

Improving Literacy Brief

Understanding Screening

CLASSIFICATION ACCURACY

Procedures aimed at classifying and predicting outcomes are important in a variety of settings. A goal in classification accuracy is to correctly identify issues that result in a later problem and situations in which the scores identify issues that do not result in a later problem.

The Transportation Security Administration (TSA) at the airport may offer a useful illustration of classification accuracy. The security line process is aimed at detecting early in the travel process items that are not allowed on flights. The scanners are set at thresholds which cause the buzzer to indicate a certain amount of “unallowable” material has been detected. When a traveler sets off the buzzer, one possibility is an item that is genuinely “not allowed” (e.g., a set of nail scissors) has been transported through the scanner. In this case, the scanner has done its job and steps are taken to resolve the issue. However, there are also instances in which the scanner “detects” something that may not be there or may not be problematic. Perhaps a harmless item which is “allowed” (e.g., some forgotten pocket change, replaced hip) sets off the detector because it shares some properties with “not allowed” items. These two scenarios illustrate a true positive (nail scissors) and a false positive (coins, hip).

On the other hand, consider a scenario in which the buzzer did not indicate any “not allowed” items. In most cases, the buzzer does not go off because there is a genuine lack of prohibited items. However, it is also possible that some prohibited items were there, and the scanner was not set at a threshold, or sensitive enough, to prompt detection. These two scenarios illustrate a true negative (nothing there to detect) and a false negative (something was there, but not detected). This example illustrates a balance that TSA attempts to strike: setting a threshold on the scanner appropriately to reliably detect items when they are actually present.

A similar accuracy is important with instruments used to measure academic progress. Classifying students is a key step in universal screening, an assessment process that helps educators identify students who are at risk for not meeting grade-level learning goals. The aim is to have tools which permit accurate classification and identification. It is very important to:

1. accurately classify a student as being at risk when they actually are at risk, or alternatively
2. accurately classify a student as not at risk when they are genuinely not at risk for academic difficulties. These are academic screening instances of a true positive and true negative.

CLASSIFICATION ACCURACY

Alternatively, it is possible that some students who are not-at-risk are classified as at-risk, and some students who are at-risk are classified as not-at-risk. These scenarios represent academic versions of a false positive and false negative, respectively. The latter example may be particularly problematic in a case when a student may miss out on critical additional support that they need.

As noted previously, a primary goal in classification accuracy is to correctly identify issues that result in a later problem versus situations in which the scores identify issues that do not result in a later problem. This goal is important whether one is considering a scanner in a TSA line or an academic screening tool.

In educational contexts, the classification procedure typically begins with an assessment of academic skills. Students' performance is reflected in scores which are then interpreted by academic professionals (e.g., teachers, administrators, school psychologists) and parents. The scores can be viewed in terms of raw scores (i.e., overall points earned) or percentile ranks (i.e., where one student score may rank in relation to their peers). They may also be used to classify a student in terms of risk for an academic problem. Typically, terms such as *not-at-risk* or *at-risk* are applied when a student scores within a range above or below a certain score on any given test. There may also be classifications in between these two, along the lines of a *marginal-risk* classification. It is important to classify students correctly, as subsequent educational plans or programming may (or may not) be made based upon these determinations of risk.

Turning again on the TSA scanner example, these risk classifications are apparent when the buzzer goes off. Ideally, there is an accurate scanner which does its job with relatively high rates of true positives (i.e., only buzzes when a prohibited item has been transported through the machinery) and true negatives (i.e., does not buzz because there was nothing there to detect). On the other hand, ideally there are relatively low rates of false positives and false negatives (i.e., the scanner does not miss anything important due to lack of sensitivity). Over the years, there has been emerging technology which has yielded greater accuracy with TSA scanners.

In the same vein, accuracy is sought with academic screening tools used for academic risk classification purposes. More specifically, sensitivity and specificity rates help gauge tests which are able to achieve true classifications at a high rate. Sensitivity is a probability that reflects the percentage of observations indicating a problem was correctly detected by the screener as being a problem. Specificity is a probability that reflects the percentage of observations indicating no problem was correctly detected by the screener as not having a problem. The National Center on Intensive Interventions (NCII) tools chart ([link below](#)) rates a screening tool highest when it has a sensitivity rate of 70% or higher and a specificity rate of at least 80%.

CLASSIFICATION ACCURACY

The sensitivity and specificity rates are useful when trying to determine which screening tools can distinguish, with relative accuracy, among at risk and not at-risk students.

SUGGESTED CITATION

Stanley, C., Petscher, Y., & Pentimonti, J. (2019). *Classification accuracy*. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy. Retrieved from improvingliteracy.org

WHERE DO I GO FROM HERE?

- For more information about identifying high quality screening tools: <https://intensiveintervention.org/>
- For more information about screening processes: <https://improvingliteracy.org/whitepaper/screening-dyslexia>

CLASSIFICATION ACCURACY

WHAT IS CLASSIFICATION ACCURACY?

Classification accuracy is an important aspect of high-quality instruments used to measure academic progress. For reading, it tells us how good a tool is at correctly distinguishing students that have a reading difficulty from students who do not.

Classification accuracy is like a Transportation Security Administration (TSA) airport line.

TRUE POSITIVE

When a traveler sets off the buzzer, one possibility is an item that is “not allowed” has come through the scanner. In this case, the scanner has done its job. This is a true positive detection.

FALSE POSITIVE

There are also instances in which the scanner “detects” something that isn’t actually there or something that is there but is not actually problematic. This is a false positive detection.

TRUE NEGATIVE

Perhaps the buzzer did not indicate any “not allowed” items. If it’s correct and there is nothing there at all, this is a true negative (nothing there to detect).

FALSE NEGATIVE

Perhaps the buzzer did not indicate any “not allowed” items, but instead it’s incorrect. In fact, there are some prohibited items and the scanner did not prompt detection. This is a false negative (something was there, but not detected).



Accurately classifying students at risk for not meeting grade level goals is a key step in universal academic screening. Tools that maximize the rates of true positives and negatives, or accurate classification and identification, are best.