ATHLETIC TRAINING EDUCATION JOURNAL
© National Athletic Trainers' Association
www.nataej.org

PERSPECTIVES IN AT EDUCATION

Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research

Laura J. Burton, PhD, Stephanie M. Mazerolle, PhD, ATC

University of Connecticut, Storrs, CT

Context: Instrument validation is an important facet of survey research methods and athletic trainers must be aware of the important underlying principles.

Objective: To discuss the process of survey development and validation, specifically the process of construct validation.

Background: Athletic training researchers frequently employ the use of survey research for topics such as clinical instruction and supervision, burnout, and professional development; however, researchers have not always used proper procedures to ensure instrument validity and reliability for the data collection process.

Description: Four major methods exist to establish the validity of an instrument: face, content, criterion related, and construct. When developing a survey to measure a previously unexplored construct (eg, an athletic trainer's attitudes toward appropriate exertional heat stroke treatment), researchers should employ a four-step process: (1) defining constructs and content domain, (2) generating and judging measurement items, (3) designing and conducting studies to develop a scale, and (4) finalizing the scale.

Clinical Advantages: Establishing the validity of a survey instrument strengthens the data yielded from the data collection process, which allows for greater confidence in the interpretation of the results from the survey.

Conclusions: Construct validity, although a time-intensive process, is necessary to ensure accuracy and validity of the survey instrument.

Key Words: Scale development, scaling procedures, instrument design

Dr. Burton is an Associate Professor of Sport Management in the Department of Kinesiology at the University of Connecticut. Please address all correspondence to Laura Burton, PhD, University of Connecticut, 2095 Hillside Road, Unit 1110, Storrs, CT 06269-1110. laura.burton@uconn.edu

Full Citation:
Burton LJ, Mazerolle SM. Survey instrument validity part I: principles of survey instrument development and validation in athletic training education research. *Athl Train Educ J*. 2011;6(1):27-35.

Survey Instrument Validity Part I: Principles of Survey Instrument Development and Validation in Athletic Training Education Research

Laura J. Burton, PhD and Stephanie M. Mazerolle, PhD, ATC

As the athletic training profession continues to expand the breadth and depth of research within the field, it is important to ensure that appropriate survey development procedures are followed as there are many research topics within athletic training that lend themselves to survey research.[1-2] In a review of the current athletic training literature, there are several studies that have utilized survey design methods for data collection: the evaluation of clinical proficiencies within athletic training education programs,[3] clinical instruction and athletic training education,[4-5] role strain and the clinical instructor,[6] job search strategies and employment issues for undergraduate programs,[7] undergraduate student burnout,[8] the assessment of student learning styles,[9] and teaching practice styles utilized to disseminate concussion related knowledge to students.[10] There are a multitude of reasons why the aforementioned studies used survey methods including allowing the researcher to gain a overview of the topic investigated,[2] access to a random sample of respondents that can allow for generalization of the results, and providing respondents with the freedom to complete the instrument when it is convenient.

An additional, and most significant, benefit to survey research is the ability to measure latent constructs. Latent constructs are variables that researchers cannot directly observe or quantify. Within athletic training, examples of surveys designed to measure latent constructs include, but are not limited to, clinical instruction and supervision,[3-6] athletic training education,[7,9] burnout,[8,11-12] work-family conflict,[13-14] concussion assessment,[15] and teaching and learning styles related to athletic training-specific content.[9-10]

Sammarone Turocy[2] introduced the basics of instrument development and implementation in a *Journal of Athletic Training* article, highlighting the specific applications of survey methods in athletic training. Therefore, the purpose of this paper is to expand upon instrument development by describing the detailed steps in establishing valid surveys for use in athletic training research.

Survey research is appealing to many athletic trainers (ATs) due to the flexibility of data collection and analysis.[2] Furthermore, survey research offers advantages in terms of economy, breadth of data collected, access to larger samples for data analysis, and anonymity, which may help to reduce socially desirable responses. Although there are some limitations to survey research, including the potential for respondents to misinterpret or misunderstand survey questions,[2] many of the potential disadvantages of survey research can be minimized by establishing valid and reliable instruments.

TYPES OF RELIABILITY AND VALIDITY

Reliability and validity are necessary entities of instrument development if researchers are to report with confidence the results obtained from the survey. Reliability refers to the consistency or repeatability of a test or measurement.[16-17] Validity refers to the degree that an instrument actually measures what it is designed or intended to measure.[16] Table 1 presents four common procedures for establishing the validity of an instrument: (1) face validity, (2) content validity, (3) criterion validity and (4) construct validity.[2] Face and content validity are qualitative measures of validity and are often employed in survey research because they are the easiest to ascertain.[16,18] Face and content validity are secured via a panel of experts who judge the survey's appearance, relevance and representativeness of its elements.[2,16] The panel of experts is comprised of individuals with expertise in the area the instrument will measure.[3,12] Face and content validity are important first steps with establishing construct validity because they establish the accuracy and connection among the questions asked and variables measured.

