General Education Assessment at JMU
Instrument Categories

Created by Dena Pastor with the assistance of
Christine DeMars, Patrick Meyer, & Steve Wise

Instruments administered for General Education program assessment at CARS can be classified into two categories based on their administration purpose. Because the instruments differ in administration purpose, they also differ in data collection scheme, types of scores reported, sample requirements, typical examinee motivation and required test security. The two categories of instruments, which we refer to as standards-based and value-added, are described in detail below.

Standards-Based Instruments

Purpose: The purpose of standards-based instruments is to determine if students are meeting JMU's expectations in various areas of proficiency. While some students may acquire the skills and knowledge necessary to achieve the expectations through their experience at JMU, it is not necessarily assumed that all students acquire such skills and knowledge strictly at JMU.

Data Collection Scheme: The data collection scheme utilized with standards-based instruments typically entails a single administration of the instruments after students have been informed of: (a) the proficiency requirement, (b) JMU's expectations for proficiency, and (c) the training available to assist students in achieving JMU's expectations.

Type of Score Reported: Students are classified into various proficiency categories (e.g., Basic, Proficient, Advanced) by comparing the student's raw or standard score on the instrument to standards of performance established by JMU faculty and administration. While individual students are informed as to which proficiency category they fall into, stakeholders are provided with the percent of students classified into each proficiency category. To assist in the interpretation of this information, stakeholders are also provided (if available) with the faculty's expectations as to the percent of students that should be in certain proficiency categories (e.g., 70% of the sophomore student population should be classified as either Proficient or Advanced).

Sample Requirements: Because competency in the proficiency areas assessed using standards-based instruments are typically required by JMU, the majority of students undertake testing. A small proportion of students may not undergo testing if they are able to have the requirement waived (e.g., students who have General Education requirements waived through an articulation agreement with VCCS).

Examinee Motivation: With our standards-based instruments, personal consequences are associated with test performance. If a student does not meet a certain level of proficiency, they are typically required to retest until proficiency is attained. Because there are personal consequences to an examinee's performance, examinee motivation to do well on the assessment tends to be high. Because examinee motivation is high, a student's test score is thought to be a fairly trustworthy indicator of what a student knows and is able to do.

Test Security: Because consequences are associated with test performance, the possibility of cheating is heightened. Steps taken to ensure test security and to minimize cheating include the use of secure computer based instruments, creation of multiple test forms, and the use of proctors. Locations where test forms or data are stored must also be kept secure. While high stakes testing yields the advantage of promoting ideal examinee motivation, costs are increased (hiring computer programmers and proctors, purchasing appropriate secure storage mechanisms). Not only are monetary costs increased, but the demands made on faculty are also high as the faculty is often asked to voluntarily create more high-quality items for additional test forms without monetary compensation.

Value-Added Instruments

Purpose: Value-added instruments are used to determine the extent to which students' skills and knowledge change as a result of completing their general education requirement at JMU. It is assumed that students will increase in their skills and knowledge as a direct result of completing the general education curriculum. It is assumed that all students, regardless of their level of proficiency as incoming freshmen, will gain in skills and knowledge as a result of completing JMU's general education curriculum.

Data Collection Scheme: To capture the extent to which proficiency changes over time, value-added instruments are typically administered at two time points: once to students as incoming freshmen (pretest; August Assessment Day) and again after students have completed 45-70 credit hours (posttest; February Assessment Day). Students are typically in various stages of completing their general education requirement at the time of posttest: some students have not fulfilled any portion of the requirement ("non-completers"), others have partially fulfilled the requirement ("partial completers"), and yet others have fulfilled the requirement ("full completers"). These various subgroups of students can be likened to different experimental groups. Non-completers serve as the control group and the partial and full completers serve as the treatment groups, which differ from one another in the amount of the treatment received. Of particular interest to stakeholders is the change over time for full completers.

Because transfer, Advanced Placement test, or International Baccalaureate test credit can be used to fulfilled general education requirements, full completers often differ from one another in the means by which the requirement was fulfilled. It is JMU's opinion that the ideal means by which to fulfill the requirement is to complete general education coursework at JMU. It is the average pretest and posttest scores of full completers who have fulfilled at least one course at JMU that are typically used to inform value-added inferences about the general education curriculum.

