Program assessment includes the evaluation of curriculum, educational experiences, and specifically designated educational outcomes. Accrediting agencies use outcome measures to assess the extent to which education programs meet standards for development of cognitive and psychomotor competence. Program assessment in higher education involves a four-component assessment loop, which was outlined in a preceding report (Figure 1).\(^1\) Educators should develop assessment instruments after having identified intended program outcomes. For accreditation by the Commission on Accreditation of Athletic Training Education (CAA TE), these outcomes are well-defined in several documents:

1. CAA TE Standards\(^2\)
2. National Athletic Trainers’ Association Educational Competencies (4th ed.)\(^3\)
3. Board of Certification Role Delineation Study (6th ed.)\(^4\)

The university’s mission and strategic plan, and the academic program’s objectives, should also be used to develop an assessment instrument. Failure to validate an assessment instrument can lead to inappropriate program modification. The purpose of this report is to review validation of assessment instruments.

**Validity**

The extent to which measurements derived from an instrument represent a construct that the instrument was intended to measure is referred to as validity.\(^5\) Program assessment instruments should be validated and statistical expertise should be available at a university to aid a program director in confirming the validity of measurements derived from instruments that were based on the Competencies\(^5\) or Role Delineation Study.\(^4\) The most appropriate statistical technique for validation of a program evaluation instrument is confirmatory factor analysis, which evaluates the consistency of responses to instrument items with defined outcomes.\(^6\) In other words, instrument validation is focused on confirmation that we are properly measuring our stated program objectives (e.g., effective clinical instructor performance, appropriate student professional behaviors, student attainment of clinical competencies).

An exploratory factor analysis identifies “themes” that may be represented within responses to a large number of instrument items.\(^6,9\) Factor loadings identify the structure of the data.\(^6,9\)
Assessment of Instrument Validity

In the early stage of instrument development, program directors and clinical coordinators are limited to assessment of the face validity and user friendliness of its design. Face validity is established by expert confirmation of information accuracy and appropriate format. Data derived from an instrument can be analyzed in summary format (i.e., descriptive statistics), which includes values for central tendency (mean, median, mode) and variability (standard deviation, standard error; Table 1). Because factor analysis requires a substantial amount of data (general rule: 10 respondents per instrument item; 10 respondents × 50 items = 500 respondents), a considerable amount of time may be required to accumulate the necessary data. Analysis prior to acquisition of the necessary volume of data can produce inaccurate results and is strongly discouraged. Thus, instrument validation may require several years of data accumulation from multiple cohorts of students. “Factor loading” is the term used to describe the correlation between a factor (component) and responses to a given item. This can confirm that a factor representing a program objective has been assessed multiple times by different instrument items. The determination of “loading” involves assessment of a correlation matrix between factors (components) and each of the instrument items. A correlation coefficient of 0.6 or higher is considered a strong relationship, which is an indication that a factor explains a substantial amount of the variability in responses to an instrument item.

When multiple instrument items load on the same factor, a similar construct has been assessed by the items. For example, multiple item responses might relate to the extent to which a clinical instructor leaves a student unsupervised. A question about supervision can be asked in an obvious way, e.g., “My approved clinical instructor (ACI) does not leave me unsupervised during my clinical experiences” and in a more indirect manner, e.g., “My ACI gives me autonomy by allowing me to cover morning weight lifting sessions all by myself.” Although multiple items may assess a similar aspect of program quality, asking a question in multiple ways may help to identify important factors.

Example of Factor Analysis

Table 2 provides an example of exploratory factor analysis output for student evaluations of a clinical instructor. The columns represent 10 factors (components) identified by the analysis of the collective item responses. The first item loads moderately on factors 3, 4, and 9 (inverse association with factors 3 and 4). Inverse correlations result from item phrasing that generates negative responses, e.g., “My clinical instructor does not . . .” associated with a positive factor, e.g., “. . . correct me in front of athletes” or vice-versa. Table 3 also provides information from the same exploratory factor analysis. Factor 1 (component 1) is strongly correlated with each of the six instrument items. This factor could be labeled as “rapport” or “interpersonal relationship” between the student and clinical instructor. Notice the inverse correlations (negative values) for the two items that were phrased in a negative manner.

Table 4 provides information derived from another aspect of the factor analysis output, which indicates that 34% of the total variance in the dataset is explained by the first factor (i.e., the “rapport” between the student and clinical instructor). Another 41% of the total variance is spread across the other nine factors (components). Collectively, the 10 factors explain 75% of the total variance within the dataset. After the number of important factors has been determined, instrument items that differentially correlate with the