on regular and meaningful feedback for correct responding and on corrective feedback for student errors.

### 3.1.1 Fluency or proficiency

Formatted: Bullets and Numbering

- (a) Definition—a condition in which the student has demonstrated accurate responses but inadequate rates or frequencies of responding (i.e., too fast or too slow)
- For example, a student may be able to fill out a job application accurately, but it takes him/her three times as long as his/her peers to complete it.
- (b) General instructional emphasis
- Focus on the kind of feedback given for correct responding. The goal is to increase the student's motivation by giving him/her immediate and positive feedback for correct responses and higher response rates. After the desired proficiencies have been demonstrated, the teacher would gradually decrease the amount and frequency of feedback.

Formatted: Bullets and Numbering

### 3.1.2 Maintenance

Formatted: Bullets and Numbering

- (a) Definition—a condition in which the student has demonstrated accurate and fluent responding, but his level of responding fails to endure over time
- For example, a student may be able to demonstrate accurate and proficient typing skills, but after a period of time, s/he is unable to demonstrate the same levels of typing
- (b) General instructional emphasis

- Focus on making instructional feedback/reinforcement even more intermittent. Also increase the delay between the student's response and the teacher's feedback. Again, the attention is focused on increasing the student's motivation to maintain acceptable performance over time.

### 3.1.3 Generalization

Formatted: Bullets and Numbering

- (a) Definition—a condition in which the student has acquired a proficient and enduring instructional response which is also observed under different instructional or training conditions
  - For example, a student may be competent at using a time clock correctly and proficiently in the classroom situation, but does not display the skill at his/her job after school.
- (b) General instructional emphasis
  - Focus on teaching the instructional behavior or skill under varied training conditions (e.g., vary instructional materials, trainers, settings, etc.)

### 3.1.4 Adaptation

Formatted: Bullets and Numbering

- (a) Definition—condition in which the student modifies a learned behavior to produce successful outcomes under varied instructional or response conditions. It is also described as the application of problem-solving sequences to novel stimulus conditions. It involves adapting a skill to meet the requirements of a new situation.
  - For example, a student may be competent at using a coin-operated washing machine, but is unable to run a similar machine that is not coin-operated
- (b) General instructional emphasis
  - Focus on teaching rule-governed behaviors under a variety of environmental conditions which require adjustments of the instructional behaviors for successful outcomes
  - Focus on teaching problem solving sequences which can be applied to a variety of responses of setting conditions

## 4 Systematic Instruction

Formatted: Bullets and Numbering

4.1 For learning to be effective and efficient, instruction must be structured, meaningful, and direct. The systematic instruction model provides a format and process for maximized learning and instructional effectiveness. It also permits the teacher to translate material directly and systematically from the IEP to actual classroom instruction. Many of the same components found on the IEP can be found in one or more of the steps of the systematic instructional model.

4.1.1 Being familiar with the steps in the instructional process will facilitate the assessment and evaluation of learner performance.

4.1.2 The systematic instruction process consists of eight essential steps (See Display T-4.1.2).

- (a) Assess student strengths and areas of need.
- For example, Howard would like to get his driver's license. We would first assess what rules-of-the-road he already knows and can demonstrate. We would also need to evaluate his prerequisite skills, such as reading and writing, his previous driving experiences and practice, etc. This information would dictate what Howard would have to learn.
- (b) Set long-term objectives.
- Identify instructional behaviors to be taught.
  - For example, a long-term objective might focus on Howard's learning to name twenty road signs by name and function when they are presented on slides.
- (c) Set short-term objectives.
- Identify instructional behaviors from a task analysis of the long-term objective
  - Given Howard's prior learning history, we might want to break the previous long-term objective into four short-term objectives. Instructionally, Howard would first learn five basic signs. The next objective would require him to learn five new signs as well as be able to recall the five signs from the previous short-term objective, etc.
- (d) Write an instructional plan.
- Focus on instructional manipulations to be conducted by the teacher to teach to the long- and short-term objectives.
  - Given the short-term objectives we have identified, we might develop an instructional plan that requires Howard to take a picture of each sign with a Polaroid camera and to create a poster for each sign which displays its meaning. This activity might be followed by a lesson using slides of the same signs and requiring Howard to name the sign and describe its function or meaning.
- (e) Write a measurement plan.
- Identify what and how student responses will be assessed and evaluated.
  - The measurement plan would consist of a simple daily probe that requires Howard to name all signs to which he has been introduced.
- (f) Implement the instructional and measurement plans.
- Once the above steps have been completed, we would put our lesson on road signs into effect and we would implement the measurement process.
- (g) Modify the instructional plan based on the student's performance data.
- Using the rules for change specified in the measurement plan, we would evaluate Howard's performance and determine if instructional changes are necessary or not. For example, if Howard labeled 60% or less of the road signs correctly for five consecutive days, we might consider changing the Polaroid and slide presentation to some other instructional format.
- (h) Evaluate the effectiveness of the total instructional plan.
- This step asks the teacher to determine if the lesson achieved what it was supposed to for Howard. We might ask ourselves if we would use the lesson or its instructional format again.

4.2 It is very likely that you have already engaged in many of these instructional steps in your work with adolescents and young adults with disabilities. However, we want to emphasize the importance of a systematic approach to instruction for students with disabilities, and how assessment and evaluation are critical to the effective and efficient functioning of systematic forms of instruction.

4.2.1 The rest of this module will focus on assessment and evaluation of academic and social behaviors. We should continually remind ourselves of the dependent relationship among our measurement and evaluation systems, our instructional methods, and the learner's performance or learning.

## | 5 Assessment and Evaluation

Formatted: Bullets and Numbering

5.1 Assessment and evaluation are two of the most important components of instructional programming, curriculum selection, and educational decision making. In this section, we will discuss and review some key assessment fundamentals. Information from this section of the module will serve as prerequisite knowledge for the sections that focus on the assessment of academic behaviors and social behaviors.

5.2 What are assessment and evaluation?

5.2.1 Assessment is defined as the measurement of some set of skills (i.e., academic or social behavioral).

(a) We can assess or measure a variety of things:

- Student learning, behaviors, and performance
- Teacher behaviors, instruction, classroom management
- Classroom instructional environment and setting
- Characteristics of other settings (e.g., family or living environment, work place, etc.)

(b) The purpose of these assessment data and procedures is to acquire information that permits the teacher and student to evaluate current progress and instructional effectiveness.

- Communication
- Accountability
- Standards for comparison or evaluation
- Evaluation of instructional effectiveness and efficiency
- Identification, classification, and placement

5.2.2 Evaluation refers to the decision-making process that we engage in when we use and analyze student performance data (i.e., assessment).

(a) We evaluate assessment data in a variety of ways:

- Within comparisons (e.g., comparing a student's previous performance with current levels)
- Between comparisons (e.g., comparing a student's performance to some standardized set of norm or criteria)

- Functional comparisons (e.g., determining whether an instructional procedure or intervention is effective in creating the desired behavioral change)

5.2.3 Assessment and evaluation processes may take one of two forms:

(a) Summative assessment and evaluation involves the measurement and analysis of student learning against a specific long-term objective at the beginning (pretest) and at the end (posttest) of instruction.

- For example, when teaching a student a complex skill such as map reading, we might administer a pretest prior to our instruction and then give a similar posttest after our instruction. This information would enable us to determine what kinds of map reading skills the student had at the beginning and end of a lesson.

(a) Formative assessment and evaluation involve the continuous measurement and analysis of student learning and performance throughout the instructional process.

Formatted: Bullets and Numbering

- In our example on map reading, student progress would be monitored and evaluated throughout the lesson—not just at the beginning and end of the lesson. This information would tell us something about the student’s success and error rates, and would allow us to make instructional adjustments before the student has experienced too much failure or frustration.

(c) Data-based decision making is an essential element of both summative and formative assessment/evaluation.

Formatted: Bullets and Numbering

- Instructional modifications frequently occur at the end of some administrative time period rather than on the basis of the prevailing response patterns of the student. As in our example of map reading, student learning and instructional programming must be continuously assessed and evaluated. Similarly, modifications indicated by these evaluations must be immediately implemented. Effective and efficient data-based decision making is characterized by the following.

- Behaviors are measurable.
  - For example, it is difficult to evaluate Howard’s knowledge of street signs if we limit ourselves to his knowing them, but we can be very efficient if we ask him to name the function of a given sign verbally and in writing.
- Measurable behaviors or expectations are indicated in behavioral objectives.
- Acceptable accuracy levels are measurable.
  - Similarly, it is very hard to determine when Howard has mastered a set of road signs when we state that he must get a “majority” of them right. When we specify “8 out of every 10 signs”, it is clear when he has mastered a set of signs.
- Specific and functional measurement procedures have been written.
  - This requirement is similar to specifying accuracy levels. Simply stated, it is easier to monitor Howard’s progress if we measure his

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

- learning slide presentations rather than driving him through the streets every day.
  - The measurement procedures directly measure the instructional behaviors.
    - This condition is rather obvious. It indicates that we should require Howard to respond in a manner similar to what he will be required to do when he takes his driver's test. Our behavioral objectives should reflect this requirement.
  - The measurement procedures are implementable on a continuous and consistent basis.
    - Our measurement and evaluation systems should allow us to monitor Howard's progress on a continuous (i.e., daily or weekly) basis. We want to be able to modify our instruction and to catch his error patterns before they become too difficult to change.
- Based on measurable criteria indicated in the behavioral objective, specific and measurable rules for instructional change must be specified (i.e., data-decision rules).
  - These data-decision rules should be expressed in the form of if-then statements.
    - Examples: If Edgar balances his personal checkbook accurately each day for five consecutive days, move to the next instructional objective.
    - If Edgar incorrectly balances his personal checkbook each day for five consecutive days, change the instructional objective to a more accomplishable level (e.g., give him another student's personal checkbook to use as a model).
  - A rule should be specified for acceptable student performance (i.e., when performance trends or patterns indicate that the student will achieve the behavioral objective by the specified time under the prevailing instructional conditions).
  - A rule should be specified for unacceptable student performance (i.e., when performance trends or patterns indicate that the student will not achieve the behavioral objective by the specified time unless the timeline is extended and/or the instructional programming is adjusted).

Formatted: Bullets and Numbering

## | 6 Kinds of Assessment

Formatted: Bullets and Numbering

6.1 Collecting and measuring student performance can be accomplished in a variety of ways. Basically, we can identify four levels of assessment (Display T-6.1).

- 6.1.1 Level one consists of written documents or archives about past student performance and/or instructional strategies. These most frequently take the form of written reports and records (e.g., the student's cumulative file).
- 6.1.2 Level two assessment focuses on information collected through interviews and verbal reports. These types of data are obtained by asking others who have

worked with a particular student or who are familiar with his or her particular academic and social behavioral strengths and areas of need and effective and ineffective instructional strategies.

