

***Chapter 5:
Evaluating and
Measuring
Learning***

Chapter 5 Objectives

After completing chapter 5, students should be able to do the following:

1. Define evaluation and explain the purposes of student evaluation.
2. Distinguish between the concepts of evaluation and measurement.
3. Explain the purposes of evaluation.
4. Compare and contrast pretest, formative, and posttest (summative) evaluation.
5. Create a plan for the appropriate use of the three different kinds of evaluation.
6. Differentiate among the competitive, noncompetitive, and performance assessment systems.

Chapter 5 Objectives—Continued

7. Explain what is meant by the possibility of bias in the assessment process.
8. Differentiate between the concepts of reliability, validity, and usability.
9. Describe the various sources of evaluative information.
10. Identify the advantages and limitations associated with the use of the different sources of evaluative information.
11. Explain the purpose of and advantages associated with the use of rating scales, checklists, and questionnaires.
12. Define the following evaluative terms: competitive evaluation, noncompetitive evaluation, performance assessment, norm-referenced evaluation, criterion-referenced evaluation, standard scores, and percentile.

The Evaluation Process

- Evaluation:
 - Process of Making a **Judgment** about Student Performance.
- Measurement:
 - Provides the **Data** for Making the Judgment.
- Evaluation is a **Four-Step** Process—Tenbrink (1986):
 - Preparation—decide what to Collect.
 - **Gather** Pertinent Data Regarding Desired Outcomes.
 - Making Judgments—compare data against selected criteria.
 - Make Reasonable **Judgments** Concerning Students' Performance.

Purpose of Evaluation

- Determine Whether Students are Learning.
- **Evaluate** how Effective you were as a Teacher.
- Identify what has to be Retaught.
- Gauge the Appropriateness of the Selected Curriculum for Students.
- Ensure Proper Placement of Students within a Program of Instruction.
- Make Sure State Guidelines for Achievement have been Met.

Evaluation Types

1. Pretest Evaluation:

- Administered Before Instruction.
- Assess Students' Prior Knowledge of a Topic for Placement Purposes.
- **Anticipate** Potential Learning Problems.
- Place Students in the Proper Course/Unit of Study.
- Assists Teachers in **Planning**—they can Focus on Identified Students' Needs.
- Identify the Pace of Instruction for Lessons.
- Evaluate the Appropriateness of Curriculum.

Evaluation Types—Continued

2. Formative Evaluation:

- Utilized During Instruction.
- Provides Feedback on Students' Progress and Learning.
- Monitors Instruction and Promotes Learning.
- Continuous Process.
- Helps Teachers Identify Students' Needs and Remediate them—Diagnostic in Nature.
- Enables Teachers to Correct Misconceptions about Students.
- Revise Curriculum.

Evaluation Types—Continued

3. Summative/Posttest Evaluation:

- **End of Instruction to Determine Student Achievement to Assign Grades.**
- **Judges Teacher Success and the Effectiveness of Curriculum.**
- **Focuses on the Cognitive Domain.**
- **Can Serve as Diagnostic Information for Teachers.**
- Sources of Summative Evaluation are as follows:
 - Homework.
 - End-of-Chapter Tests.
 - Classroom Projects.
 - Standardized Achievement Tests.
 - Lack of Response to Questioning.
 - Spot-Check Students' Work.

Characteristics of Pretest, Formative, and Posttest Evaluation

TABLE 5.1 Characteristics of Pretest, Formative, and Posttest Evaluation

	Pretest	Formative	Posttest (Summative)
Purpose	To identify difficulties and place students	To promote learning through feedback	To assess overall achievement
Nature	Many questions related to general knowledge	Few questions related to specifics of instruction	Many questions related to specific and general knowledge
Frequency of Administration	Varied—usually before instruction	Frequently—usually during instruction	Once—usually final phase of instruction

