

Assessment Case Studies

Use these bookmarks to navigate to your case study:

- [Elementary Math](#)
- [Elementary Language Arts](#)
- [Middle School Social Studies](#)
- [Middle School Physical Education](#)
- [High School Music \(Instrumental\)](#)
- [High School Math](#)
- [High School Choir](#)
- [High School Health](#)
- [High School Science \(Biology\)](#)
- [High School Social Studies \(U.S. History\)](#)
- [High School English](#)

Elementary Math (Grade 3 – Multiplication) ([homepage](#))

Scenario:

Ms. Hernandez teaches 3rd-grade math. The unit objective states that students will *demonstrate conceptual understanding of multiplication as repeated addition and equal groups*.

For the unit assessment, students complete a **30-question worksheet in five minutes**. All problems are **word problems** (e.g., “There are 6 bags with 4 apples in each bag. How many apples are there?”). Students are required to show their work using both **drawings and written explanations**.

Ms. Hernandez believes this approach provides a more “authentic” view of student thinking. However, several students receive reading intervention services, and about one-third of the class consists of English language learners. During the assessment, some students appear rushed, skip problems, or struggle to finish writing explanations even when they can correctly solve problems orally.

Ms. Hernandez notices that students who read fluently tend to score higher, even though similar students demonstrate comparable multiplication understanding during small-group instruction.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

Elementary Language Arts (Grade 4 – Paragraph Writing)

[\(homepage\)](#)

Scenario:

Mr. Thompson's learning goal is for students to *write a coherent paragraph that includes a topic sentence, supporting details, and a concluding sentence*.

Students are given 40 minutes in class to write a paragraph about their favorite place. Mr. Thompson grades the paragraphs holistically, using his professional judgment rather than a rubric. He explains to students that he “knows good writing when he sees it.”

When grading, he realizes that his mood affects scoring; after reading several strong papers in a row, average papers seem weaker. He also notices that students with neat handwriting and stronger spelling receive higher scores, even when their paragraph structure is unclear.

When parents inquire about grades, Mr. Thompson struggles to explain why two similar paragraphs received different scores.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

Middle School Social Studies (Grade 7 – American Revolution) ([homepage](#))

Scenario:

Ms. Walker wants students to *explain multiple causes of the American Revolution and how they interacted*.

Due to time constraints, she selects a **20-question multiple-choice quiz** from the textbook’s test bank. Most questions assess students’ recall of dates, events, and definitions (e.g., “What year was the Stamp Act passed?”).

After grading, Ms. Walker notices that students who memorized vocabulary performed well, but several students who participated actively in discussions about causes and consequences scored poorly. When she meets with students, some can verbally explain the causes of the Revolution but struggle with the quiz questions.

Ms. Walker begins to wonder whether the quiz reflects what students actually understand—or simply what they memorized.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

Middle School Physical Education (Grade 6 – Overhand Throwing) ([homepage](#))

Scenario:

Ms. Lopez's learning goal is for students to *demonstrate correct overhand throwing technique*.

During a busy class period, each student gets **one attempt** to throw a ball while Ms. Lopez oversees multiple stations. Students throw from varying distances due to space limitations. Ms. Lopez assigns scores based on quick observations while managing behavior and equipment.

Some students are nervous and rush their throw, while others wait longer and feel more comfortable. Ms. Lopez later wonders whether a single attempt under these conditions provides an accurate picture of student skill.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Music (Instrumental) ([homepage](#))

Scenario:

Mr. Davis teaches a high school band. His final assessment requires students to perform a solo piece. The stated objective is to assess “**Musical Expression and Technical Proficiency.**”

Mr. Davis uses a rubric with two categories, each scored from 1–10:

- **Technical Accuracy** (correct notes, rhythm, tempo)
- **Musicality** (emotion, interpretation, phrasing)

Over time, Mr. Davis notices that he tends to assign higher “Musicality” scores to students who select musical pieces he personally enjoys, even when phrasing or expression is inconsistent. Additionally, students who perform early in the day—after warming up—often receive higher technical scores than students who play later, especially right before lunch when fatigue sets in.

Performances occur over three class periods, and Mr. Davis does not record performances for later review.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Social Studies (Grade 11 – Document Analysis)

[\(homepage\)](#)

Scenario:

Mrs. Reed’s objective is for students to *analyze historical documents and evaluate competing historical perspectives*.

Students are assigned a **five-page take-home essay** analyzing primary source documents. No drafts or checkpoints are required, and students receive the essay prompt three weeks before the due date. The rubric includes broad categories such as “depth of analysis” and “quality of argument.”

Mrs. Reed notices wide variation in essay quality. Some papers include advanced arguments and polished writing, while others reflect minimal analysis. Several students later admit they received help from parents or tutors, while others worked independently.

When grading, Mrs. Reed wonders whether the assessment measures students’ analytical skills—or access to outside support.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Math (Algebra II – Quadratic Equations)

([homepage](#))

Scenario:

Mr. Smith aims to assess students' ability to *solve quadratic equations using multiple methods*.