- 6.1.3 Formal and informal written tests make up level three assessments. Most level three assessment devices are standardized, norm-referenced tests or informal, curriculum-based tests. They may be obtained through the usual commercial vendor, or they may be teacher made. Checklists and other forms of behavioral or instructional inventories are also included in this level of assessment.
- 6.1.4 The last, most direct, and objective level of assessment includes direct observational techniques. Procedures employed at this level rely on the teacher making direct observations of behaviors as they are occurring and recording them directly in some systematic manner. It is this level of assessment that we will be emphasizing in this module, especially with respect to the assessment and evaluation of social skills.

## | 7 Assessment Guidelines

← - - - - Formatted: Bullets and Numbering

7.1 When developing assessment materials and procedures, it is important that the assessments of performance are as systematic and objective as possible. The goal is to maximize your confidence in the appropriateness of the results for their intended use or purpose. In this section of the module, we will discuss a simple four step sequence.

7.2 Step one: Describe the assessment situation (Display T-7.2).

7.2.1 What is the reason for assessment?

(a) Management decisions

- Identifying student strengths and needs
- Matching appropriate instruction and curriculum to student needs
- Determining current level of functioning and predicting future performance

← - - - - Formatted: Bullets and Numbering

(b) Selection decisions

- Determination of appropriate program placement, modification, or service delivery requirements

← - - - - Formatted: Bullets and Numbering

(c) Programmatic decisions

- Determination of overall achievement level of students in program
- Determination of program components requiring modification
- Determination of program additions, deletions, or expansions

← - - - - Formatted: Bullets and Numbering

7.2.2 Who will conduct the assessment and who are the decision-makers?

7.2.3 What knowledge and/or set of skills will be assessed? “What prerequisite skills are being assumed?”

7.2.4 Who are the students to be tested?

- (a) Age/grade; number; major learning difficulty; and other related characteristics ← Formatted: Bullets and Numbering

### 7.3 Step two: Plan the test exercise.

#### 7.3.1 Describe the tasks planned to sample student skills.

- (a) Describe the test activity ← Formatted: Bullets and Numbering
- (b) Determine whether to test formally or informally
- (c) Determine the number of performance samples required
- (d) Determine who will supervise the assessment

#### 7.3.2 Under what conditions will student performance be assessed?

- (a) In the classroom by observing everyday activities? ← Formatted: Bullets and Numbering
- Are these activities readily available?
  - Are resources available to develop the assessment procedures and materials and to conduct the assessment?
  - Do the everyday settings provide sufficient opportunities for the desired behavior?
  - Can the setting be used to assess more than one behavior or generalized to other settings?
- (b) Design a specific test exercise and situation that simulates the desired performance conditions. ← Formatted: Bullets and Numbering
- Can the desired simulation be realistically created?
  - Are resources available to develop and evaluate the exercises?
  - Does the test exercise or situation allow each student an opportunity to respond to the same test activity?

#### 7.3.3 How many assessment tasks or samples of performance are needed to judge a student's skill?

- (a) Design one assessment device or procedure and apply repeatedly.
- (b) Design and use several difference assessments to assess the same performance skill at one time.
- (c) Use different assessments over time.
- (d) Suggestions.
- Gather as much assessment materials as required for the intended purpose.
  - Gather sufficient data to enable objective and valid evaluations
  - Repeated measures over time have the advantage of showing change or trends over time. They can provide a picture of the learning or behavior change progress.

#### 7.3.4 Who will supervise or administer the assessment activity?

- (a) Consider the teacher or other educator as supervisor if ← Formatted: Bullets and Numbering
- student has a vested interest in demonstrating high performance for the teacher (i.e., for grades).
  - standardized testing conditions are required.
  - teacher or aide has the time.
  - a skilled person is required to administer or evaluate performance.

- (b) Consider unsupervised conditions if:
- students have no reason to misrepresent their performance.
  - standardized testing conditions are not required.
  - insufficient time is available for supervised or individualized administration.
  - permanent products result for later scoring or analysis.

← - - - - Formatted: Bullets and Numbering

7.4 Step three: Describe the performance to be evaluated.

7.4.1 What kind of performance will be evaluated?

- (a) The outcome or product of behavior
- Use if a lasting tangible product results
  - Use if quality or performance is more important than quantity or speed
  - Use with time to have some record of proficiency or fluency
- (b) The actual behavior or performance
- Use if process or action is more important than product of behavior
  - Use when rate and/or quantity of performance is important
  - Use when conditions under which behavior occurs is important
  - Use when time and resources are available to conduct procedures

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

7.4.2 What aspects of the behavior or performance skills will be evaluated?

- (a) For a directly observed behavior or skill, examine its dimensions.
- Rate or frequency
  - Duration or length
  - Latency or initiation time
  - Locus or location
  - Force or intensity
  - Topography or shape
  - Accuracy or standard
  - Functional effect or result
- (b) Suggestions
- Define performance criteria or dimensions prior to administering the assessment.
  - Focus on observable behaviors.
  - With products, focus on quality and attributes.
  - Be certain that measurement unit matches relevant dimension of the skill, performance, or product.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

7.4.3 Will the assessment be unobtrusive or public?

- (a) Public assessment procedures are associated with a variety of reactive effects.
- Motivation
  - Competition
  - Stress and anxiety
- (b) Suggestions and considerations
- Consider privacy rights of student

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

- Consider natural vs. contrived performance effects
- Control for anxiety

#### 7.5 Step four: Design a plan for rating and recording performance.

##### 7.5.1 How will the assessment data be used?

###### (a) Will actual student performance data be used?

- Often most useful for diagnosis, skill mastery assessment, and evaluation of instructional effectiveness

← --- Formatted: Bullets and Numbering

###### (b) Will student performance be summarized or manipulated?

- Applying a statistic (i.e., mean, median, etc.)
- Standardization of data (i.e., percentiles, standard scores, etc.)
- Most often useful for grading, grouping, and placement decisions

← --- Formatted: Bullets and Numbering

###### (c) Considerations

- Resources available to observe, summarize, rate, and manipulate performance data

← --- Formatted: Bullets and Numbering

##### 7.5.2 Who will rate or evaluate performance?

###### (a) Consider teacher (or other educator) if

- specialized knowledge in analyzing student response or applying evaluation procedures are required.
- uniform standards and high accuracy required.
- teacher has time and resources.

← --- Formatted: Bullets and Numbering

###### (b) Consider peer or self-rating if

- slight variations from rater to rater are acceptable.
- students have no vested interest in outcome.
- resources are limited.
- students have received some training in purpose and implementation of assessment and evaluation procedures.

← --- Formatted: Bullets and Numbering

##### 7.5.3 How will assessment data be recorded?

###### (a) Consider direct observation procedures.

- Potentially most accurate and objective since no indirect measures and few judgments are employed
- Gives direct picture of performance

← --- Formatted: Bullets and Numbering

###### (b) Consider checklists or rating scales.

- Combine observation with evaluation
- Easy to develop and/or administer
- Useful for establishing frame of reference for understanding performance

← --- Formatted: Bullets and Numbering

###### (c) Consider anecdotal records.

- Useful for unique events or behavior that cannot be adapted to systematic measurement procedures
- Can be left open-ended and flexible

← --- Formatted: Bullets and Numbering

##### 7.5.4 How are assessment data to be interpreted?

- (a) Compare student’s data with that of other students (between comparison).
  - For ranking, grouping, guidance, and placement purposes
- (b) Compare student’s data with a preestablished standard or criterion of acceptable performance.
  - Diagnosing and certifying skill mastery (between comparison)
  - Evaluating instructional effectiveness (functional comparison)
- (c) Compare student’s data with self (within comparison).
  - Show learner progress and trends over time
  - Make predictions about future performance
- (d) Compare student’s performance against standards established by social environment (between comparison).
  - Social validation (i.e., meeting expectations of others such as parents, employer, etc.)

← --- Formatted: Bullets and Numbering

← --- Formatted: Bullets and Numbering

← --- Formatted: Bullets and Numbering

← --- Formatted: Bullets and Numbering

## | 8 Technical Aspects of Assessment

← --- Formatted: Bullets and Numbering

8.1 The evaluation of the technical aspects of assessment procedures and instruments requires that certain prerequisites be considered.

8.1.1 Clear statement of purpose

- (a) What decisions are to be made.
- (b) Who the decision maker(s) will be.
- (c) What skills will be assessed.

← --- Formatted: Bullets and Numbering

8.1.2 Clear communication of purpose of test (i.e., instructions)

- (a) Explain what and how test is to be taken or administered.
- (b) Specify expectations about performance.

← --- Formatted: Bullets and Numbering

8.1.3 Bias free (i.e., test student performance, not attitudes of evaluator)

- (a) No measure is completely without bias, but precautions can be taken to reduce its effects.
- (b) Control for sexual, racial, cultural, etc., differences.
- (c) Provide equal performance opportunities for all.
- (d) Use explicit and performance based training and scoring criteria.
- (e) Write it down, do not depend on a mental notebook approach.
- (f) When possible, rate each student “blind” (i.e., without knowledge of student identity).
- (g) Test in student’s primary language.

← --- Formatted: Bullets and Numbering

8.2 Confidence. Whenever we assess a student’s performance, either indirectly or directly, we must concern ourselves with two critical issues: the sensitivity of our measuring devices or procedures, and accuracy with which we use them. In this section of the module, we will examine some of the technical aspects of assessment and evaluation that affect our degree of confidence in what and how we measure student performance.

8.3 Error

8.3.1 Whenever we observe or measure something, some degree of error is *always* present.

- (a) The concern we must address is *how much* error.
- (b) Two statistics allow us to evaluate the amount of error present and the degree of confidence we can place on the measurement.
  - Reliability coefficients
  - Standard deviations
- (c) We will not be discussing how to compute these statistics, but instead we will attempt to understand what they represent and how we can use them.

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

8.3.2 Reliability coefficients

- (a) Reliability refers to the degree of agreement in repeated measures or assessments of something. It is a reflection of the dependability of the assessment information.
- (b) The reliability coefficient refers to the proportion of variability or error in a set of scores that reflects true differences among the individuals.
- (c) We can derive reliability coefficients in three basic ways:
  - Test-retest (index of stability)
    - Administrate test to same population approximately two weeks later with same instrument and procedures.
    - Scores from both administrations are correlated.
    - Measure the degree of stability of instrument over time.
  - Alternate form
    - Two alternate forms of an assessment device administered to the same population of students
    - To determine at what level they measure the same trait or skills
    - Can be administered at the same time (concurrent) or at a future time, similar to test-retest (predictive)
  - Internal consistency
    - If all items in a test measure the same trait or skill, then by splitting the items in half, the correlation between the two halves should be very high or close to 1.
    - Another common form is coefficient alpha, the average correlation between the items of a test.
- (d) Assessment procedures with reliability scores less than 0.9 (1.0 = perfect reliability) should be used with great caution when making individualized decisions (i.e., for a specific student). Group or curriculum decisions (i.e., with data for a group of students) should be made with caution with reliability scores less than 0.8.
  - Sources of low reliability
    - Low-quality assessments
    - No clear purpose or performance criteria stated
    - Insufficient performance samples gathered
    - Carelessly and inexplicitly scored
    - Poor test administration practices

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

- Student and/or observer/tester distractions
    - Non-uniform exercises, responses, and ratings
  - Characteristics of the student
    - Poor student motivation
    - High test anxiety
- (e) Uses of reliability coefficients
- Estimate test's relative freedom from measurement error
  - Estimate of individual's true score
  - To determine standard error of measurement

### 8.3.3 Standard error of measurement (SEM)

- (a) SEM gives estimate of amount of error that might be associated with a student's true score.
- (b) SEM is based on standard deviation of the error distributed around a true score.
- (c) Measurement error is always present for a person's score or measurement of performance.
- SEM gives information about the degree of certainty or confidence you may have in interpreting or using the score.
  - The larger the score, the greater the degree of uncertainty, the less sure we can be about the person's score.