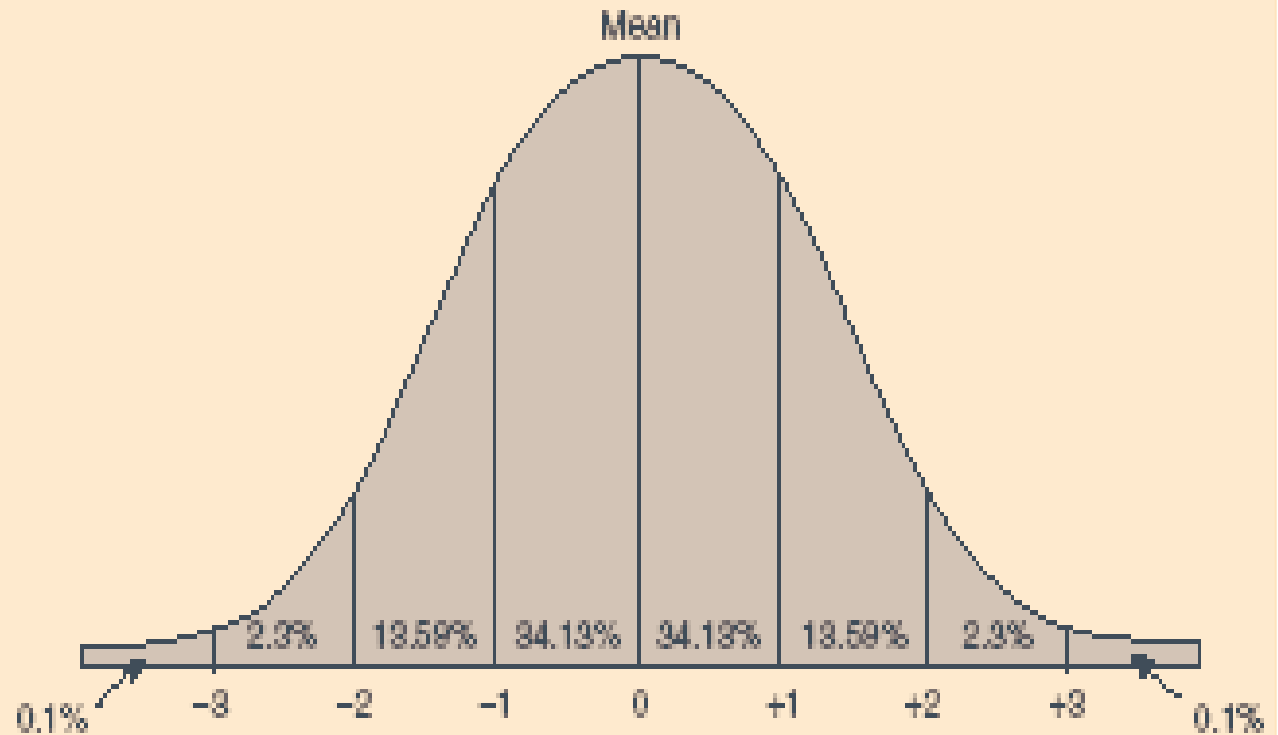
Systems of Evaluation

1. Competitive Evaluation:

- **Compare** student performance to the performance of others (Norming Group) who are similar on the curve.
- This Comparison is **Competitive**—it ranks and sorts.
- **Normal Curve/Natural Curve:**
 - Natural Distribution of all Sorts of things in Nature.
 - Standard Deviation is a Measure of how much Scores Spread out Around the Mean.
 - It is a Mathematical Construct Divided into Equal Segments.
 - Mean is the Average of Whole Population on some Attribute.
 - It is Utilized by Schools to Report the Results of Standardized Tests.