After establishing face and content validity, researchers must establish the instrument's criterion-related validity and construct validity before using it for quantitative analyses. Criterion-related and factorial validity are both empirical measures that follow a rigorous process with assessing the construct validity of an instrument.[18] Criterion-related validity testing demonstrates the accuracy of the measure by comparing it to a previously established and valid instrument or some other external criterion (eg, comparing scores from a short and long version of the same instrument).[16,18] Construct validity is the degree to which an operational measure correlates with the theoretical concept investigated. Construct validity provides the researcher with confidence that a survey actually measures what it is intended to measure. Construct validity allows researchers to draw legitimate conclusions from their findings when and if no criterion or content has been accepted as adequate.[2,16,19]

A review of research in the athletic training literature specifically using survey methods reveals consistent use of face and content validity. However, in the literature,[2-7,15] there is a noted lack of research that has assessed construct validity, which is a necessary and empirically rigorous validation process.[16] Although face and content validity are important considerations in establishing validity, an instrument's construct validity must also be established to enhance the psychometric properties of the instrument. Furthermore, although many of the instruments presented in the athletic training literature are established and valid measures in other populations, those same measures may not transfer to the domain of athletic training. That is, those measures may not assess the same constructs within an athletic training population that are measured in other populations (eg, nursing, coaching). Completing the steps of assessing construct validity can help ensure that the instrument is appropriate for the population investigated. Following, we present the process of construct validation using factor analysis and introduce the steps needed to determine factorial (convergent and discriminant) validity and criterion-related (concurrent and predictive) validity.

Table 1. Methods of Survey Validation

Type	Descripton	Purpose
Face	Evaluation of an instrument's appearance by a group of experts and/or potential participants.	Establishing an instrument's ease of use, clarity, and readability.
Content	Evaluation of an instrument's representativeness of the topic to be studied by a group of experts.	Establishing an instrument's credibility, accuracy, relevance, and breadth of knowledge regarding the domain.
Criterion	Evaluation of an instrument's correlation to another that is deemed unquestionable or identified as the gold standard.	Establishing an instrument's selection over another or establishing the predictability of the measure for a future criterion.
Construct	Evaluation of an instrument's ability to relate to other variables or the degree to which it follows a pattern predicted by a theory.	Establishing an instrument's ability to evaluate the construct it was developed to measure.

THE PROCESS OF CONSTRUCT VALIDATION

Construct validation is necessary to help researchers establish that the survey items actually measure the constructs they propose to measure. (Table 2 provides definitions of important terms used throughout this article.) The steps of construct validation follow the development of content validation and include the steps of factorial validity and criterion-related validity. Factor analysis is most often associated with securing construct validity.[2] It is important to mention, however, that establishing content validity involves more than just factor analysis.[16] Therefore, to adequately address the process of construct validity and to expand upon Turocy's[2] outline of the survey validation process, we will address the steps necessary for scale development and describe the steps to establish construct validity.

Steps of Instrument Development

Instrument development includes four steps.[16] Step one consists of defining constructs and determining domain content. Step two involves generating items for the survey and judging the appropriateness of the items. Step three is to design and conduct studies to test the scale. Lastly, step four involves finalizing the scale based on data collected in the third step.

Step one, defining constructs, begins with a thorough exploration of relevant literature in the domain.[4] In some cases an instrument may exist from a different domain, which may be applicable to athletic training. For example, Henning et al,[20] in their examination of peer-assisted learning, and Clapper and Harris,[12] in their examination of burnout, used pre-existing scales from other research domains to study the respective phenomena in athletic training. If using a pre-existing scale, researchers must establish the instrument's reliability and validity for their sample. In the case where an appropriate scale does not exist, researchers need to use a panel of experts to assist in developing items that measure each construct they plan to investigate. It is also necessary to provide definitions for each construct[3] before item generation begins.

The second step of the instrument development process is item generation and judgment of appropriate items for the survey. Item development, one of the more time-consuming steps for survey research,[2] warrants a strong understanding of the current literature, existing scales (eg, 5- versus 7-point Likert scale), and research agenda/purpose. The type of question researchers use (eg, open ended, close ended, positive vs. negative, double barreled, scale item) will be based on the type of information needed, such as attitudes, beliefs, or behaviors.[2] Another important consideration with item development is data analysis. Open-ended questions may require coding and interpretation due to the variety of potential responses, whereas scale-items and close-ended questions allow for a more streamlined analysis.[2] Regardless of the type of question used, it is important to ensure a logical sequencing of questions, keep language neutral and non-leading, ask only one question at a time, and not underestimate the knowledge of the respondent.[2] Once item development is complete, researchers use a group of experts to review the new instrument for clarity, readability and recommended changes. After establishing face and content validity, the next step in the development of a psychometrically sound instrument is to pilot test the instrument and reduce the number of items in the survey.