← - - - - Formatted: Bullets and Numbering

## 8.4 Validity

8.4.1 Validity reflects the extent or degree to which a procedure measures what it is intended to measure.

- (a) Used to determine the kinds of inferences that can be made from a measurement or performance score
- For example, the extent that bias influences the measurement of a skill
- (b) Validity is based on a wide array of information.
- The collection of this information is called validation.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

8.4.2 Three basic kinds of validity

- (a) Content validity—examination of the appropriateness of the items used
- Appropriateness of types of items included (i.e., their relation to the topic being assessed)
  - Completeness of the item sample (i.e., are all aspects of the topic being addressed?)
  - Method by which items measure content or skill
- (b) Criterion-related validity—extent to which a person's performance on a specific criterion (e.g., grades) can be estimated from that person's true score
- Concurrent criterion-related validity—relation to current criterion scores
  - Predictive criterion-related validity—predicting future criterion scores

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

(c) Construct validity—estimating the validity or confidence of the construct or hypothesis that may underlie a test or score

Formatted: Bullets and Numbering

- Determined inferentially by examining the
  - Definition of the construct
  - Theory behind the construct
  - Empirical research on the construct
- For example, intelligence, learning ability, and self-esteem are all constructs that may underlie performance on a measure.

Formatted: Bullets and Numbering

#### 8.4.3 Guidelines for maximizing validity

- (a) Provide a clear statement of purpose of assessment.
- (b) Maintain a clear relationship between behavior or product being measured and the characteristic or construct being examined.
- (c) Use different assessment formats to measure the skill or behavior.
- (d) Emphasize direct observation or measurement of behavior.
- (e) Conduct unobtrusive measurement procedures to avoid reactivity and reduce bias.
- (f) Provide a clear and observable definition of behavior or skill.
  - Definition process should occur before items are selected rather than vice versa.
- (g) Provide a clear understanding and definition of assumptions and hypothesis that underlie domain to be measured.

Formatted: Bullets and Numbering

### 8.5 Norms, or representativeness

#### 8.5.1 Test scores or measures are used in a variety of ways.

- (a) They are frequently used to standardize groups of students.
- (b) They are also used to evaluate a pupil's performance against some standard or norm.

Formatted: Bullets and Numbering

#### 8.5.2 When selecting or constructing tests, it is important to evaluate the representativeness of the population of subjects used to norm or standardize a test.

- (a) Number of people
- (b) Age
- (c) Sex
- (d) Grade
- (e) Cultural background of parents/families
- (f) Geographic factors (e.g., region of country)
- (g) Race
- (h) Intelligence/achievement
- (i) Date of norm determination
- (j) Special population characteristics (e.g., disabilities)
- (k) Proportion of kinds of people used

Formatted: Bullets and Numbering

#### 8.5.3 Guidelines

- (a) If norm sample is different than students being assessed, limited generalizations and comparisons can be made without qualification.
- (b) Focus on *within-student* assessments rather than *between-student* assessments.

← - - - - Formatted: Bullets and Numbering

## 9 Basic Strategies for Assessing Academic Performance

9.1 In this section, we will examine some basic strategies for assessing academic performance.

9.1.1 Three major assessment approaches will be described.

- (a) Diagnosis or within-student assessment
- (b) Peer comparisons or between-student assessment
- (c) Mastery domain or student/curriculum assessment

← - - - - Formatted: Bullets and Numbering

9.1.2 The following discussions are not intended to be comprehensive, but instead, representative of the approach. It will be your responsibility to apply these strategies to your own teaching situations.

### 9.2 Diagnosis

9.2.1 Diagnosis is defined as the development of an assessment situation that is designed to analyze a student's current academic skills and inferring error patterns.

9.2.2 The purpose of diagnosis is to determine how the student responds.

- (a) This type of analysis requires the presentation of test materials which are essentially open-ended.
- (b) The actual administration and scoring of the test is of less concern.
  - For example, in math, a series of computation problems were developed and presented on paper. These problems represented all four operations with whole numbers, fractions, and decimals. Marilyn was directed to solve as many problems as possible, to show all of her work, and to leave no problems unattempted. Her teacher then reviewed her work looking for particular problem types that were consistently missed. Part of this analysis also included assessing which problem types were completed without errors and which were not attempted at all.

← - - - - Formatted: Bullets and Numbering

9.2.3 The diagnosis of academic skills can be viewed as a two-tiered approach.

- (a) The first tier, or survey level, consists of the development of materials that span well below and well above the estimated level of proficiency of the student.
  - A number of guidelines should be considered in constructing academic test materials.
    - Take random samples from a wide range of problems.
    - Teachers should give directions to the student that indicate the test includes a wide range of difficult problems.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

- Students should also be informed that they are not expected to answer all of the items correctly.
  - For example, in developing a spelling test for Roland, the teacher included words that ranged from very easy to very difficult. High frequency, irregular words and lower frequency, regular words were also included. Enough of these materials were developed (e.g., a total of 75 words) to provide Roland with plenty of opportunities to make mistakes. He was told to write the words on a blank sheet of paper when the teacher pronounced them, and to attempt each word even though he might not be familiar with all of the words
- (b) The second tier, or specific level, consists of the analysis of the survey level assessment data. ← - - - Formatted: Bullets and Numbering
- The purpose of this analysis is to formulate an appropriate and specific follow-up assessment.
    - That will pinpoint specific correct and error response patterns. ← - - - Formatted: Bullets and Numbering
    - This analysis requires the detailed specification of particular item types from a well-defined domain of objectives.
    - The domain from which item types are derived can be defined in one of two manners.
      - One approach is to task analyze the survey level stimulus materials. This analysis requires a table of specifications consisting of two categories: The behavioral dimensions or responses required of the student, and the material content which is ranked from difficult to progressively simpler material. The behavior or response requirements move from simple to more difficult: identification; accuracy (percentage correct); mastery (accuracy within circumscribed time limits (i.e., rate); and automaticity (mastery in the presence of distractors). The task analysis approach requires the systematic testing of the student in all of the above response requirements. Find the cell which is most appropriate for the student. For example, in reading, an informal assessment was developed from a sampling of a wide range of materials. Pages were copied from a *Readers Digest*; the sports, society and front pages of a local newspaper; pages from more specialized magazines; and finally reference manuals. These samples represented increasingly more difficult reading material. These tasks were presented to Jim and he was asked to read from these materials and then describe their content by answering multiple-choice questions, writing short answers, and paraphrasing the author.
      - In the second approach, an error analysis is conducted by reviewing the student's errors; these errors are classified into particular types and materials are developed that contain the items in error. For example, after being presented assessment materials, an analysis of Al's errors revealed difficulties with sequenced material (i.e., two or three events were described out of sequence).

A set of passages were then developed which emphasized sequences of events. Al was asked to read the passages and describe the order of events. His performance was analyzed to determine his response strategies and to prepare an instructional procedure to remediate the sequencing difficulty.

- 9.2.4 Materials for this type of assessment can be published or teacher-made as long as they meet the above criteria.

### 9.3 Peer comparison or relative standing

- 9.3.1 In this approach to academic assessment, the student's performance is compared to the performance of his/her peers, and a determination of relative standing is made.

- (a) The emphasis is placed on the standardized administration of tests.

- A constant measurement task is used.
- The administration and scoring procedures are controlled and standardized.

← - - - - Formatted: Bullets and Numbering

- (b) Two types of materials can be used in this testing approach..

- Published tests

- These tests are usually developed in the private market and sold through publishing and testing companies.
- They usually stand alone and contain all of the materials needed to test and score the student.
- They may be thought of as a form of survey level assessment in that a broad range of material is usually included, with very few items within any one objective or domain area. The important questions to address in the selection of this type of test are: What are the behaviors sampled? What is the format of the student's response (often a multiple-choice format is used)? For what purpose are the results to be used (for published tests, it is usually limited to the determination of general placement decisions)? Is the norm group appropriate? Is the test technically adequate (i.e., reliable and valid)? For example, Mr. Reschly wanted to find out the relative level of skill proficiency of the students in his math class. He selected the SRA math achievement test because it included both computation problems as well as story problems, it was well normed, and it was both reliable and valid. He secured enough copies of the student protocol or answer forms and a copy of the directions to administer the test from SRA. He then reviewed the test, practiced its administration, and on Monday gave the test to his class. After he scored the tests, he converted them to a standard score, using a table at the back of the administration manual.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

- Standardized teacher-made tests

- In this case, standardized refers to the consistency of the administration and scoring of the test, not to whether or not it is norm-referenced or published.

← - - - - Formatted: Bullets and Numbering

- A standardized teacher-made test may be constructed if the following guidelines are satisfied. The test administration directions and scoring rules must be specified in an explicit and clear fashion. A scripted set of instructions should be considered. All student response protocols should be developed a priori with a guide for scoring the responses. Results should be quantified and reported in a manner that positions or describes the student in reference to an appropriate group. For example, Mrs. Fuchs randomly selected 100 commonly used words from the American Heritage Word Frequency book. She created sentences in which these words would be needed, but were not included (providing a form of a cloze-like test). A total of 100 such sentences were developed and typed onto five sheets of paper. She then wrote a set of directions. First, she listed a general set of considerations in administering the test (e.g., the number of students to test, the format of the testing situation, etc.). Second, she wrote out exactly what the tester should say to the students. Finally, she included a scoring key for determining the accuracy of the student's responses. This key also included a conversion of the student's score to an appropriate metric which made the score comparable across another group of students and similar time period.

#### 9.4 Domain-referenced

9.4.1 The third form of academic assessment is domain-referenced testing.

(a) This approach is characterized by three principal components:

- A well-defined domain from which items are sampled
- A procedure for sampling the items
- A criterion for determining mastery

← --- Formatted: Bullets and Numbering

(b) Criterion-referenced and domain-referenced are terms which can be used interchangeably.

- When criterion-referenced is used, *criterion* refers to the measurement area in which the eventual prediction or comparison will be made. It does not refer to the standard or level of successful functioning for mastery.

