Key Test Building Strategies to Measure Learning

Study Island

Create a test blueprint

- Choose a learning timeframe: full academic year, between units, between lessons.
- Decide on essential understandings within that timeframe.
- Decide on topics related to those essential understandings.

Determine appropriate test length

- Depending on how many topics are identified in your blueprint, each topic may only be measured by 1-8 items. (e.g., if there are 10 important topics to cover with 60 items, each topic may have around 6 items).
- The 60 for 60 Rule: If you have 60 minutes of testing time, you should not include more than 60 multiple-choice items.
- Include a few extra items in the pre-test to increase options on the post-test.

Note: This is a bit of a goldilocks dilemma—the more items you have, the more accurate and reliable your assessment will be (all things equal). However, keep in mind that eventually test takers get tired of testing and their performance decreases. Ideally, the test should be short enough to avoid test-taking fatigue.

Leverage Study Island Items

Use Study Island Test Builder to find items for students to demonstrate skills in the topic.

- Use the filter tool in Test Builder to quickly find items that are necessary to assess the topics related to your essential understandings.
    - Filters: State, Subject, Grade, Topic, Item Type, Item Depth of Knowledge (DOK)
- Spread item difficulty: Pick a range from easy to difficult to get a good measure of low to high achieving students.
    - What makes an item difficult?
        - Technology enhanced items are often more difficult than multiple-choice items.
        - Items with more complex scenarios are often more difficult.
        - Items with attractive distractors (that is, incorrect answers that seem plausible) will be more difficult than items with obviously incorrect answers.

Study Island

Evaluate PRE-TEST performance

Evaluate how examinees and test items performed to determine how to adjust the pre-test, or if you need to build a new one.

- Adjust: Drop items where more than 80% of students got it right
    - Will be difficult to see growth if you don't
    - If more than 20% of items—or whole topics—need to be dropped, select and administer more pre-test items.
- Re-build: Floor and Ceiling Effect
    - Responding to the floor effect: If most students got most questions wrong, create a new pre-test with easier items.
    - Responding to the ceiling effect: If most students got most questions right, create a new pre-test with harder items.

Note: All students should not receive exactly the same score – some fluctuation in student scores is expected.

Evaluate POST-TEST performance

- Build your post-test using items retained from the pre-test.
- Ideally test difficulty will have decreased (scores increased).
- Variability is good: If most students get 100% correct on the post-test, the items are not difficult enough to accurately gauge how much growth students achieved.
- If you observe a pronounced ceiling effect, be aware that growth is likely to be underestimated.
- If an item must be dropped from the post-test, also drop it from the pre-test.

Calculate Growth

- Once the final set of pre-test and post-test items are identified and administered, all students should have their pre-test scores and post-test scores computed.
- The post-test and pre-test scores (number or percent correct values) can then be tested for significant differences, and/or compared using effect size statistics[1].

[1] For specifics on significance testing and effect size calculation, see: https://datatab.net/tutorial/paired-t-test

edmentum
Edmentum.com
800.447.5286
info@edmentum.com
5.2.2023

5600 W 83rd Street
Suite 300, 8200 Tower
Bloomington, MN 55437