To pilot test an instrument, researchers must consider sample size, sample composition, initial item reliability estimates, and the type of validity-related surveys to include in the study's design.[16] Current sample size recommendations for pilot testing an instrument include either a minimum of 10 respondents per instrument item[17] or a minimum overall sample of 300 respondents.[21-22] This recommendation can pose a significant challenge for athletic training researchers; therefore, researchers should consider survey length and available populations early in the survey design process.

After determining the appropriate size of the sample, the next step is to identify an appropriate sample of the population. The sample should be selected from the population of interest for the survey. For example, even if researchers have easy access to athletic training students, they should not be the sample used for pilot testing an instrument whose population of interest is professional ATs. The level of experience and education of students does not

Table 2. Operational Definitions and Instrument Validity

Term	Operational Definition
Latent Construct	Outcomes that are not directly measurable.
Factor	Measure used to understand a latent construct.
Exploratory Factor Analysis	Statistical technique used to explore the possible underlying factor structure of a set of observed variables.
Confirmatory Factor Analysis	Statistical technique used to verify the factor structure of a set of observed variables.
Principle Factor Analysis	A factor extraction method that assumes items in the analysis can be calculated by the extracted factors.
Principle Axis Factoring	A factor extraction method that compares squared multiple correlations (R^2) of each item to all other items included in the analysis.
Maximum Likelihood Method	A factor extraction method that assumes the distribution for each item is normal and all eigenvalues are greater than zero.
Criterion Related Validity	Accuracy of a measure or procedure by comparing it with another measure or procedure that is valid.
Factorial Validity	A form of construct validity that is established through factor analysis.
Orthogonal Rotation	Factor rotation method that assumes factors are independent of each other (ie, uncorrelated).
Oblique Rotation	Factor rotation method that assumes factors are not independent of each other (ie, correlated).
Unidimensionality	A single item is helping the researcher to understand only one latent construct being examined in the survey or assessing only a single construct or latent factor.
Eigenvalue	A single value that represents the amount of variance in all of the items that can be explained by a factor.
Item-to-Total Correlations	Correlations of one item to all remaining items in the instrument.
Inter-item Correlations	Correlations between all items within an instrument.

match the experience and education of professional ATs; therefore, use of students may compromise the instrument's validity.

After pilot testing the instrument, researchers must analyze individual survey items to examine whether the items are unidimensional. Survey item unidimensionality means a single item is helping the researcher understand or assess only one latent construct measured by the survey. A survey item should be unidimensional and help to explain only one construct, not multiple constructs. For example, if the intent of a survey is to address peer education as part of the athletic training education experience, the items designed to examine peer education must address only peer education and not another aspect of the educational experience (eg, interaction with clinical instructors).

Researchers can assess unidimensionality by inspecting inter-item correlations and corrected item-to-total correlations. Inter-item correlations for items intended to measure the same construct should be moderate but not too high (ie, between .30-.60).[23] Inter-item correlations that are high suggest that each of the items are not contributing something unique to the construct, and therefore, are not unidimensional. The concept of unidimensionality also applies to the latent construct the instrument is measuring. To assess unidimensionality of constructs in an instrument, researchers should use confirmatory factor analysis, which is beyond the scope of this article. After assessing dimensionality of items in the pilot study, researchers will use exploratory factor analysis to establish construct validity of the instrument. We discuss exploratory factor analysis (EFA) as a tool to assess construct validity in the next section.

Process of Exploratory Factor Analysis

After assessing the dimensionality of survey items, researchers can use factor analysis as one analytic tool to assess construct validity. Again, construct validity provides evidence that the items in the survey actually measure the constructs they are proposed to represent. Researchers use factor analysis to "examine empirically the interrelationships among the items and to identify clusters of items that share sufficient variation to justify their existence as a factor or construct to be measured by the instrument."[19(p108)] One of the objectives of survey research is to develop a parsimonious survey that will best explain the constructs under investigation. EFA is an important tool for instrument development because it allows researchers to develop a survey that contains the minimum number of items needed to understand the constructs best. That is, EFA provides the researcher with information that will help reduce the number of items in a proposed survey so that the remaining items can best explain the constructs under investigation. As an example, EFA can help the athletic training education researcher develop a survey that uses the fewest number of questions to measure and understand those factors that contribute best to the athletic training education experience. A second benefit of EFA is to examine the underlying factors that structure the instrument.[19] Using the example above, EFA helps researchers determine the most important factors contributing to the athletic training education experience.