← --- Formatted: Bullets and Numbering

(c) The purpose of this type of test is to assess a student's performance upon learning program content.

- If constructed appropriately, domain-referenced tests can be extremely sensitive to a student's learning of program content and to levels change within short periods of time.
- This type of test is frequently used to determine progress within a curriculum. If a student successfully passes a content unit, the student moves to the next unit.
- For example, Mr. Burch wanted to test his class on all key vocabulary terms used in a unit on banking. He systematically wrote down all words in the lesson. A total of fifteen words were identified and incorporated into a word definition test, which he used for both a pretest and a posttest.

← --- Formatted: Bullets and Numbering

(d) How is a *passing* score or acceptable level of performance determined?

← --- Formatted: Bullets and Numbering

- A variety of strategies can be used in establishing a minimum mastery score.
  - Establish the score based on performance of students who are known to be successful.
  - Approximate score (e.g., 90%) based on current learning or performance rates.
  - Give the test to a group of students and establish the passing score as the median or 75th percentile of these students.
  - Identify levels of performance from the natural environment that are necessary for success in those environments.
- (e) How are items selected for criterion-referenced tests?
- Sample all items. This strategy is acceptable if the domain is limited and manageable.
  - Sample representative items. This strategy is appropriate when duplication of several items is present in the total domain. For example, if the skill being tested is vowel diagraphs (ea, ee, ie, ou), not all words including these diagraphs need to be included. Rather, the test would include a certain number of words with each of the diagraphs.
  - Use Random Sampling. This strategy is necessary when the domain is sufficiently large to preclude either of the above two strategies

Formatted: Bullets and Numbering

## | 10 What to Assess

Formatted: Bullets and Numbering

10.1 In this section, we will discuss strategies and guidelines for defining the item selection procedure.

Formatted: Bullets and Numbering

10.1.1 Since the content of assessment or curriculum is discussed in a separate module, it will not be emphasized. The focus will be placed on the selection of item formats for assessing the curriculum.

### 10.2 Selection Type

Formatted: Bullets and Numbering

10.2.1 Selection type items are self-contained and do not allow the student to go beyond the item in answering the problem.

Formatted: Bullets and Numbering

#### 10.2.2 Multiple choice

Formatted: Bullets and Numbering

- (a) Multiple choice items must be answered by selecting a response from a set of possible (alternative) responses.
- (b) Multiple choice items consist of two basic components:
  - A stem (or stimulus) which the student must use to identify an appropriate and completing response
  - A set of alternatives (distractors) from which the correct response must be selected
- (c) The advantage is ease in scoring. The disadvantage is the investment required to construct the test.

(d) Multiple choice items are useful for literal comprehension and application. They are less useful for the assessment of analysis, synthesis, or evaluation skills.

(e) The following guidelines should be followed in constructing multiple choice type items:

- Provide extensive direction in answering the item.
- Select distractors or alternatives to represent the range of possible responses.
- Clearly present the task and the expected response.
- Use a question in the stem whenever possible.
- List the alternatives in a column.
- Avoid the use of negatives in the stem.
- Ensure that all alternatives are constructed in a parallel manner.
- Avoid clues in the alternatives.
- Use distractors of equal difficulty for the naïve student.
- Avoid trick questions.
- Vary the serial position of the correct choice.

Formatted: Bullets and Numbering

#### 10.2.3 True-false items

(a) True-False items present a declarative statement that purports to be an accurate reflection of the state of something. The student is expected to agree (mark it true) or disagree (mark it false).

(b) A number of issues should be considered when constructing and using T/F selection type items.

- Clearly specify a situation which is entirely true or false.
- Confine the length and complexity of the sentence.
- Avoid specific determiners (always, never).
- Avoid ambiguity, vagueness, or open-ended issues.
- Make all statements plausible.
- Direct the statement in a positive manner.
- Avoid double negatives.
- Use the same approximate number of true and false statements.

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

#### 10.2.4 Matching items

(a) Matching items include a list of stimulus features and possible responses that are related on a one-to-one basis. The student is directed to pair them together.

(b) When constructing or using matching type items, consider the following.

- Keep both lists homogeneous.
- Clearly establish the basis for the relationship.
- Keep the list short (from 5 to 8 items in each list).
- Arrange the lists in a logical and orderly fashion.
- Include responses which match more than one item or don't match any.
- Present both lists in the same visual field.
- Direct the student which way to match.

Formatted: Bullets and Numbering

### 10.2.5 Classification

Formatted: Bullets and Numbering

- (a) Classification type selection items provide the student with elements of two sets, and request a categorization of all members of one set within members of the other set.
- When constructing or using classification type selection items, consider the following guidelines.
  - Be certain that the sets are mutually exclusive
  - Include an exhaustive list of elements in a set

### 10.3 Supply type items

Formatted: Bullets and Numbering

10.3.1 Supply type items direct the student to complete the response by providing the correct answer rather than selecting it. The stimulus item provides varying degrees of freedom in structuring the student's response.

### 10.3.2 Cloze items

Formatted: Bullets and Numbering

- (a) In cloze supply type items, the stimulus item contains all but a minimal amount of information that the student is requested to complete. Fill-in-the-blank type items are most commonly used.
- (b) When developing or using cloze supply type items, the following issues should be considered.
- Make certain only one response is correct for each blank.
  - Include enough information in the stimulus to ensure comparability across student responses.
  - Leave the blank space near the end of the sentence.
  - Omit only significant words central to the area of knowledge which is being tested.
  - Use consistent size blanks across all items.
  - Avoid structuring the answer through syntax (i.e., use of singular/plural).

Formatted: Bullets and Numbering

### 10.3.3 Short answer

Formatted: Bullets and Numbering

- (a) Short answer supply type items provide the student with a question or directive to respond with a short phrase, with little structure or constraint on the content of the answer.
- (b) When using or developing short answer supply type items, the following issues should be considered. The question or statement should specify a concise and simple answer. Include rules for responding in the directions desired.

### 10.3.4 Extended answers

Formatted: Bullets and Numbering

- (a) Extended answer supply type items present a stimulus item in the form of a question or directive and require the student to provide an extended answer that is longer than a phrase or one sentence. Few restrictions are provided for structuring the response. This is useful for attending to higher order issues (i.e., analysis, synthesis, comparisons, and evaluations).

- (b) When using or writing extended answer items, keep the following guidelines in mind. ← Formatted: Bullets and Numbering
- Make the statement or question clear and concise.
  - Provide the weights or points for each of the items.
  - Provide clear directions for structuring the student's response.

| 11 How to Assess Academic Performance ← Formatted: Bullets and Numbering

11.1 In this section, we will discuss some guidelines and considerations related to assessing academic performance. Some of this material was discussed briefly in the section on assessment fundamentals. However, the emphasis in this section will be on the assessment of academic performance.

11.2 Administration of the test

11.2.1 Regardless of the type of test being administered or the subject area being assessed, a number of test administration considerations should be kept in mind to improve the accuracy and utility of the results.

11.2.2 Timing of the test

(a) Many assessments do not examine how much time a student requires to complete a test or activity, limiting measurement to number or percentage correct and incorrect. ← Formatted: Bullets and Numbering

- Learning is not just the acquisition of accurate information or skills. Learning also involves the student's fluency or proficiency at a task or skill.
- For example, a student may be 100% accurate at filling out a job application, but if it takes the student more than 30 minutes to complete it, s/he may not be able to compete with more efficient applicants.

(b) Recording how long it takes for a student to complete a skill accurately is important for a number of reasons. ← Formatted: Bullets and Numbering

- Having time allows the teacher to calculate a student's rate of performance. This calculation simply expands the summary information available.

- Formula:

$$\frac{\text{\# of behaviors}}{\text{amount of time}} = \text{rate}$$

- Examples: Words read per minute ← Formatted: Bullets and Numbering

- Number of tasks completed per hour
- Number of widgets assembled per hour

- Having time allows the teacher to determine the duration of time required to complete a task.

- Duration refers to the amount of time from start to finish required to complete a single task or activity. ← Formatted: Bullets and Numbering

- For example, how long it takes to balance a checkbook.
- Timed assessments allow the teacher to evaluate relative proficiency across students.
  - Allowing for an evaluation of instructional effectiveness and efficiency
  - It is possible to evaluate whether sufficient *opportunity to respond* has been provided for both the highest and lowest performing students.

Formatted: Bullets and Numbering

#### 11.2.3 Developing standardized directions

- (a) When administering academic assessments, uniform directions or instructions should be provided.
- (b) The purpose of such standardization is to:
  - direct the students to respond to the essential features of the test question. This assistance increases the likelihood that the content of the test is indeed valid (i.e., that the student is responding to the relevant features of the item).
  - ensure comparability across all students. That is, to insure that no one's performance is differentially affected by the manner in which the test is given
  - ensure comparability across time. As test results are compared from pre- to post-test outcomes, this standardization of instruction attempts to increase the probability that the change is not a result of variation in administration.

Formatted: Bullets and Numbering

#### 11.2.4 Physical conditions

- (a) Since the purpose of testing is often to determine *optimal performance* it is important that physical conditions be considered during assessments. The validity of a test is often times influenced by the conditions under which the assessments are administered.
- (b) The following factors should be considered:
  - Lighting of the room must be adequate to read and write, but not distracting.
  - Noise must be low threshold, providing an undisturbed assessment.
  - Interruptions must be avoided to keep the student's attention on the testing.
  - Comfortable surroundings (heating and ventilation) must be present to maximize performance.
  - Materials must be present and organized to minimize waiting on the part of the student. This is especially important in group testing and with students who are easily distracted by long transitions or preparation activities.

Formatted: Bullets and Numbering

### 11.3 Scoring the test

11.3.1 Prior to the actual administration of the test, plans should be made for the scoring of student responses.

11.3.2 A number of issues need to be considered when planning for the scoring of student performance on tests or other assessment instruments.

- (a) Who will score the test?
- (b) How to score missing or unanswered items--are they to be treated as incorrect or absent?
- (c) What type of scoring keys will be established?
- (d) Will interscorer reliability be determined? If so, what procedures will be used? This simply examines the consistency of results scored by different individuals. Low interscorer reliability can be an indication of a high degree of judgment involved in the scoring of a test.

Formatted: Bullets and Numbering

11.3.3 In preparing to score a test, it is important to determine the unit of analysis.

- (a) This issue refers to the size or amount of a learner's performance that will be used to determine a score.
  - The unit of analysis is important in terms of the scaling of the test and determining the range of possible scores.
  - Many tests may comprise subunits, each with a subtotal. The teacher will need to determine whether the student's score will be reported for each subtest or summed across subtests and reported for the total test.
  - For example, A standardized test that accompanies a commercially prepared social studies curriculum contains five major units; each unit has five assessment items. The publishers state that four correct out of five are needed for each unit to be considered passing (or mastery). Because this small number of items does not adequately assess the curriculum for each unit, serious scoring and validity problems exist. So Mr. McNabb has decided to base mastery on the total test score, using 21 out of the total 25 as the criterion for mastery.