Normal Probability Curve

Figure 5.1
Normal Probability
Curve

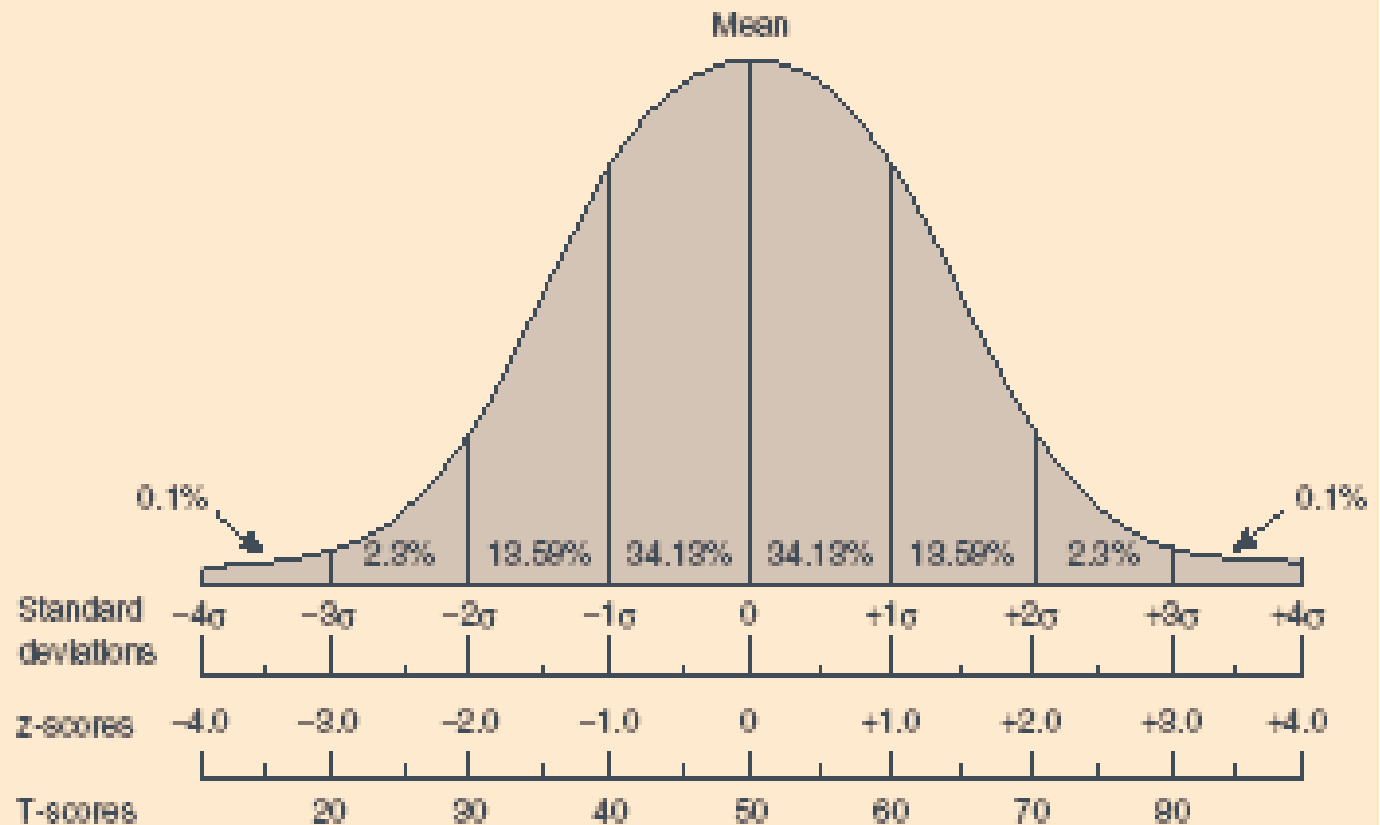


Systems of Evaluation—Continued

- **Competitive Evaluation Continued:**
 - **Standard Score:**
 - Provide a Standard Scale by which Raw Scores can be Converted.
 - Allowing the Comparison of Different Scores on Different Assessment Instruments.
 - **Four Standard Scores** are as follows:
 - **Z Scores**—correspond to the Standard Deviation.
 - **T Scores**—correspond to the Standard Deviation.
 - **Stanine Scores**—Standard Nine—groups scores into nine categories.
 - **Percentile**—percentage of population whose scores fall at/below that score.