Before moving forward in factor analysis, there are options with EFA the researcher must resolve in accordance with the objectives of the particular study. These options include what type of extraction method to use (eg, principle component analysis, principle axis factoring, maximum likelihood), the number of factors to extract, the rotation method to use (eg, orthogonal or oblique), and which items to retain and remove from the instrument (Figure 1).

Decisions in Factor Analysis

Principle component analysis (PCA) is the desired method of EFA to use when researchers are interested in reducing the number of items for an instrument.[16] PCA is useful when researchers have initially developed a survey with several questions (items) and want to reduce the number of questions to best measure the constructs with the fewest number of questions. PCA maximizes all variance in the items, so those items that do not contribute to the understanding of the factor (ie, those with little explained variance) are evaluated for deletion. Another extraction method is principle axis factoring (PAF), which researchers may use when they want to determine the underlying factors related to a set of items.

Following selection of the appropriate extraction method, researchers must evaluate the adequacy of the sample prior to interpreting the EFA output. Sampling adequacy provides the researcher with information regarding the grouping of survey items. Grouping items into a set of interpretable factors can better explain the constructs under investigation. Measures of sampling adequacy evaluate how strongly an item is correlated with other items in the EFA correlation matrix. Measures of sampling adequacy help researchers assess whether the items used in the survey have some type of relationship to one another. Researchers can assess sampling adequacy by examining the Kaiser-Meyer-Olkin (KMO) output provided in the factor analysis. A KMO correlation above .60-.70 is considered adequate for analyzing the EFA output.[16] In addition to examining KMO for sampling adequacy, researchers must also evaluate the correlation matrix of all survey items to determine if the matrix can be analyzed using factor analysis. If the correlation matrix is an identity matrix (ie, there is no relationship among the items),[24] it cannot be analyzed. Bartlett's test of sphericity provides a chi-square output that must be significant, which indicates the matrix is not an identity matrix. If the KMO correlation indicates sample adequacy and Bartlett's test of sphericity indicates the item correlation matrix is not an identity matrix, researchers can move forward with the factor analysis.[16]

The next decision researchers need to make in conducting an EFA is selection of a factor rotational method. Rotational methods (orthogonal or oblique) are tools used within EFA to help make factors in the instrument easier to interpret.[16] The overall goal for instrument development is for the instrument to have a simple structure. A simple structure indicates that each item in the survey helps to explain one and only one particular construct. Using the example of assessing the athletic training education experience, the survey should examine only those factors that are relevant to the athletic training education experience (eg, peer education, clinical instructor interaction, in-class instruction). The questions used in the survey should examine only the factor(s) they are intended to measure. For example, a question about peer education should measure only the factor peer education and not other factors in the survey (eg, clinical instructor interaction). Varimax rotation is the most common form of orthogonal rotation for EFA and will often provide a simple structure.[23] Oblique rotation (Oblimin with Kaiser Normalization) is another option that will provide a simple structure, but it also allows factors to remain correlated in the analysis. Both methods of rotation will provide researchers with information regarding which survey items to retain or delete. Researchers should use both rotational methods to determine the most meaningful solution.[19] After selecting the appropriate method of rotation, researchers will identify the best number of factors to investigate the latent constructs in the instrument.

Factor identification is the process used to ascertain the number of factors to keep in a survey. During the early stages of instrument development, researchers may have an understanding of how many factors are necessary to explain the latent constructs the instrument will measure. Therefore, researchers may determine *a priori* the number of factors in the survey, which can serve as a guide to factor selection. EFA, in addition to the researcher's knowledge prior to instrument development, can identify the number of factors that best explain the latent constructs the survey will measure. EFA provides information regarding the combination of instrument items whose shared correlations explain the greatest total variance (ie, Factor 1). The first factor identified will provide the most information to the researcher regarding the latent constructs the survey measures. For example, EFA can help to identify what is the most important factor to understanding the athletic training education experience.

EFA continues to identify the combination of items that explains the greatest total variance remaining (ie, Factor 2) until it accounts for all item shared correlations.[16] Each subsequent factor will contain additional information to help explain the latent constructs the survey will measure. However, the researcher must decide how many factors should be used to measure the latent constructs. Again, researchers may have determined *a priori* the appropriate number of factors to explain the latent constructs. For example, researchers could have determined during survey development that peer education, clinical instructor interaction, and in-class instruction interaction are the three most important factors to assessing the athletic training education experience. This *a priori* information will assist with determining the appropriate number of factors to retain in the instrument, but researchers have additional tools to help make this decision.