The unit test includes:

- 5 multiple-choice questions
- 2 extended-response problems requiring multiple steps

Mr. Smith gives partial credit but does not use a clear marking scheme. Students who make small arithmetic errors sometimes lose most of the points, depending on how rushed Mr. Smith feels while grading.

When students review their tests, they notice inconsistencies in scoring similar mistakes. Mr. Smith finds himself second-guessing his grading decisions and worries about fairness.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Choir ([homepage](#))

Scenario: Ms. Patel wants students to *demonstrate accurate pitch and blend within a choral ensemble*.

Grades are largely based on Ms. Patel's impressions during rehearsals. Students who stand near the center of the choir tend to receive higher grades because they are easier to hear. Quieter students who blend well but sing softly sometimes receive lower scores.

Students are unclear how their individual performance is assessed within the group context. Ms. Patel values musicality but struggles to assess individuals fairly in a large ensemble.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Health ([homepage](#))

Scenario:

Ms. Turner teaches 10th-grade Health. Near the end of the semester, students complete a written assessment intended to measure their understanding of **decision-making skills related to nutrition, substance use, and mental health**.

The assessment consists of:

- 40 multiple-choice questions taken from the textbook publisher's test bank
- 2 short-answer questions asking students to describe a "healthy choice they made recently"

Ms. Turner administers the test during first period on a Monday. Several students arrive late, and some miss the test due to an assembly later that same day. Students who miss the test are allowed to take it after school with no time limit.

For the short-answer questions, Ms. Turner grades holistically and admits that she tends to give higher scores to students whose choices align with her personal beliefs about health (e.g., students who discuss diet and exercise over video game use or screen time).

She notices that students who perform well on the multiple-choice section often struggle to describe real-life applications of health concepts, while some students who actively participate in class discussions perform poorly on the test.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Science (Biology) ([homepage](#))

Scenario:

Mrs. Jones teaches 11th-grade Biology. To assess student understanding of **cellular respiration**, she gives a unit test that includes vocabulary matching, diagram labeling, and one extended-response question asking students to “explain how ATP is produced.”

The extended-response question is worth 40% of the test grade. Mrs. Jones grades the responses without a rubric, relying instead on his “gut feeling” of whether a student truly understands the concept.

Because grading extended responses is time-consuming, she grades some responses late at night after practice and others during her prep period. She later realizes he may have been stricter with students whose handwriting was difficult to read or whose responses were longer but less organized.

Additionally, English learners and students with IEPs receive extended time, but the test language itself is dense and highly technical. Several students who understand the process orally struggle to express their knowledge in writing.

Mrs. Jones is disappointed that test scores are lower than expected, especially among students who performed well during labs.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School Social Studies (U.S. History) ([homepage](#))

Scenario:

Ms. Alvarez teaches U.S. History. Her summative assessment for a unit on the **Civil Rights Movement** is a timed in-class essay responding to the prompt:

“Describe the most important leader of the Civil Rights Movement and explain why their contributions mattered most.”

Although the unit covered multiple perspectives, Ms. Alvarez strongly emphasizes Martin Luther King Jr. in her lectures. Students who focus on lesser-known figures (e.g., Ella Baker, Malcolm X, or local activists) often receive lower scores, even if their arguments are well supported.

Ms. Alvarez uses a general essay rubric but does not share it with students beforehand. She notices that essays written during first-period classes tend to score higher than essays written during last-period classes, when students are tired and rushing to finish.

She also grades essays alphabetically, sometimes reading several weaker essays in a row, which she suspects impacts her expectations for the next paper.

Ms. Alvarez wonders whether the assessment truly reflects students’ understanding of the movement—or simply their essay-writing stamina and ability to guess what she wants.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-

High School English ([homepage](#))

Scenario:

Mr. Lawson teaches 10th-grade English. For a literature unit on *To Kill a Mockingbird*, he assesses student learning through a creative project. Students may choose one of the following formats:

- Write a traditional literary analysis essay
- Create a visual art piece representing a major theme
- Perform a dramatic monologue from a character's perspective

The stated goal of the assessment is to measure **literary analysis and thematic understanding**.

Mr. Lawson uses one rubric for all formats. However, the criteria are broad and ambiguous, with categories such as:

- “Depth of Understanding”
- “Creativity”
- “Quality”

Students who are artistically inclined tend to receive higher scores, even though their projects sometimes lack textual evidence. Meanwhile, strong writers occasionally receive lower scores if their essays feel “uninspired.”

Mr. Lawson also realizes that he enjoys grading creative projects more than essays and may be more generous with projects he personally finds interesting.

After grades are posted, several students ask how the same learning goal could be fairly evaluated across such different formats.

1. To what extent is the assessment **reliable**? Please justify your answer.
 2. To what extent is the assessment **valid**? Please justify your answer.
 3. What is one concrete, practical step the teacher can take to improve the assessment? Please reference validity and/or reliability.
-