

Evaluation funnel

The following evaluation funnel was designed to demonstrate how different assessment techniques are used to make decisions. People make observations (assessments) all of the time. Based on those observations and how we interpret the information, various decisions are made (evaluation). The repercussions for some of those decisions are relatively minor, however as the 'stakes' increase we have to use observations that are increasingly accurate and objective.

It should be noted that the funnel's intent is to communicate issues concerning decisions concerning assessment and evaluation. Qualities of the various techniques listed are not exhaustive, and different audiences would rate the stakes of various "questions" differently. Furthermore, there is no universally accepted 'gold standard' for assessment qualities.

Stakes/Qualities

Questions

Assessment Examples

Low to Medium

Feasible,
Easy, quick
to administer
by anyone;
frequently
conducted

Subjective
Different observers
may have
different opinions.

The observer defines
what is 'normal'

Did Billy understand the lesson that I just taught? Is Kelsey paying attention? Is Gary asleep or thinking deeply? Is this reading passage difficult for Richard? Is Mary on the right page? Can Daniel already understand the vocabulary words in the unit that I am about to present? What do I need to teach Sarah for her to do long division more correctly? Does Nathan know the difference between a letter and a word?

Which sounds has Burton mastered and which ones still need teaching? Has Elizabeth learned the concepts in this social studies unit? Is Julia ready to go on to the next basal reader? Does Clara have the prerequisite skills for the next math unit?

Watching the student

Asking questions

Teacher made tests and quizzes

Worksheets

Informal testing and observation

Test, teach, test

Medium to High

Increased structure As the stakes increase the format of the test becomes increasingly structured or standardized.

Increasingly objective Different raters render similar judgments.

Measurement qualities moderately robust. Often have research support

May begin to depart from direct overlap with what is taught

How should I describe Clara's skills in reading to her parents on her report card? Are the procedures that I'm using to teach science effective for my class in general?

How are Syracuse City Schools 3rd graders performing in reading compared to other 3rd graders? How can we improve this program?

Should Emily be evaluated for a disability? Which students should be included in this remedial program?

Work samples that are evaluated by some type of criterion

Specific, well constructed rubrics

Assessments conducted in a standardized format

Reading Comprehension Assessment (RCA)

Curriculum Based Assessment (CBA)

Group standardized tests (PEPs, IOWA, DRP)

High Stakes Assessment

Proven measurement qualities Research supporting reliability, validity and generalizability of results.

Frequently administered individuality to assure reliable results (desirable)

Objective and unbiased

May be less "feasible" Take more time to administer and interpret (in the case of commercial norm-referenced tests)

Results can be communicated clearly and accurately to a wide audience

Is this expensive program effective? Does Michael have a disability? Is Edward (an at risk student) making progress in response to the intervention?

Individually administered norm referenced assessment (i.e., Woodcock Reading, WIAT, K-TEA)

Curriculum Based Assessment (CBA)

Monitoring of Basic Skills Program (Computerized Cloze Procedure)

Researched questionnaires/checklists

Repeated, systematic student observation