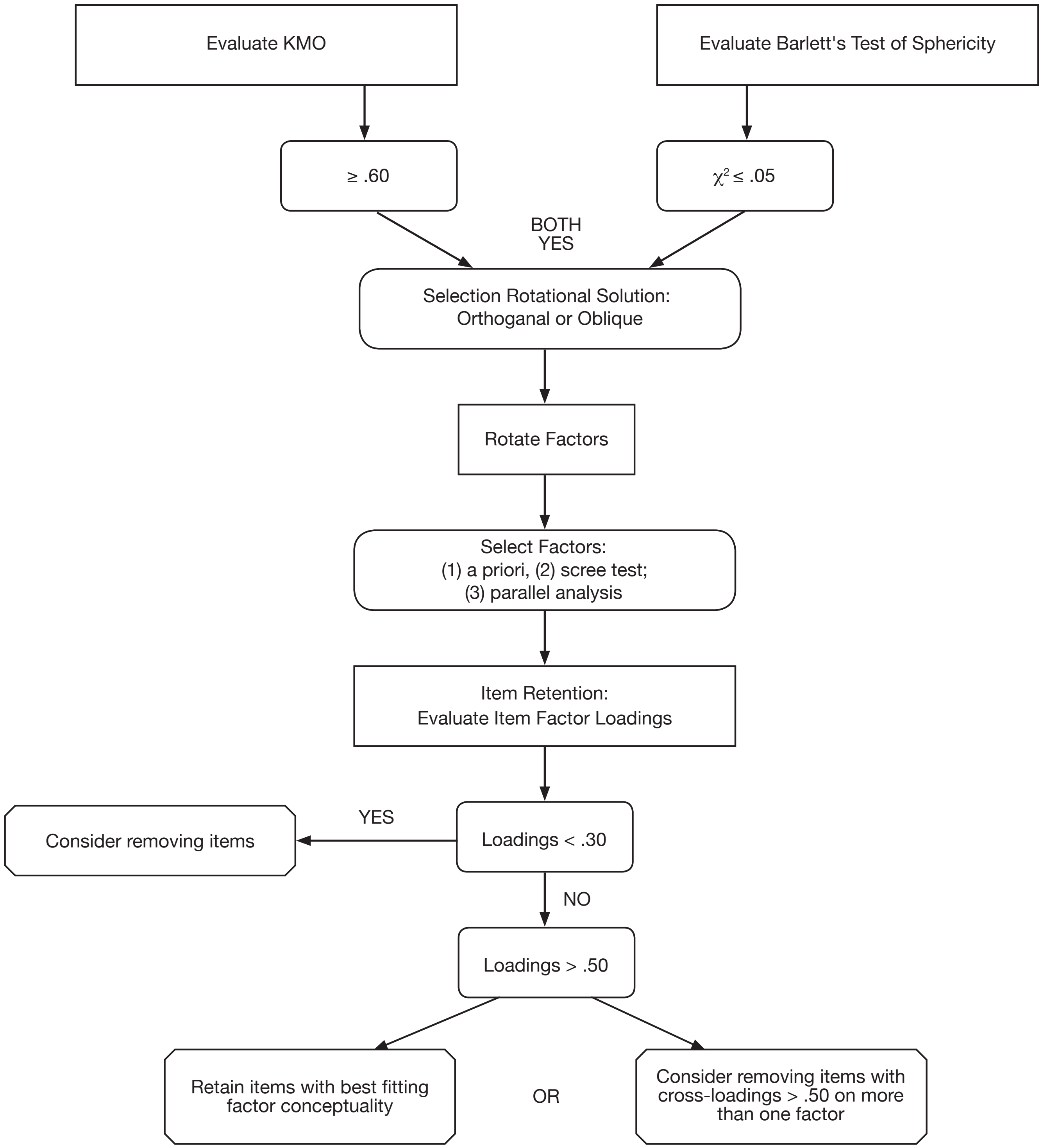

Figure 1. Selection of Principle Component Analysis (PCA) or Exploratory Factor Analysis (EFA)

There are three tools available to researchers to help determine the appropriate number of factors to retain. The first tool is the Kaiser criterion,[24] which recommends that researchers select factors with eigenvalues greater than 1.0. A second tool is the scree test. The scree test examines the scree plot, which is a plot of the eigenvalues along an x-y axis. The point at which the curve decreases and straightens out (ie, the "elbow" of the graph) is the point where researchers should include all factors before and at the elbow (Figure 2). Parallel analysis is the third tool that researchers can use to graphically represent factors to consider retaining in the instrument. For additional information regarding parallel analysis, we recommend reviewing the article on factor retention by Hayton, Allen, and Scarpello.[25] There is not one method of factor retention recommended over the others, and researchers are encouraged to use all three methods when determining the number of factors to retain in the instrument.