Formatted: Bullets and Numbering

11.3.4 Scaling properties

- (a) Scaling refers to the range of number of items used to score a student's performance.
- (b) The guiding rule should be to maximize the scale as much as possible. For example, Mr. Staley has administered a spelling test to his top group of students. He sampled words from the newspaper, administering one word in rolling dictation every 10 seconds, for a total time of 2 minutes. He decided to score the words by counting the number of correct letter sequences, rather than the total number of words spelled correctly. The range for the former unit (correct letter sequences) was 56 to 223, while for the latter unit (words correct), the scores ranged from 6 to 17. Clearly, greater sensitivity exists with correct letter sequences. Other examples include the use of correct digits vs. problems in math, and number of words written vs. number of sentences in written expression.

Formatted: Bullets and Numbering

### 11.3.5 Instructional relevance

- (a) Instructional relevance refers to the nature of the unit for scoring test responses.
- (b) The unit of behavior should be meaningful and have content validity of its own. For example, A sample of Donovan's writing was analyzed to determine if he had any significant *deficits* (i.e., his writing contained significantly simpler syntax and structure than used by age/grade peers). His writing was analyzed for several variables: the number of syllables per word, the number of mature words, the number of correct word sequences, and the mean number of words per sentence-unit. Of all of these indices, the last one is probably the most clearly related to a measure of *complexity* and syntax and therefore has the greatest instructional relevance.

← - - - - Formatted: Bullets and Numbering

### 11.3.6 Sensitivity to change

- (a) To use data in the formative evaluation of programs, it is important that the data reflect changes in student performance over relatively short periods of time. For example, in the writing situation above, in which a student's writing sample was scored for the number of words per sentence-unit, a problem may exist in how sensitive the measure is to performance change. Especially over brief periods of time, it is not unusual to find that students increase the number of words per sentence-unit at a maximum rate of only 1.5 words per year. Instructional relevance may conflict with sensitivity.

### 11.3.7 Cost-effectiveness

- (a) Clearly all measurement data must be useful and usable in a manner that facilitates their incorporation into the daily classroom routine.
- (b) The following represent questions that assess the cost-effectiveness of a measurement system.
  - How much time is taken in the administration, scoring, and reporting of the scores?
  - How much money is expended?
  - What is the impact on the program for making decisions with and without the data? Advantages? Disadvantages?
  - What are the short- and long-term effects of daily assessment procedures?
  - What alternative procedures are available? What information will be gained or lost?

← - - - - Formatted: Bullets and Numbering

## 11.4 Reporting the outcome

### 11.4.1 Using percentage

- (a) Percentage is useful for two basic reasons:
  - Makes data across different test or subtests comparable when the number of items are not equal. For example, Mr. Wehby wanted to compare Paul's performance on the math test from the latest unit against his performance on the previous unit. Paul's score was the

← - - - - Formatted: Bullets and Numbering

same for both tests—an obtained score of 70. However, the latest test included only 75 items while the earlier test had a total of 100 items. To make a fair comparison across two situations that have presented the student different opportunities to respond, Mr. Wehby has decided to use percentage correct. Paul’s performance on the first test was 93% correct, and on the later test, only 70% correct. Clearly he had done better on the previous test.

- Focuses performance on accuracy. Percentage correct is essentially a measure of accuracy. It is calculated by the following formula:

$$\frac{\# \text{ correct}}{\# \text{ correct} + \# \text{ incorrect}} \times 100 = \%$$

- (b) When using percentage, a number of cautions should be kept in mind.
  - Percentage can be quite insensitive when there are fewer than 10 items. A one-item change will cause a large change in percentage.
  - The distribution of students’ scores is not even, and not directly comparable across the entire range of raw scores. The difference between raw scores near the middle of the distribution (from 40 – 60 percent) are not the same as the comparable difference of the raw scores near either of the tails (10 to 20% and 90 to 100%)

← - - - - Formatted: Bullets and Numbering

#### 11.4.2 Using absolute correct and incorrect scores

- (a) When the number of items a student gets correct or incorrect is of importance, reporting straight correct and incorrect raw scores may be useful, and are, especially useful when the focus is on the amount of learning or number of tasks completed.
- (b) A few cautions about the straightforward reporting of raw scores need to be noted. A student’s score is not anchored or related to the average or variation of scores earned by other students. When students have different opportunities to respond, comparisons on number correct or incorrect cannot be made. Raw scores do not always reflect the number of opportunities to respond.

#### 11.4.3 Using rate

- (a) Rate or frequency of performance is one of the most preferable scores for reporting performance.
- (b) Rate scores have the following advantages: They may discriminate different proficiency levels (i.e., how fast or fluent a student is performing) that are not visible with accuracy alone; and they ensure comparability of data by controlling for the *opportunity to respond*.

← - - - - Formatted: Bullets and Numbering

## | 12 What Skills to Assess

← - - - - Formatted: Bullets and Numbering

### 12.1 Introduction

- 12.1.1 In the previous part of this module, the focus was on the assessment of academic skills. The discussion centered on arranging materials and settings in a manner that allow the student to show certain academic performance and, in turn, allow teachers to evaluate student learning and instructional effectiveness.
- 12.1.2 Many of the general principles and guidelines discussed in that section may be applied to the assessment of social skills; for instance,
- (a) Including time in assessments;
  - (b) Analyzing *opportunity to respond*;
  - (c) Insuring measurement of reliability and validity; and
  - (d) Assessment approaches.
- 12.1.3 A major difference in this part of the module will be a focus on the direct observation of social behaviors within the context of *natural* setting conditions.
- (a) We will look at social skills within the context of the antecedents and consequences of the setting in which they occur.
  - (b) Behaviors do not occur in a vacuum. The term *social skill* usually means that the behaviors are occurring within some kind of social context.
- 12.1.4 Another difference will be a focus on interpersonal, school, family, and community behaviors.
- 12.2 To start, we need to look at social behaviors in an objective and positive manner (i.e., we must work on removing our own biases and stereotypes).
- 12.2.1 One way this can be accomplished is by examining social behaviors with respect to their frequency of occurrence and to the setting in which they occur.
- 12.2.2 A simple three-part categorization format might be considered .
- (a) Behavioral excesses are those social behaviors that a student exhibits that occur too much or too often. These excesses could be behaviors that need to be eliminated from the student’s behavioral repertoire, such as hitting others, throwing objects, etc. These excesses could also be behaviors that are acceptable in their basic form, but are unacceptable at their present rates. For example, talking excessively, raising hand too often, driving too fast, etc.
  - (b) Behavioral deficits are analogous to excesses except that these are behaviors that occur at levels that are too infrequent and generally need to be increased; for example, turning in homework on time, raising hand before asking a question, etc.
  - (c) Behavioral assets are behaviors that a student currently possesses that are functionally and socially useful to them. These are important behaviors to assess because they might function as acceptable behaviors that could be

taught to replace undesirable ones, and because they help teachers and others see the students as a person with strengths.

- 12.2.3 Teachers can remain more objective and positive by defining what social behaviors they are assessing or teaching in *operational*, or observable, terms.
- (a) This means describing what is being assessed in terms that another person can reliably identify the presence of that behavior. For example, “getting along with others” is not an operational definition because it could be used to label a wide variety of social skill behaviors. One operational definition for this general area might be “talking with two peers for at least 15 minutes” or “using the computer for equal amount of time when working on math problems with one other peer.”

- 12.2.4 A third strategy for remaining objective during the assessment of social behaviors is to apply the *fair pair* principle.

- (a) Like focusing on behavioral assets, the fair pair principle emphasizes the identification and assessment of behaviors that will need to be eliminated.
- (b) It is important to keep in mind that the initial and continued elimination or reduction of behavioral excesses will be facilitated if a behavior can be identified to replace the excess and if that behavior will result in positive reinforcement for the student.
- (c) We cannot expect a behavior to be removed if the student does not have a behavior to replace it. It is sometimes useful to think of behaviors as being communicative. If a student only has one way to communicate (i.e., set the occasion for reinforcement), s/he will continue to use that behavior even if it is socially undesirable.

Formatted: Bullets and Numbering

- 12.2.5 A final guide to help us remain positive and objective in our assessment of social behaviors is to focus on socially desirable behaviors

- (a) This statement implies that we need to assess and teach behaviors that are of functional use to the student if they are increased, decreased, and/or maintained.
- (b) We need to insure that the behaviors we assess to be taught or reinforced will cause the student to be perceived more favorably in the social environment and to be more likely to be reinforced by relevant others.

### 12.3 What social skills can we assess?

12.3.1 The answer is simply any social skills that can be measured.

12.3.2 Rather than categorizing the many behaviors that can be assessed, in this situation we will discuss the characteristics of those skills that can be assessed, changed, or taught and the setting in which those behaviors are observed. There are six simple components to be assessed (display T-12.3.2).

- (a) Setting--the place or conditions under which a social skill is observed or expected to occur. For example, a teacher might be interested in assessing the students' ability to ask for materials in a socially appropriate manner such as "please, may I borrow..." The setting conditions might be during a group science activity, at the cash register of a store, or in the industrial arts area. This behavior is appropriate across a number of settings, whereas hand raising may be functionally appropriate only during large group discussions at school or other similar settings
- (b) Pre-disposing factors--those skills, characteristics, or conditions that are associated with a specific student and that affect his/her ability to change or learn a social skill. Often times we look at a student's prior learning history to help us identify contributing influences. For example, being able to learn how to say "please, may I borrow" may be affected by a student's knowledge about borrowing, or his/her ability to identify (learning history) when to use that type of statement.
- (c) Precipitating or triggering factors--those things or events that signal or indicate to a student to engage in a given social behavior. These triggering factors for most students signal some kind of socially appropriate adaptive response, whereas, these same factors may trigger maladaptive behaviors from other students. For example, when he is out of paper, Doug responds by saying "please, may I borrow", but Molly, on the other hand, will grab paper from a neighbor when she needs another piece.
- (d) Contributing factors--those things or events that follow a student's actions that serve to increase or decrease it. Again, these contributing factors may be associated with adaptive as well as maladaptive behaviors. For example, Doug says, "please, may I borrow" because many students and adults will loan him what he wants and will say something positive in return. Molly grabs items because she also gets what she wants...and she manages to get some kind of reaction or attention from the person she has grabbed from.
- (e) Time--when, how long, or how fast a given social behavior is observed. For example, Terry says, "please, may I borrow" but she says it after she has already grabbed it from her neighbor. Dave has a similar timing problem in that he says the phrase approximately 12 times an hour. In his case, the social skill is accurate, but it occurs too often.
- (f) Social behavior--the actual behavior to be changed or taught. For the purposes of this module, a social behavior can be an appropriate social skill or an inappropriate behavior that needs to be eliminated. It functions as a means of moving effectively through the social environment.