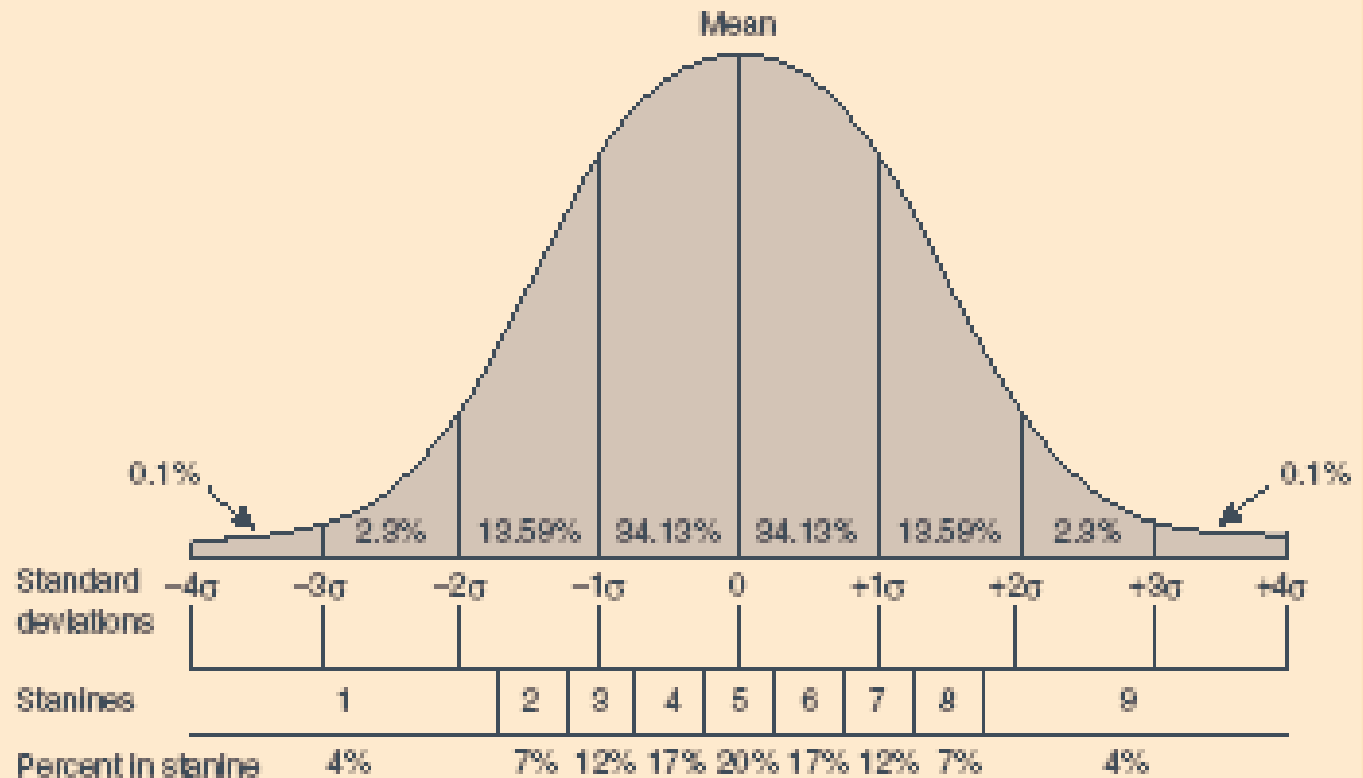
The Characteristics of the Normal Curve

Figure 5.2
The Characteristics
of the Normal Curve



Normal Distribution Curve and Stanine Scores

Figure 5.3
Normal Distribution Curve
and Stanine Scores



Systems of Evaluation—Continued

2. Noncompetitive Evaluation:

- Don't Require **Interstudent** Comparisons.
- Students are Assessed on whether they have Mastered a Specific Body of Knowledge and Skills—the Criterion.
- **Authentic** Assessment:
 - Real-World Situations.
 - Require Students to Apply **Relevant Skills** and **Knowledge**.
- **Performance** Assessment:
 - Students Demonstrate Behaviors that the Assessor wants to Measure.