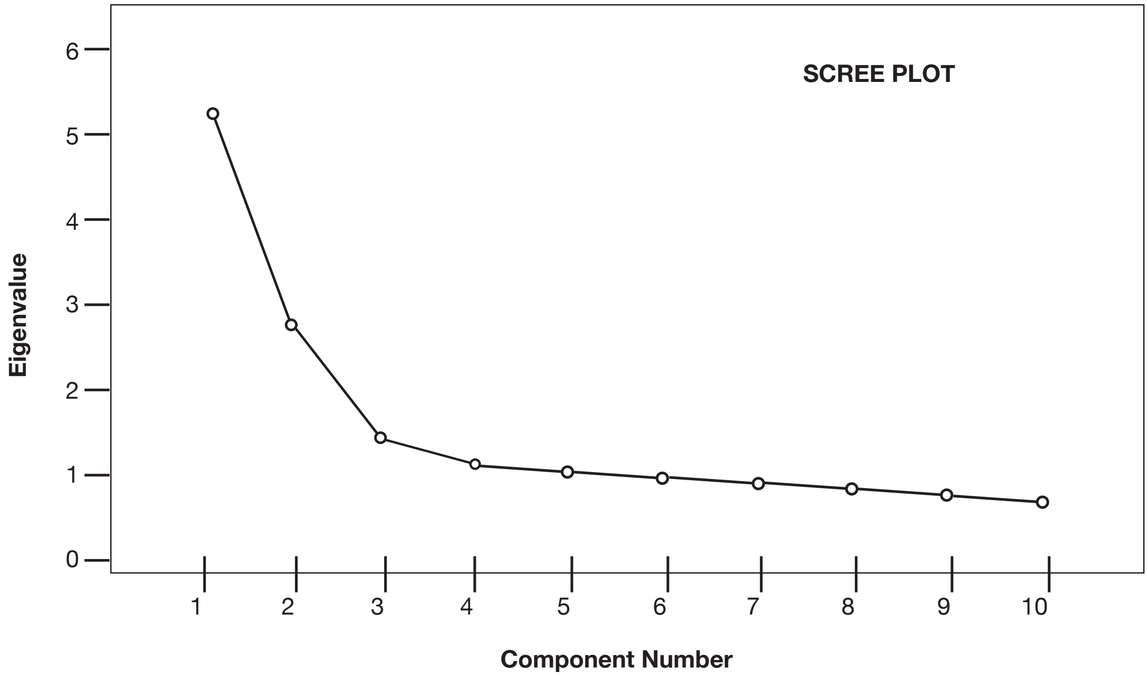

Figure 2. Factor Selection Using a Scree Plot. A three factor solution is suggested by the scree plot. The curve of the line ends at the third factor.

After researchers select an appropriate number of factors for retention (based on a prior theory, eigenvalues, scree test, and parallel analysis), they must use item trimming and item retention to move the survey into its final form. Researchers will review the output for factor loadings to determine which items should remain in the survey and which items should be reviewed for deletion. Using the assessment of athletic training education experiences as an example, if the survey contains eight items intended to measure peer education, the researcher will evaluate the loading of these items on the peer education factor. Instrument items should load on only one factor at .50 or higher with no cross loadings (ie, loading on more than one factor) of >.30.[21] Returning to the athletic training education example, items that measure peer education but fail to load on the peer education factor per this criteria should be considered for deletion.

If researchers select an oblique rotation (as described earlier) to help simplify the instrument, they will review the pattern matrix output of factor loadings to determine which items to retain or delete from the survey. However, if researchers use an orthogonal rotation, they will use the factor matrix output to determine which items to retain or delete. In addition to examining item loadings in the pattern matrix or factor matrix output, researchers should also consider both the face and content validity of each item. Researchers may consider some instrument items important enough to analyze further regardless of their loadings in either the pattern or factor matrix output. Researchers may also use reliability analysis to determine item retention or deletion.[16] An additional analytical tool available to researchers is the corrected item-to-total correlation provided in the survey reliability procedures.[16] These correlations reflect how correlated one item is to the remaining items in the set of items. Researchers should consider removing items with low corrected item-to-total correlations because these items may not provide meaningful information to help explain the latent constructs under investigation.