← - - - Formatted: Bullets and Numbering

12.3.3 A social behavior has six different dimensions on which we might focus our assessment (display T-12.3.3).

- (a) Topography--the shape or characteristics that a behavior has. For example, shaking hands has a variety of topographies: a formal, business-related shake; an informal, friendly greeting; a "high-five"; etc.

- (b) Duration--how long a behavior lasts. For example, does the handshake last for the appropriate length of time of about 2 or 3 seconds, or does it last for 10 seconds?
- (c) Latency--how long it takes for the behavior to be initiated after an appropriate triggering event has occurred. For example, after Mr. Compton says "Hello, nice to meet you," does Don shake hands within 2 seconds or does he wait until the initial greetings are concluded and the topic has changed?
- (d) Intensity--the amount of force a given social behavior might be characterized. For example, a socially acceptable handshake is firm and steady. The dead-fish grip or the iron-man clamp are generally perceived as socially unacceptable.
- (e) Locus--where the behavior occurs. For example, Jeff has a good sturdy handshake, but he only uses it when he knows the other person and when he is at home.
- (f) Frequency = how often or how fast the behavior occurs. For example, Bill likes to shake hands and will sometimes shake hands 10 to 15 times an hour.

12.4 What we assess can cover a broad range of social skills and behaviors. Rather than categorizing these behaviors, we have discussed some general guidelines that might help us identify what to assess in an objective and systematic manner.

### | 13 How to Assess Social Skills

← - - - - Formatted: Bullets and Numbering

- 13.1 Following the focus of the previous section, we will discuss how to assess social skills from a generic approach. We will emphasize procedures and guidelines that should be applicable across a wide variety of social skills and behaviors. These procedures and guidelines should also be useful in adapting commercially prepared assessment devices to your individual assessment purposes. Following the direct interventionist approach that serves as a foundation for our work in special education, we will continue to focus on observable behaviors.
- 13.2 First, a reminder of some basics. We already mentioned that we will focus on observable behaviors, but this is also related to being performance-based. We want to see social skills in action. Therefore, we must assess social skill action rather than promises (i.e., observe the student when greeting someone to see if they shake hands; do not ask the student, "What do you do when you greet someone?" out of context). As discussed earlier, we will assess social skills in a formative fashion. To get the best measure of social skill acquisition and maintenance, we should assess regularly over time rather than at the beginning and end of an instructional time period.

### | 14 Indirect Observation Strategies

← - - - - Formatted: Bullets and Numbering

- 14.1 There are a number of ways to assess social skills without directly observing the social behavior.

14.1.1 Anecdotal reports can take two basic forms.

(a) Written records

- These written records represent a person's interpretation and description of a student's social skills. However, the reader should be careful about the meaning derived from such reports.
- A reader's interpretation is based on the *writer's* interpretation of the student's behavior. The reader is developing second-hand impressions. Some written reports are based on interviews with other relevant individuals in the student's life which dilutes the validity even further.
- An author's written impressions have self-fulfilling effects and may cause the reader to infer more than is appropriate. For example, when Mr. Wall read about Beth's "aggressive behavior", he assumed that Beth was physically aggressive (i.e., hit others and destroyed property). However, Beth is actually only verbal aggressive (i.e., she talks-back and tends to call people derogatory names).
- Written reports are frequently based on a student's prior behavioral history, and may or may not reflect the student's current social skill levels.

Formatted: Bullets and Numbering

(b) Structured and unstructured verbal interaction

- Like written reports, these indirect assessments provide information about a student's social skills through another person's interpretations and descriptions. Similar use precautions apply.
- Verbal reports are frequently shaped or influenced by the interaction between the people discussing the student. For example, a person's social or professional status may cause the speaker to embellish a description to meet their expectations about the listener.

Formatted: Bullets and Numbering

14.1.2 Checklists, inventories, and rating scales are other commonly used formats for assessing social behaviors.

(a) These assessment devices have the following characteristics:

- Usually a first step in assessing social skills and behaviors
- Frequently used to identify social skill problems rather than strengths
- Often used to identify feelings and attitudes as well as personality characteristics (e.g., self-concept, self-esteem)
- Generally filled out by someone who has observed the student over time and knows the student well such as a parent, teacher, peer, or the student him- or herself
- Behavioral statements or descriptions are presented and the rater must judge to what degree the statement describes the student, based on opinion rather than direct measures.

Formatted: Bullets and Numbering

(b) A number of weaknesses are associated with checklists and rating scales that affect their reliability and validity.

- Assessment outcomes are subjectively based and dependent upon the raters observation skills, memory, and ability to judge or rate behavior.
- Tend to emphasize the assessment of inappropriate behaviors

Formatted: Bullets and Numbering

- Frequently assumed to measure “hypothetical constructs” (i.e., non-specific personality states that are assumed to exist though not directly observable)
  - Because they tend to be general screening devices with a limited sample of response choices, they may not accurately reflect the specific behavioral strengths and weaknesses of a student.
  - Tend not to be directly translatable into educational programming
  - Usually requires additional in-depth assessment and evaluation.
- (c) Checklists and rating scales should be used if
- a general screening is desired;
  - direct measurement procedures cannot be implemented; and
  - subjective interpretations can be tolerated.

← - - - - Formatted: Bullets and Numbering

14.1.3 Sociograms and other similar procedures are sometimes used to assess social skills.

- (a) In general, these procedures require students to rate themselves along a single criterion or continuum of descriptors in relation to their peers or social environment.
- (b) Intended purposes:
- To identify work habits
  - To determine relative popularity/unpopularity, acceptance/rejection, impulsivity/compulsivity, cooperation/noncompliance
  - To identify level of self-concept or self-esteem, self-image
  - To identify feelings/attitude toward school, home, community, school, etc.
- (c) Sociograms generally ask the student to
- name other students who best match some descriptor (e.g., name three students who are good at math); and
  - rank order themselves and/or students (e.g., most to least popular).
- (d) Cautions
- Confidentiality must be maintained.
  - Ratings are subjectively based and evaluated.
  - Do not translate into teachable objectives.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

14.2 The above indirect strategies for assessing social skills and behaviors can be useful. However, a clear understanding of their limitations and purpose is necessary.

14.2.1 Their use tends to be limited to the general screening of student social skills so that relative ratings or comparisons can be made. They are not useful in mastery determinations of the planning and monitoring of interventions.

14.2.2 Because they are indirect measures, they are characterized by subjective interpretation.

| 15 Direct Assessment Strategies of Social Skills

Formatted: Bullets and Numbering

15.1 In this section of the module, we will discuss basic strategies for assessing social skills through the use of direct observation procedures.

15.1.1 As a reminder, the effective and reliable use of direct observation systems is dependent upon

- (a) a clear and operational definition of what social skill is being assessed
- (b) formative evaluation; and
- (c) an understanding of the direct relation between the social behavior and the setting or environmental conditions.

Formatted: Bullets and Numbering

15.1.2 *Direct* refers to assessing a social behavior as it is occurring.

- (a) Unlike previously mentioned procedures, subjective interpretations and judgments are reduced.
- (b) Objective and descriptive data are provided, thus reducing the need to infer some underlying, unobservable hypothetical construct that may not be directly manipulable.

Formatted: Bullets and Numbering

15.2 Types of data recording—Data recording can be conducted by two basic means: continuous recording or time sampling.

| 15.2.1 Continuous recording

Formatted: Bullets and Numbering

- (a) A set amount of time is specified before the observation session begins.
- (b) All occurrences within this time frame are recorded.
- (c) Three main types of data can be collected.

- Event
  - The frequency or occurrence of a behavior
  - Reported as a frequency (i.e., # of occurrences) or a rate (i.e., # of occurrences per unit of time)
- Duration
  - The length of time a behavior is displayed, from its initiation to its termination
  - Reported as time (in minutes and seconds) for each occurrence, average occurrence, or sum of occurrences (which can be used to calculate percentage of time displaying behavior)
- Latency
  - The elapsed time between the onset of a stimulus (such as the teacher giving a direction) and the initiation of a behavior (such as the student starting the assignment)
  - Reported as time for each occurrence or average time

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

- (d) A fourth type of data is intensity, or severity
  - The degree to which a behavior is displayed
  - Is used less often because it is generally difficult to operationalized and therefore not applicable to direct observation

- Indirect or qualitative measures are often better suited, but the cautions stated previously should be considered

### 15.2.2 Time sampling

← --- Formatted: Bullets and Numbering

- (a) Time sampling is useful if it is not practical to conduct a continuous observation.
- (b) The observation period is broken down into a series of shorter intervals.
- (c) Rather than record frequency, duration, or latency, the observer records whether or not the behavior was displayed during the interval.
- (d) There are three ways to record the presence or absence of a behavior:
  - Whole interval—The behavior coded as present if it is displayed for the entire interval
  - Partial interval—The behavior is coded as present if it is displayed at any point during the interval; this is useful for non-continuous or sporadically displayed behavior.
  - Momentary—The behavior is coded as present if it is displayed at the moment the interval begins.

## 15.3 Functional behavioral assessment (FBA)

15.3.1 An FBA is conducted in order to determine possible causes (antecedents) and effects (consequences) of a disruptive or undesirable behavior. This information can then be used to plan positive behavioral interventions in which prosocial behaviors are used to replace the undesirable one(s).

15.3.2 Once the problem has been verified (i.e., the seriousness of the behaviors is corroborated by others), the problem behavior should be operationally defined.

(a) Information should be collected on

← --- Formatted: Bullets and Numbering

- When the behavior does/does not occur (e.g., specific days or times of the day, during certain activities or classes)
- Where the behavior does/does not occur (e.g., lunchroom, classroom)
- Conditions when the behavior does/does not occur (e.g., when doing individual seatwork)
- Number and identity of people present when the behavior does/does not occur (e.g., Mrs. McCorkle the math teacher, groups of 10 or more people)
- Events that occur before the behavior (e.g., teacher gives the student a direction)
- Events that occur after the behavior (e.g., student is sent to lock up)
- Other behaviors associated with the occurrence of the problem behavior (e.g., negative peer interactions)

15.3.3 Data should now be collected on possible functions of the behavior.

- (a) Data collection is conducted using a variety of indirect and direct methods. The specific methods used should be tailored to the behavior of interest and the resources available.
- (b) There are several ways of categorizing behavior, for illustration we will present two. ← - - - Formatted: Bullets and Numbering
- First, it is important to determine the function of the behavior.
    - To get something—The behavior may serve the purpose of getting something for the student. The “things” may include (but are not limited to): attention (from peers or adults); tangible goods (e.g., food, stickers, money).
    - To avoid something—The behavior may serve the purpose of avoiding a situation that the student finds unpleasant or threatening. Avoidance behavior occurs before the undesirable situation (e.g., every day, 10 minutes before math class, Edna develops a stomachache.). Escape behavior occurs after the undesirable situation has been initiated (e.g., during seatwork, Edna throws her pencil at the teacher, getting sent out of the room.). It should be noted that the function of avoiding something, may also have an underlying function. That is, the student may wish to avoid the situation because he/she is frustrated, bored, tired, hungry, etc. Determining *why* the student wishes to avoid the situation is as important as determining that the student wants to avoid the situation.
  - Second, it is important to determine whether the problem behavior represents a skill deficit or a performance deficit. A skill deficit occurs when the student does not know how to perform a prosocial behavior (e.g., Ron does not know how to give someone an appropriate compliment). The problem behavior has replaced the prosocial one to achieve the desired function. A performance deficit occurs when the student knows how to perform the behavior, but does not know strategies for determining when to perform it. It can also occur if the student chooses not to perform the prosocial behavior (e.g., if the problem behavior is easier, gets better results, etc.). A performance deficit can lead to a behavioral excess. For example, Willy knows how to shake hands, but is unsure of when it is appropriate to do so. As a result, he shakes hands very frequently.