Systems of Evaluation—Continued

- **Noncompetitive Evaluation Continued:**
 - **Portfolios:**
 - Systematic, Organized Collection of **Evidence**.
 - Documents **Growth** and **Development** Progress Toward Reaching Specified Goals and Objectives .
 - Enables Students to Display their Skills and Accomplishments to Others.
 - Teachers use Scoring Rubrics to Examine Portfolios.
 - Helps Students Communicate Learning to their Caregivers.
 - **Electronic Portfolios** use Technology to Facilitate the Process of Maintaining a Portfolio.
 - Teacher Portfolios Enable Teachers to Share their Duties, Expertise, and Growth in Teaching.

Problems with Evaluation

- Certain Groups are **Not** Assessed Fairly because of Predetermined Notions:
 - Girls fare **Better** than Boys—School Structure Rewards Passive Socialization Girls Exhibit.
 - Boys do Better in Middle School and High School Mathematics and Sciences because they Receive Specific Feedback in their Schoolwork.
 - Girls Receive Non-Specific Feedback in their Schoolwork and this Affects the Grades in Mathematics and Sciences.
 - Poor and Minority Students Do Poorly because of the Influence that Cultural and Economic Factors might have on some Teachers.

Problems with Evaluation—Continued

- Creative Children Might do Poorly because they Don't Test well.
- Children with Behavior Problems and Children w/Exceptionalities also Don't Fare well because of Predetermined Notions.
- Explicitly Stated Performance Expectations Prior to Evaluation Procedures Might Offset Predetermined Notions.
- Purpose of the Evaluation should be made Clear to Students before Instruction Begins.
- Assessment Results are Not to be Shared Indiscriminately.

Measurement Accuracy

- **Reliability:**
 - Consistency a Measurement Device Gives the Same Results when the Measurement is Repeated.
- Factors that **Affect** Reliability:
 - Poor Construction of the Measurement Instrument—*sampling error*.
 - Attribute being Measured May Vary over Time—*trait instability*.
 - Scoring/Recording Inaccuracies will Affect Scores—*scoring error*.
 - Motivation, Health, and Luck can Cause Variance.

Measurement Accuracy—Continued

- Techniques to **Increase** Reliability are as follows:
 - Increase the Number of Evaluative Items.
 - Establish Optimum Item Difficulty—utilize test items of Moderate Difficulty (give Moderate Spread of Scores that allow judgments of Student’s Performance in Relation to Others).
 - Write Clear Items and Directions.
 - Administer the Evaluative Instrument Carefully.
 - Score Objectively.

Measurement Accuracy—Continued

- **Validity:**
 - Whether the Evaluative Instrument Measures what it is Suppose to Measure.
 - Content Validity Requires that a Teacher's Test Matches the Actual Class Instruction.
 - Teachers Should Make Sure that their Test Items Match their Stated Learning Objectives.

Measurement Accuracy—Continued

- **Usability:**
 - How well a Measurement Device is Suited for gathering Desired Information.
 - Test should be:
 - Easy to Administer and Score.
 - Fall within Budget Limitations.
 - Suitable to the Test Conditions.
 - Have Appropriate Degree of Difficulty.
- Measurement Devices Used for **Making Decisions** must be: Reliable, Valid and Suitable.