In addition to an examination of factor loadings of individual instrument items, researchers must also consider the number of items in a survey to complete the item selection process. The number of items appropriate for a survey will depend on the content domain examined. For example, instruments assessing a broad domain (eg, AT behavior regarding diagnosis and treatment of exertional heat stroke) will require more items to adequately investigate the domain.[16] As a general rule, each factor within a survey should contain a minimum of four items. The quality of survey items should guide this process, however. Researchers should write items in such a way that each item assesses a distinct aspect of the factor it represents. Items worded too closely will not improve the content validity of the instrument even though closely worded items can yield higher inter-item correlations and an increased coefficient alpha (reliability).[16]

Sample size must always be an important consideration in EFA. If the sample size is inadequate, researchers may be unable to interpret the results of the factor solution or may not be able to repeat the solution when using a different sample. As aforementioned, an adequate sample should contain at least 10 respondents per survey item or a minimum of 300 respondents. Researchers will use the measure of sampling adequacy (KMO) to evaluate sample size.

Researchers should also be aware of unstable factor solutions that may occur in exploratory factor analysis. Statistical programs (eg, SPSS) provide information, specifically the number of iterations needed to reach a solution, on the output analysis of factor analysis. A high number of iterations (more than 200) needed to reach a solution implies a potentially unstable solution, which suggests not interpreting the factor analysis output. An additional process researchers may use to examine the stability of the factor solution is to run both a principle component analysis and a principle axis factoring. Different solutions from these extraction methods may indicate an unstable solution. When using a smaller sample (ie, samples of less than 300 participants), it is beneficial to run both extraction methods to evaluate the stability of the factor solution.[19]

Exploratory factor analysis is an important statistical analysis that researchers can use to evaluate an instrument's construct validity. The next section highlights additional measures researchers may use to examine construct validity further.

ADDITIONAL FORMS OF CONSTRUCT VALIDITY

Convergent and Discriminant Validity

Convergent validity measures how closely related the developing survey is to a preexisting survey that measures a similar construct. Researchers can determine convergent validity by selecting a survey that is similar to the survey under development and then administer both surveys to the same sample. Within athletic training education, researchers could use surveys developed in allied health education (eg, physical therapy, nursing) to evaluate convergent validity. The survey under development will demonstrate convergent validity if there are significant and strong correlations between its measures and the measures from a previously validated survey that assesses a similar construct. For example, in a study investigating undergraduate burnout by Riter et al,[8] the researchers utilized the Mashalch Burnout Inventory to collect data. Using the same sample, the researchers could administer the Athletic Training Burnout Inventory (ABTI) to assess convergent validity.

In contrast, discriminant validity requires that the instrument not be highly correlated with the measures from another instrument that is supposedly different conceptually.[16] To measure discriminant validity, researchers can follow the same procedures as convergent validity but administer a survey that is known to measure a different construct. The multitrait-multimethod matrix technique (MTMM technique) is the method most often used by researchers to assess convergent and discriminant validity. For additional information regarding the procedures for the MTMM technique, we encourage readers to review survey development texts.[16,19]

Predictive and Concurrent Validity

Predictive and concurrent validity are measures of criterion-related validity, which address the question "What is the relationship between scores on the instrument and some external criterion that provides a more direct measure of the targeted characteristics?"[19(p189)] Both types of validity help researchers evaluate how well the instrument under development is able to measure its intended constructs. Researchers can assess concurrent validity by administering the instrument to a specific sample and use additional data from the same sample to measure the construct under development.

Predictive validity refers to the ability of the instrument to predict future behavior. To establish predictive validity, the instrument is administered to a sample with additional measures of a different predictive external criterion obtained at a point later in time.[19] Researchers administer this predictive measure to the sample at a time following the administration of the instrument under development. Measures from the new instrument and the predictive criterion are compared to determine how well the new instrument predicted the respective behavior.

REPORTING RESULTS OF EFA

A final step in the process of instrument development is reporting the results of the process in a peer-reviewed journal. When writing up the results of the EFA to support construct validity of the survey, there are specific output tables and results to report in the manuscript. To begin, researchers should report in one table individual item descriptive statistics including means and standard deviations. In addition, researcher should report the KMO and Bartlett's test of sphericity results. An additional table should report eigenvalues and variance explained by each factor. Finally, researchers should report the pattern or structure matrix (rotated matrix) with factor loadings for each item retained in the survey, as well as the correlations among the factors. It is important that researchers report the findings of construct validity in peer-reviewed journals because this step will improve the instrument development process within the respective domain.