## 15.4 Functional analysis and functional relations

15.4.1 Functional analysis is a procedure for directly observing a student’s behavior and the setting conditions associated with the behaviors.

- (a) Environmental events are systematically manipulated so that relations to the behavior can be experimentally evaluated. ← - - - Formatted: Bullets and Numbering
- (b) It consists of five major components:
- The social skill or behavior of interest
  - Antecedent or preceding events to the behavior

- Consequence or following events to the behavior
  - The function, or purpose, of the behavior
  - Time
- (c) Cause-effect relations can be developed.
- (d) Predictable behavioral chains can be identified.

← - - - - Formatted: Bullets and Numbering

15.4.2 Functional relations are derived from an analysis and the testing of information from a functional analysis.

- (a) A functional relation is defined as a statement describing a predictable relation between two variables, usually the student's behavior and antecedent or consequence events.
- The behavior of interest is called the dependent variable
  - The affecting variables (i.e., antecedents or consequences) are called independent variables.
  - A functional relation exists when a predictable statement can be made about the relation between dependent and independent variables.
- (b) Functional relations are first derived from testable explanations. A testable explanation is a specific statement about possible functional relations in which observable dependent and independent variables are indicated.
- Example: "Whenever the teacher has a transition between lessons that exceed five minutes, the number of talk-outs increases threefold."
  - Example: "If Ellen sits next to Jim during math class, Jim does not finish his work and will talk with Ellen."
- (c) A danger is the creation of explanatory fictions. An explanatory fiction is a non-specific statement about possible functional relations in which the independent or dependent variables are not observable or manipulable. Sometimes the independent variable is a restatement of the dependent variable. Examples include:
- "Whenever Calvin is hyperactive, he talks back at the teacher." Because hyperactivity can take so many different forms, it is not observable.
  - "Stu fails to make friends because he is emotionally disturbed."
  - "Bluto doesn't get along with adults because he is a juvenile delinquent."
- (d) A testable explanation is a functional relation if a systematic manipulation of the independent variable produces a predictable change in the dependent variable. If an undesirable, unpredictable, and/or non-adaptive functional relation is identified, the teacher's task is to change the testable explanation or develop a more appropriate one. For example
- If we change Ellen and Greg's seats so that Greg is next to Jim, and if Jim does not talk with or touch Greg, then our testable explanatory statement about Jim sitting next to Ellen is a functional relation.
  - If Jim talks with or touches Greg, then our testable explanation will need to be adjusted so that it will account for Jim's behaviors and the setting conditions.

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

← - - - - Formatted: Bullets and Numbering

- (e) Functional relation statements are not only useful in assessing social skill behaviors, but also in identifying a possible place to start interventions or teaching strategies. We can identify whether acceptable replacement behaviors exist in the student's behavioral repertoire and whether these behaviors occur at the right times and places.

## 15.5 Comparison of FBA and functional analysis

### 15.5.1 Similarities

- (a) Both examine the function of the behavior.
- (b) Both explore the antecedents and consequences of the behavior.
- (c) Both involve setting variables (e.g., time of day, type of activity that behavior occurs in).

← - - - - Formatted: Bullets and Numbering

### 15.5.2 Differences

- (a) FBA can incorporate a variety of assessment strategies; functional analysis is a particular type of assessment strategy.
- (b) A functional analysis can be part of an FBA.
- (c) FBA does not involve the systematic manipulation of independent variables, whereas functional analysis does.

← - - - - Formatted: Bullets and Numbering

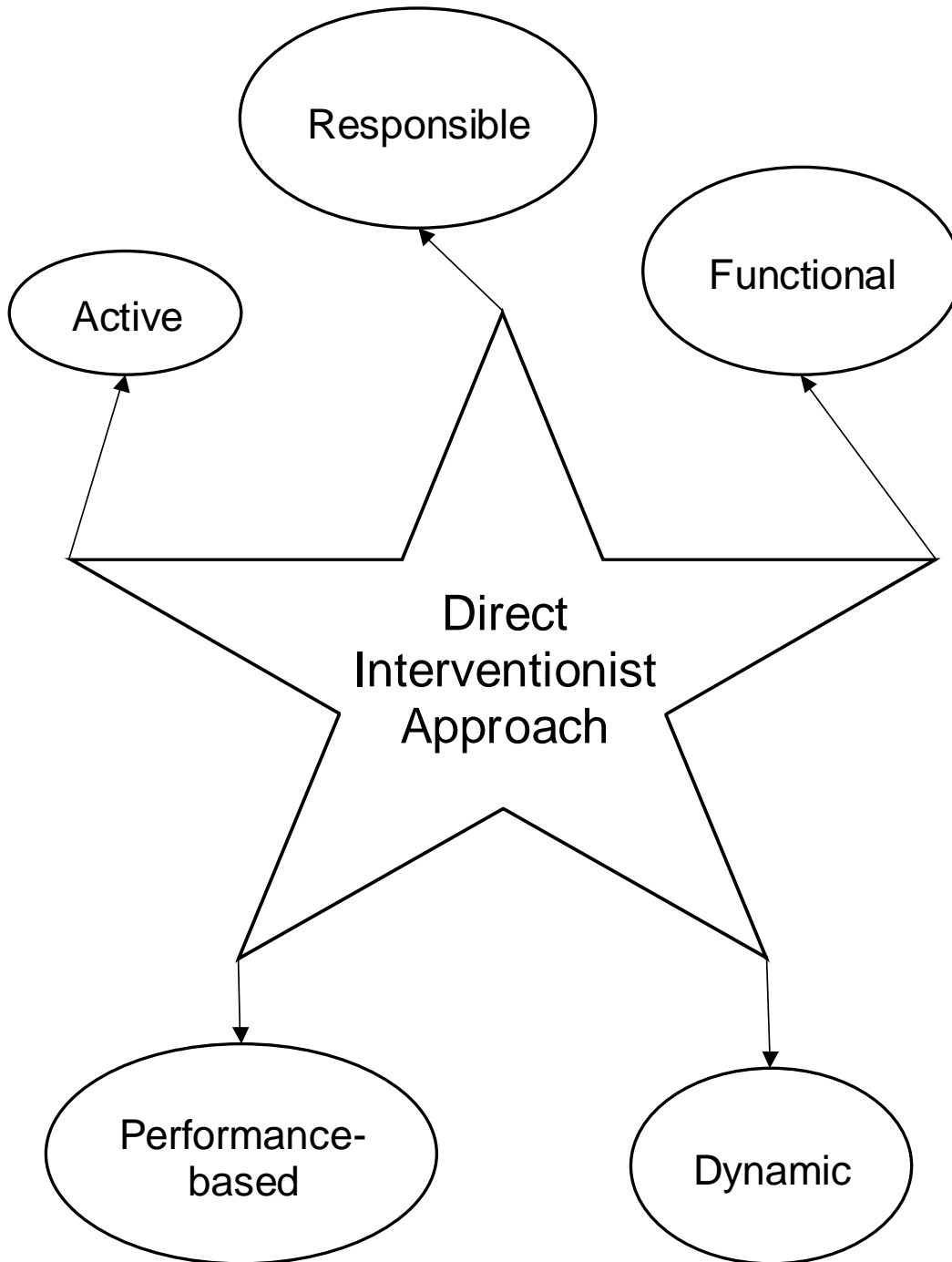
## References

- Anastazi, A. & Urbana, S. (1997). *Psychological testing*. Upper Saddle River, NJ. Prentice Hall.
- Carter, M., & Kemp, C. R. (1996). Strategies for task-analysis in special education. *Educational Psychology, 16*(2), 155-170.
- Gable, R. A., Quinn, M. M., Rutherford, R. B., Howell, K. W., & Hoffman, C. C. (1998). *Addressing student problem behavior—part II: Conducting a functional behavioral assessment (3rdEd.)*. Washington, DC: American Institutes for Research, Center for Effective Collaboration and Practice.
- Hintze, J. M., & Shapiro, E. S. (1995). Best practices in the systematic observation of classroom behavior. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology III*. Washington, DC: National Association of School Psychologists.
- Howell, K. W., & Davidson, M. J. (1997). Programming: Aligning teacher thought processes with the curriculum. In J. W. Lloyd, E. J. Kameenui, & D. Chard (Eds.) *Issues in Educating Students with Disabilities*. Mahwah, NJ: Lawrence Erlbaum.
- Howell, K. W., Fox, S. L., Zucker, S. H., & Moorehead, M. K. (2000). *Resources for implementing curriculum-based evaluation*. Belmont, CA: Wadsworth/Thompson Learning.
- Howell, K. W. & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making (3rd Ed.)*. Belmont, CA: Wadsworth/Thompson Learning.
- Kame'enui, E., Simmons, D., & Cornachione, C. (2000). *A practical guide to reading assessments*. Washington, DC: U. S. Department of Education.
- Kaplan, J. S. (2000). *Beyond functional assessment: A social-cognitive approach to the evaluation of behavior problems in children and youth*. Austin: Pro-Ed.

- Kratochwill, T. R., Sheridan, S. M., Carlson, J., & Lasecki, K. L. (1999). Advances in behavioral assessment. In C. R. Reynolds & T. B. Gutkin (Eds.) *The handbook of school psychology (3rd Ed.)*. New York: John Wiley & Sons, Inc.
- National College of Juvenile and Family Court Judges (1988). *Learning disabilities and the juvenile justice system*. Reno, NV: author.
- Nelson, C. M., Rutherford, R. B., & Wolford, B. I. (1987). *Special education in the criminal justice system*. Columbus, OH: Merrill.
- Salvia, J. & Ysseldyke, J. E. (1995). *Assessment (6th Ed.)*. Boston: Houghton Mifflin.
- Sattler, J. (1991). *Assessment of children: Revised and updated third edition*. San Diego: Author.
- Suzuki, L. A., Meller, P. J., & Ponterotto, J. G. (Eds.). (1996). *Handbook of multicultural assessment: Clinical, psychological, and educational applications*. San Francisco: Jossey-Bass.