Assessment Concepts

TABLE 5.2 Assessment Concepts

Concept	Description
Diagnostic Evaluation	Evaluation administered prior to instruction for placement purposes
Formative Evaluation	The use of evaluation in supplying feedback during the course of a program
Summative Evaluation	A judgment made at the end of a project that determines whether it has been successful or not, and commonly used to give grades
Competitive Evaluation	Evaluation that forces students to compete with each other
Noncompetitive Evaluation	Evaluation that does not force students to compete with each other
Performance Assessment	Assessment in which students demonstrate the behaviors to be measured
Portfolio	A systematic, organized collection of evidence that documents growth and development and that represents progress made toward reaching specified goals and objectives
Standard Scores	A score based on the number of standard deviations an individual is from the mean
Percentile	The point on a distribution of scores below which a given percentage of individuals fall
Reliability	The extent to which individual differences are measured consistently, or the coefficient of stability of scores
Validity	The extent to which measurement corresponds with criteria—that is, the ability of a device to measure what it is supposed to measure
Usability	The suitability of a measurement device for collecting desired data

Information Sources

- **Cumulative Records:**

- **Student Information** Collected over their Years in School.
- Contain: Academic Records, Health Records Family Data, Vital Statistics, test scores and Behavioral Comments from Past Teachers and Administrators.
- Teachers should Exercise Great Care when Reviewing these Records—Test Scores may not accurately portray the real student.
- **Federal Legislation** Protects their Confidentiality and Provides Caregivers with the Right to Examine and Challenge their Contents.

Personal Contact

- Teachers and **Daily Observations** of Students:
 - Observations Must have a **Clear Purpose**.
 - Interaction w/ Students Provides Relevant Information about Student:
 - Work Habits.
- Social Adjustment:
 - Personality Traits.
 - Strengths and Weaknesses.
 - Academic Progress.
 - Interests.
 - Beliefs and Opinions.

Facilitates Planning and Instructional
- Rating Scales, Checklists and Anecdotal Records Facilitate the Accurate Recording of this Information.

Analysis

- Teacher **Analysis** of Work and Observations:
 - Formal and Takes Place During/After Instruction.
 - Less Pressure than a Formal Test.
 - Thorough Analysis Helps Correct Any **Misconceptions** of Students' Abilities.
 - Provide Insights into Students' Thinking Processes and Skill Acquisition.
 - Students Can Also **Reflect** on their Work and Assist in Critiquing the Work of their Peers.
 - Student Reflection and Evaluation is More Productive when the Criteria is Made Known and Understood Before Instruction Begins.
 - Samples of Student should be kept in a File and Students should be Allowed to Review it and Reflect and Improve on their Work.

Open-Ended Themes and Diaries

- Social Climate and Out-School **Differences** in Students' Lives Can Add to the Understanding Teachers have of Students.
 - Curriculum can be Revised.
 - Daily Classroom Work can be Adjusted.
- **Sources** for this Information can come from:
 - Writing Activities that Address these Issues.
 - Periodic-Schedule Diary Writing Concerning Concerns/Feelings.
 - Rapport between Teacher and Students must be Developed for this to take Place.
 - Teachers should Read w/Care these Reflections and Respect the Privacy they Warrant.

Conferences

- **Influences** on Students are: Caregivers, Neighborhoods, Communities.
- **Involving** Caregivers:
 - **Scheduled** and **Planned** Conferences:
 - **Clarify Problems** Children might be having in Class.
 - Atmosphere should be **Positive** w/Caregivers as Equal Partners.
 - Clears up Caregiver's **Misconceptions**.
 - **Student-Led Conferences** are Enable the Caregivers and Teacher to get Know each other better.
- **Other Sources of Information** for Teachers are:
 - Past Teachers and Administrators.
 - School's Support Personnel.

Framework for a Successful Conference

TABLE 5.3 Framework for a Successful Conference

Step 1	<i>Plan ahead.</i> Establish your purpose. Plan what you intend to say, what information you want to obtain, what your concerns are. Plan what your next step will be in the classroom as a result of the conference.
Step 2	<i>Starting the conference.</i> Be positive. Begin the conference with a positive statement.
Step 3	<i>Holding the conference.</i> Establish a positive sharing relationship. Be an active listener. Be accepting with regard to input and advice. Establish a partnership, so all concerned can work toward a common goal.
Step 4	<i>Ending the conference.</i> End the conference on a positive note. Communicate the fact that in working together the common goals will be reached.
Step 5	<i>Conduct follow-up contact.</i> Keep all parties informed. Send notes and make phone calls to share successes and/or further concerns.

