EPSY 5221 Test Design

**Common words used in the measurement field:**

A Test: a task or series of tasks used to obtain systematic observations presumed to be representative of educational or psychological traits or attributes

Teacher-made test: constructed by teachers for use within their own classroom, tailored to classroom learning objectives

Standardized test: constructed by test specialists working with curriculum experts and teachers; samples of behavior are gathered under uniform procedures (fixed items/tasks, administration procedures, and scoring) – no connotation re: standards

Norm-Referenced test: designed to measure individual differences in achievement, intelligence, interests, attitudes, or personality; relative to others (within group or between group norms)

Criterion-Referenced test: relate a student’s score on an achievement test to a domain of knowledge rather than to the scores of others, or some normative distribution

Individual tests are designed to be administered to one person at a time

Group tests are designed to be administered to many persons at a time

Interim assessments are often used to measure academic progress, but some refer to these as so-called formative assessments – two purposes that are not compatible; these do not necessarily follow a specified schedule. Typically they are employed for formative uses. These are often standardized, common-course, and end-of-course exams. Others argue that these fall between so-called summative and formative assessments, with the intent to aggregate information beyond the classroom level.

“Assessments administered during instruction to evaluate students’ knowledge and skills relative to a specific set of academic goals in order to inform policymaker or educator decisions at the classroom, school, or district level. The specific interim assessment designs are driven by the purposes and intended uses, but the results of any interim assessment must be reported in a manner allowing aggregation across students, occasions, or concepts” (p. 6). See: Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*(32), 5-13.

Benchmark assessments, sometimes called interim assessments, are given periodically, beginning with the start of the school year. These are intended to provide periodic information on the achievement of content standards in the midst of instructional cycles. Some suggest that there are two forms of benchmark assessments, interim and formative. Others argue that benchmark, diagnostic, formative, and predictive assessments are all under the umbrella of interim assessments.

Objective test/scoring: has clear and unambiguous scoring criteria—independent scorers can agree on the number of points answers should receive

Subjective test/scoring: relatively vague descriptive scoring criteria where independent scorers may not agree

Power test: generous time limits so most students can attempt every item

Speeded test: sever time limits where items are so easy that few students are expected to make errors

Generic Steps in Planning Assessment Program

1. Define Purpose
2. Determine Resources Available
3. Test selection or Development [non-trivial tasks]
4. Administrative decisions
   1. Timing – Fall, Spring (depends on purpose)
   2. Test preparation activities
5. Plan for Dissemination and Use of Results

|  |  |  |
| --- | --- | --- |
|  | Teacher-Made | Standardized |
| Specificity of Objectives | Objectives are specific to the needs of students in a given classroom | Objectives are general to the needs of students in most classrooms |
| Content | May come from any area of the curriculum. Items may be added, eliminated, or modified as desired | Items are fixed and are not modifiable; only the most common areas of the curriculum are surveyed |
| Rules for administration and scoring | Determined by the teacher. Should be uniform within the class and can be adapted to the particular needs of students. | Determined by the test publisher. Must be followed exactly as stated in the test manual. |
| Norms | No norms are provided, although teachers may generate their own. | Norms are provided by the publisher to compare class or student performance to similar and different age and grade groups. |
| Evaluation of tests | Test quality is assessed by the teacher | Data on the quality of instruments are provided by the publisher. |

A Case for Review

A small community college suddenly found itself under intense pressure from local politicians and the State Education Authority to raise admissions standards. The college had always used high school grades, letters of recommendation, and a personal interview to select students for admission. Some students went on to 4-year colleges after completing the 2-year program, but many went to work in the community after graduation. A large number of students did not complete the 2-year program, and this caused the State Educational Authority to become concerned.

After a lengthy meeting of the faculty and trustees, the community college decided to change its admissions policy radically. High school grades, letters of recommendation, and a personal interview were all discontinued in favor of the Scholastic Assessment Tests, formerly the Scholastic Aptitude Test (SAT). The new policy was viewed as a way to achieve an objective standard of high quality for admission. The community college officials also thought the SAT would be helpful in their communication with the community because the single entrance requirement would be easier to explain.

After announcing the change in admissions policy, college officials were surprised to receive complaints from high school seniors and their parents as well as from professional testing experts. The college administrators were puzzled by the controversy generated by their change and wondered what steps to take next.

1. What should the community college administration have done to gain a clearer understanding of the dropout problem?
2. Should the SAT have been used as the sole criterion for admission?
3. In keeping with the mission of the community college, what kind of admissions screening procedures should be used?

How does measurement and evaluation inform decision making of this type?