Type of Score Reported: Scores are typically not reported to individual students since mechanisms are not currently in place to ensure that those scores are informative and accurately interpreted. Instead, scores are reported for groups of students. In order to understand the index that is used to convey change over time to stakeholders, a hypothetical example is needed. Suppose that the difference between the average pretest and posttest raw scores of the JMU full completers was 5 points. A 5 point difference may seem sizeable if the total number of points possible on the instrument were 30, but insignificant if the total number of points were 100. Because the difference in average raw scores between pretest

and posttest is dependent on the raw score scale, the standardized mean difference is used to capture, on average, the extent to which scores change over time. The specific standardized mean difference index employed is Cohen's *d* defined as follows:

$$d = \frac{M_{post} - M_{pre}}{SD_{pre}},$$

with M_{post} representing the average posttest raw score, M_{pre} the average pretest raw score and SD_{pre} the standard deviation of the pretest raw scores. The interpretation of the resulting *d* is best conveyed using an example: if $d = 0.5$, the average posttest scores are half a standard deviation unit higher than average pretest scores. In addition to Cohen's *d* and the descriptive statistics for the pretest and posttest scores, the correlation between pretest and posttest scores should also be reported so that meta-analytic studies that combine the standardized mean difference across cohorts and measures can be pursued.

Sample Requirements: When value-added interpretations are sought, the data collection scheme involves two, as opposed to one, administrations of an instrument. The increased number of measurement occasions often makes it cost-prohibitive for higher education institutions to assess all students. At JMU we minimize costs and maximize our ability to generalize to the entire student body by administering a value-added instrument at two time points to a *random* sample of students. The size of these random samples is large enough to make valid inferences about student learning, but generally much smaller than the samples used for standards-based assessments.

Test Score Consequences for the Examinee: Because different students are completing different value-added instruments, high stakes cannot be associated with test scores. For instance, suppose that standards were available to classify students into one of three proficiency categories categories (e.g., Basic, Proficient, Advanced) on all value-added instruments. It may seem reasonable to require all students to score in at least the Proficient category at posttest on whichever test is randomly assigned. However, due to the inherent differences in the difficulty of various general education clusters, it may be easy to be classified as Proficient in one cluster and relatively difficult in another. The student administered the instrument in the difficult cluster would be at a disadvantage in trying to meet the requirement of Proficiency. To ensure fairness, either all students would need to be administered all cluster instruments (which, as aforementioned, is cost-prohibitive) or the standards for proficiency would need to be forced to be comparable across subareas, a phenomena unlikely, if not impossible, to have happen.

The above discussion was not meant to imply that we should *not* strive to have students be proficient in all subareas of general education. Instead the discussion was meant to elucidate why personal consequences cannot be imposed on test scores when different examinees are completing different instruments. The inability to tie high personal consequences to test scores is more a function of the data collection scheme, which is considered necessary due to its cost efficiency. The above discussion was also not meant to imply that we should *not* set standards on our value-added instruments. Standards on value-added instruments add incredibly useful information when utilized with value-added instruments. For instance, although Cohen's *d* may reveal a considerable increase in proficiency over time, it could be that the majority of students are still not "proficient" at the time of posttest. Using the standardized mean difference in conjunction with the percentage of students meeting proficiency standards is an ideal reporting mechanism that we are

currently working toward providing. It should be kept in mind that if the expectation is for all students to meet proficiency at posttest, the need for value-added assessments is diminished since the same end is desired of all students, regardless of their pretest level.

Although personal stakes are not associated with student's test scores, students are required to participate on Assessment Days, where the value-added instruments are completed. Stakes are associated with participation. If students do not participate in Assessment Day their course registration is blocked. Because there are no personal consequences tied to an individual's test score, there is concern as to how valid these scores are as an indicator of what a student knows and is able to do. A large body of research at JMU is currently focused on how to statistically accommodate the low-stakes nature of the test when computing test scores and there is encouraging evidence that our students' motivation to do well on these tests is reasonably high for most students.

Test Security: Because our value-added instruments are low-stakes for the examinees the possibility of cheating is minor compared to the high-stakes standards-based tests. Even still, similar efforts are made to ensure test security. The only security mechanism not fully in place is the existence of multiple forms of our value-added instruments. If the decision was ever made to increase the stakes associated with our value-added instruments, the creation of multiple forms would be essential. Thus, both the financial costs and the costs associated with the time faculty need for creating new items would need to be considered.