CONCLUSION

Construct validity, as one method of survey validation, is a necessary step in the research process to ensure that a multi-item survey instrument accurately measures the constructs under investigation. The survey development articles presented in the athletic training literature have used less rigorous, qualitative measures (face validity and content validity) to determine validity. To ensure a psychometrically sound instrument, however, researchers must also establish construct validity. To demonstrate construct validity in an instrument with established validity in other populations, not all steps of instrument development process must be completed as outlined in this paper. At a minimum though, researchers should use factor analysis to establish one aspect of the instrument's construct validity with the intended population. While securing construct validity may appear to be a daunting process, it is a necessary process.

REFERENCES

1. Turocy PS. Overview of athletic training education research publications. *J Athl Train*. 2002;37(suppl 4):S162-S167.
2. Turocy PS. Survey research in athletic training: the scientific method of development and implementation. *J Athl Train*. 2002;37(suppl 4):S174-S179.

3. Walker SE, Weidner TG, Armstrong KJ. Evaluation of athletic training students' clinical proficiencies. *J Athl Train*. 2008;43(4):386–395.
4. Lauber CA, Toth PE, Leary PA, Martin RD, Killian CB. Program directors' and clinical instructors' perceptions of important clinical-instructor behavior categories in the delivery of athletic training clinical instruction. *J Athl Train*. 2003;38(4):336–341.
5. Laurent T, Weidner TG. Clinical instructors' and student athletic trainers' perceptions of helpful clinical instructor characteristics. *J Athl Train*. 2001;36(1):58–61.
6. Henning JM, Weidner TG. Role strain in collegiate athletic training approved clinical instructors. *J Athl Train*. 2008;43(3):275–283.
7. Stigler VG, Meador R, Tsuchiya M. Job search and employment-related issues in athletic training education programs. *J Athl Train*. 1999;34(4):368–374.
8. Riter TS, Kaiser DA, Hopkins T, Pennington TR, Chamberlain R, Eggett D. Presence of burnout in undergraduate athletic training students at one western US university. *Athl Train Educ J*. 2008;2(Apr-Jun):57-66.
9. Stradley SL, Buckely BD, Kaminski TW, Horodyski M, Fleming D, Janelle CM. A nationwide learning-style assessment of undergraduate athletic training students in CAAHEP-accredited athletic training programs. *J Athl Train*. 2002;37(suppl 4):S141-S146.
10. Covassin T, Elbin R, Stiller-Ostrowski JL. Current sport-related concussion teaching and clinical practices of sports medicine professionals. *J Athl Train*. 2009;44(4):400–404.
11. Hendrix AE, Acevedo EO, Hebert E. An examination of stress and burnout in certified athletic trainers at Division I-A universities. *J Athl Train*. 2000;35(2):139-144.
12. Clapper DC, Harris LL. Reliability and validity of an instrument to describe burnout among collegiate athletic trainers. *J Athl Train*. 2008;43(1):62–69.
13. Mazerolle SM, Bruening JE, Casa DJ. Work-family conflict part I: Antecedents of work-family conflict in Division I-A athletic trainers. *J Athl Train*. 2008;43(5):505-512.
14. Mazerolle SM, Bruening JE, Casa DJ. Burton LJ. Work-family conflict part II: The impact on job and life satisfaction in Division I-A athletic trainers. *J Athl Train*. 2008;43(5):512-522.
15. Piland SG, Motl RW, Ferrara MS, Peterson CL. Evidence for the factorial and construct validity of a self-report concussion symptoms survey. *J Athl Train*. 2003;38(2):104–112.
16. Netemeyer RG, Bearden WO, Sharma S. *Scaling Procedures: Issues and Applications*. London: Sage; 2003.
17. Nunnaly J. *Psychometric Theory*. New York, NY: McGraw-Hill; 1978.
18. Arnold BL, Gansneder BM, Perrin DH. *Research Methods in Athletic Training*. Philadelphia, PA: FA Davis; 2005.
19. Gable RK. *Instrument Development in the Affective Domain: Measuring Attitudes and Values in Corporate and School Settings*. 2nd ed. Boston, MA: Sage; 1993.
20. Henning JM, Weidner TG, James J. Peer-assisted learning in the athletic training clinical setting. *J Athl Train*. 2006;41(1):102-108.
21. Comery AL, Lee HB. *A First Course in Factor Analysis*. Hillsdale, NJ: Erlbaum; 1992.
22. Tabachnik BG, Fidell LS. *Using Multivariate Statistics*. 4th ed. Needham Heights, MA: Allyn and Bacon; 1992.
23. Pett MA, Lackey NR, & Sullivan JJ. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*. Thousand Oaks, CA: Sage; 2003.
24. Kraiser H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*. 1958; 23:187-200.
25. Hayton, JC Allen, DG Scarpello V. Factor retention decisions in exploratory factor analysis: a tutorial on parallel analysis. *Organizational Res Methods*. 2004;7(2):191-205.