Testing

- **Tests:**
 - Series of **Tasks** to Obtain Systematic Information.
 - Presumed to be Representative of an Educational Attribute/Series of Attributes.
- **Tests** are Classified as follows:
 - Administration—group/Individually.
 - Scoring Procedures—objectively/subjectively.
 - Response Emphasis—speed/accuracy.
 - Response Type—performance/paper-and-pencil.
 - Comparison Groups—teacher-made/standardized.

Testing—Continued

- **Norm-Referenced** Test Interpretation:
 - Compare a Student's Score w/a **Norming Group**—a representative sample.
 - **Compares** Individuals w/one Another.
 - Utilized when Schools need to Know how their Students Performed in Relation to others at the same Age/Grade Level.
- **Criterion-Referenced** Test Interpretation:
 - Compare an Individual's Score against a **Predetermined** Standard/Absolute Standard—such as 60 per cent/80 per cent Correct.
 - Teacher made Tests are Criterion-Referenced Tests—students are compared to a criterion established by the teacher.

Testing—Continued

- **Classroom Tests** are:
 - Pretests and Posttests.
 - Chapter Exams.
 - Unit Exams.
 - Midterms and Final Exams.
 - Quizzes.
 - Standardized High Stakes Tests.
- **Criticisms of Tests** are:
 - Tests Superficial Learning.
 - Emphasis on Knowing than Thinking.
 - Verbalization is Rewarded Not Doing.
 - Based on Teacher Expectations Not on Student's Beliefs and Values.

Techniques to Increase Reliability and Validity During Student Observations

1. Utilizing Rating Scales—specific characteristics or qualities that are arranged in order of quality.
2. Checklists—list of criteria against which a student’s performance or end product is judged.
3. Questionnaires—device that helps teachers measure students’ attitudes, feelings, and opinions.

Final Words on Evaluating and Measuring Learning

- **Remember** that to Obtain a Complete Picture of Students, Teachers Need a **Variety of Sources** of Information.
- **Complete** Pictures Affect:
 - **Quality** of Teaching and Planning
 - **Curriculum** that is Used

Rating Scale for Group Involvement

Figure 5.6
Rating Scale for
Group Involvement

RATING SCALE FOR GROUP INVOLVEMENT

Name _____ Date _____

Directions: Rate the student's involvement on each attribute by placing an X on the appropriate position on the scale.

1. Participation _____
0 1 2 3 4 5
Uninvolved Involved

2. Cooperativeness _____
0 1 2 3 4 5
Inadequate Adequate

3. Productivity _____
0 1 2 3 4 5
Inadequate Adequate

4. Self-Discipline _____
0 1 2 3 4 5
Inadequate Adequate

5. Motivation _____
0 1 2 3 4 5
Inadequate Adequate

Information Sources and Evaluation Instruments

TABLE 5.4 Information Sources and Evaluation Instruments

Concept	Description
Cumulative Record	A file that holds information collected on students during the school years
Personal Contact	The collection of data through daily interactions and observations of students
Analysis	An examination of students' work
Open-Ended Themes and Diaries	The periodic writing of students on in-school and out-of-school topics
Conferences	A meeting between individuals regarding issues of common concern to both parties
Testing	A task or series of tasks used in obtaining systematic observations
Rating Scale	A set of characteristics arranged in order of quality so that performance can be better judged
Checklist	A listing of characteristics used for indicating the presence or absence of identified attributes
Questionnaire	A set of statements to which students react in order to examine their attitudes, feelings, and opinions

Reflection



This Picture is taken from page 180 of this Chapter. The student appears to be visibly upset with the grade on this assignment.



Based on your reading of this Chapter, **as his teacher** what might you have done differently to **avoid** his present reaction?

The End!