## **TRANSPARENCIES**

This transparency presentation summarizes the content of this module. It does not include all the information contained within this module and should be used to supplement its implementation.



T-1.2

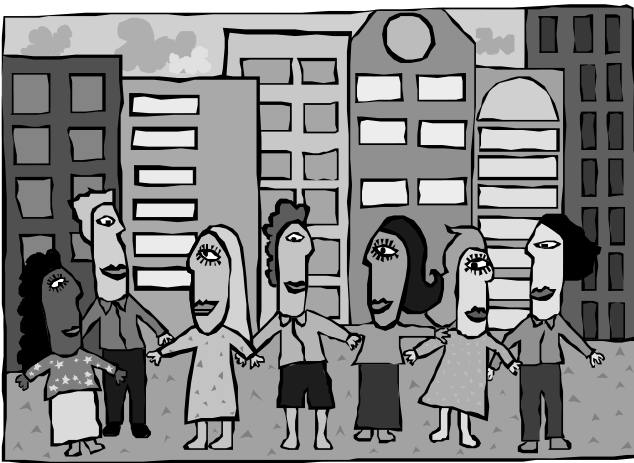
# Critical Components of the Instructional Setting



Student



Teacher



Social Environment of the Adolescent/Young Adult

T-2.1.3

# Stages of Learning

Acquisition



Fluency or Proficiency



Maintenance



Generalization



Adaptation

# Eight Steps in the Systematic Instruction Model

Assess student's strengths and needs

Set long-term objectives

Set short-term objectives

Write an instructional plan

Write a measurement plan

Implement both plans

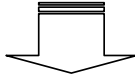
Modify instruction based on data

Evaluate effectiveness of plan

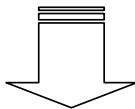


# *Levels of Assessment*

Written Documents  
or Archives



Interviews and  
Verbal Reports

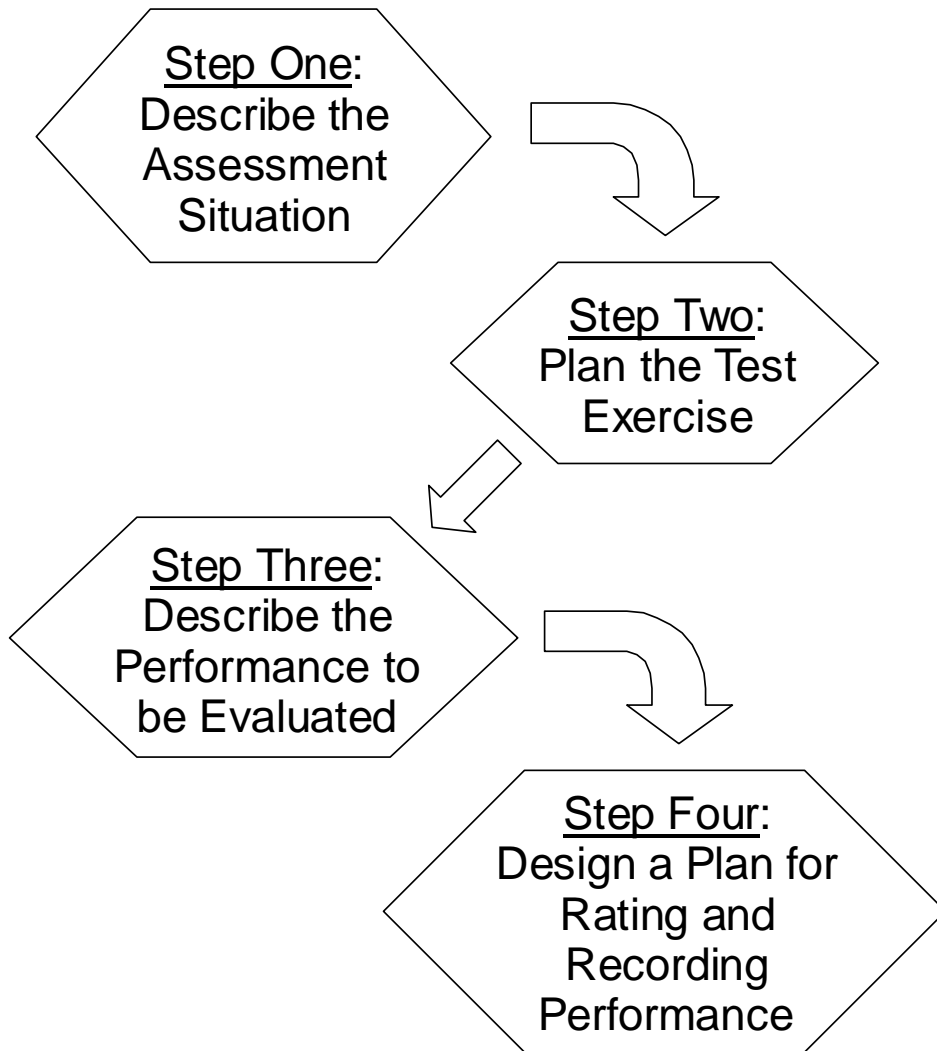


Formal and Informal  
Written Tests



Direct Observation

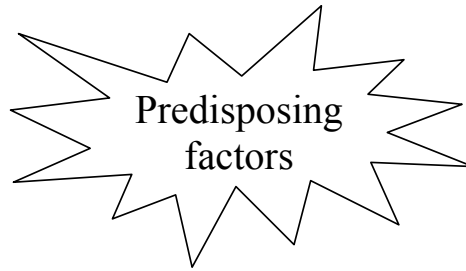
# Assessment Guidelines



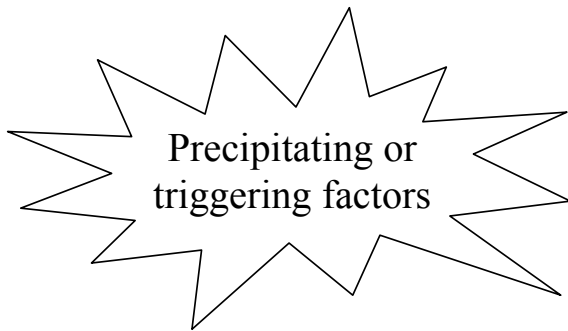
# What to Assess?



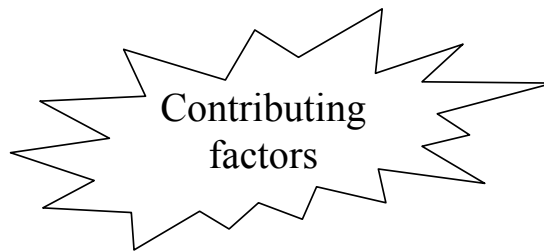
Setting



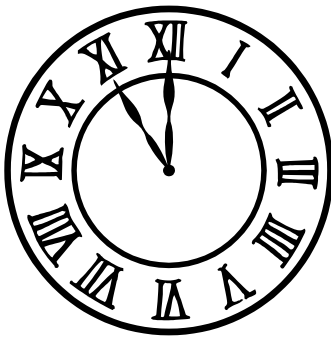
Predisposing  
factors



Precipitating or  
triggering factors



Contributing  
factors

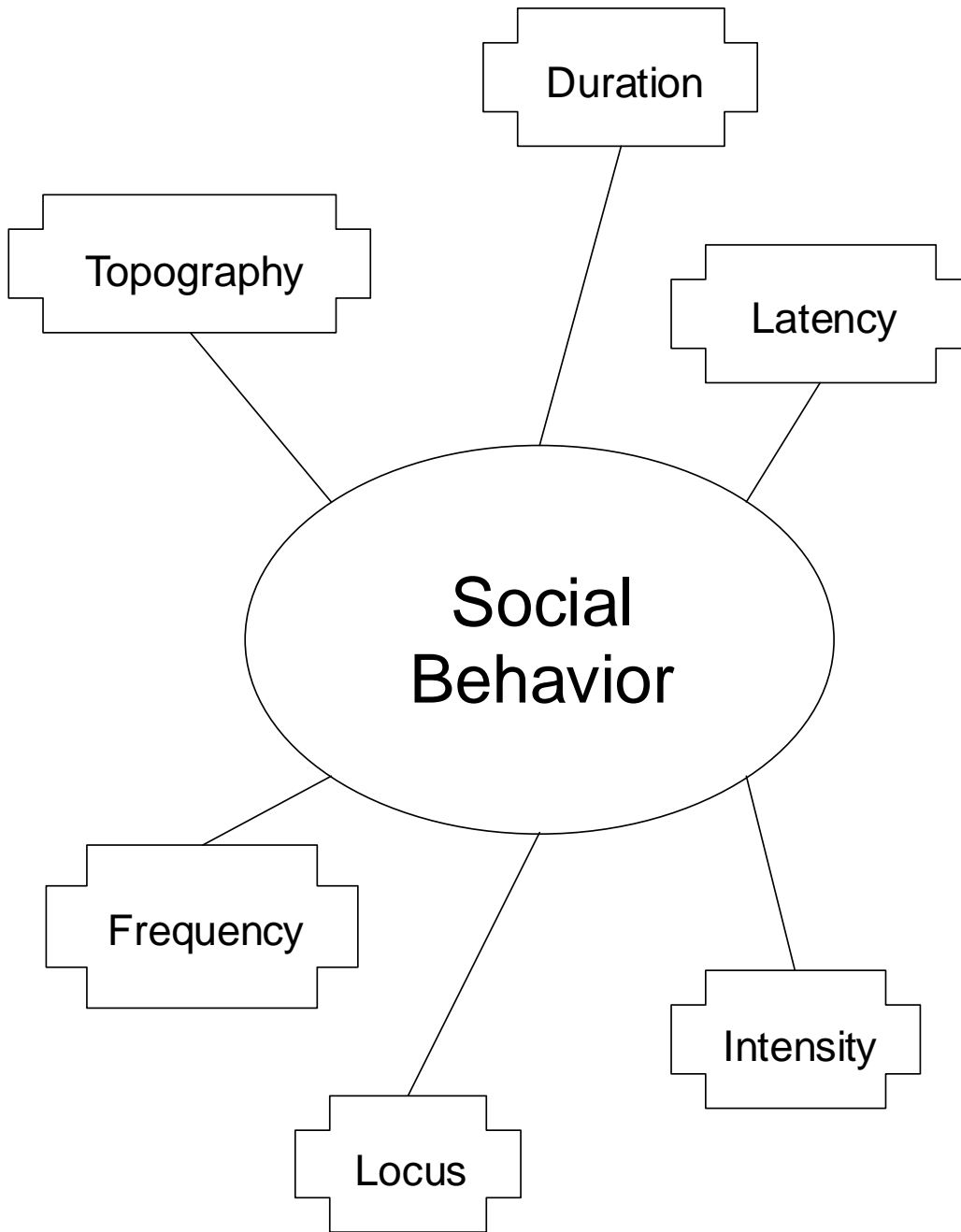


Time



Social behavior

T-12.3.2



T-12.3.3

