[Journal of International Agricultural and Extension Education](https://newprairiepress.org/jiaee)

Volume 27 | Issue 2 Article 3

4-1-2020

Scale Development and Validation: Methodology and Recommendations

Kevan W. Lamm
University of Georgia

Alexa J. Lamm
University of Georgia

Don Edgar
University of Georgia

Recommended Citation

Lamm, K. W., Lamm, A. J., & Edgar, D. (2020). Scale Development and Validation: Methodology and Recommendations. *Journal of International Agricultural and Extension Education, 27*(2), 24-35. https://doi.org/10.4148/2831-5960.1115

Scale Development and Validation: Methodology and Recommendations

Abstract

The importance of valid and reliable data and its collection is fundamental to empirical research; however, there remain inconsistent approaches to creating robust scales capable of capturing both valid and reliable data, particularly within international agricultural and extension education contexts. Robust scale development consists of five areas for validation: content, response process, internal structure, external structure, and consequential. The purpose of this guide was to provide methodological recommendations to improve scale development rigor and adoption and to provide a set of functional principles to aid researchers and practitioners interested in capturing data through developed, or adapted, scales. Additionally, the information summarized provide a benchmark upon which to evaluate the rigor and validity of reported scale results. A consistent framework should provide a common lexicon upon which to examine scales and associated results. Proper scale development and validation will help ensure research findings accurately describe intended underlying concepts, particularly within an international agricultural and extension education context.

Keywords

scale development, validity, quantitative analysis

Scale Development and Validation: Methodology and Recommendations

Kevan W. Lamm
University of Georgia

Alexa J. Lamm
University of Georgia

Don Edgar
University of Georgia

Abstract

The importance of valid and reliable data and its collection is fundamental to empirical research; however, there remain inconsistent approaches to creating robust scales capable of capturing both valid and reliable data, particularly within international agricultural and extension education contexts. Robust scale development consists of five areas for validation: content, response process, internal structure, external structure, and consequential. The purpose of this guide was to provide methodological recommendations to improve scale development rigor and adoption and to provide a set of functional principles to aid researchers and practitioners interested in capturing data through developed, or adapted, scales. Additionally, the information summarized provide a benchmark upon which to evaluate the rigor and validity of reported scale results. A consistent framework should provide a common lexicon upon which to examine scales and associated results. Proper scale development and validation will help ensure research findings accurately describe intended underlying concepts, particularly within an international agricultural and extension education context.

*Keywords***:** scale development, validity, quantitative analysis

Introduction

Data, its collection, its analysis, and the meaning that is drawn from it is ubiquitous within research and evaluation across disciplines and is fundamental activity of science (DeVellis, 2017). However, despite the consistent agreement on the importance of valid and reliable data, there have been challenges associated with data collection instruments, particularly scales (Emmerson & Neely, 1988). Frequently, data are collected in a manner that jeopardizes both validity and interpretability (McKnight, McKnight, Sidani, & Figueredo, 2007). These potential issues are noteworthy as they represent the fundamental building blocks of analysis, borrowing a classic warning from computer programming, “garbage in, garbage out” (Lordo, 2001).

There has been a notable gap in the international agricultural and extension education literature providing a set of scale development guidelines and methodologies. In social science, research measurements are most often derived from theory which leads to conceptualization of problems toward measurement (DeVellis, 2017). Without a common understanding of these principles, a potential for misinterpretation and missed opportunities are inherent towards accurate measurement of what is to be measured. Historically, one of the main limitations with scale development is the perception of opaqueness, that is clarity between measures (Nunnally & Bernstein, 1967). Therefore, scales that are developed in complementary disciplines are frequently employed (DeVellis, 2017); however, the nuance and uniqueness of efforts within a particular context are frequently omitted. Within the international agricultural and extension education literature, the need for contextually relevant scales has been identified: “A standardized scale should help to provide a common measure of capacity among RAS [rural advisory service] networks and to facilitate knowledge sharing using a standard set of capacity items” (Lamm, Lamm, Davis, & Swaroop, 2018, p. 52).

In an age when there is increasing emphasis on data collection and analysis, the utility and appropriateness of scales used to collect data are vital. These trends are reinforced by factors such as a culture towards increased scrutiny and accountability in international settings (Taras, Rowney, & Steel 2009). Additionally, the fidelity of scales to appropriately measure that for which they are intended and in the context in which they are intended, is a persistent trend (DeVellis, 2017). These trends are amplified within international agricultural and extension contexts where evaluation, capacity assessments, capacity building, and other activities require high levels of collaboration among many local and international actors (Davis & Sulaiman, 2014).

Literature Review and Methodological Recommendations

This work is based on the scale development recommendations of Crocker and Algina (1986) as well as Messick (1995). Overall, there are five main areas for validation: content, response process, internal structure, external structure, and consequential. Adhering to an established set of methodological process steps helps to ensure consistency and predictability across scale development endeavors (Crocker & Algina, 1986).

Content Validity

Content validity refers to the ability of the scale to appropriately measure what it has been intended to measure. Initially, the researcher or team should consult the literature through an exhaustive search to determine aspects that will construct the scale. Once the review of literature has been completed, experts should review the proposed scale to confirm that the intent and purpose of the scale is valid. Any revisions, edits, or changes should be made to align further with respects to the expert's opinion. DeVellis (2017) refers to this as sampling adequacy in which the content domain is revealed through a specific set of items. This is critical to further steps in scale development as it sets the parameters of what is to be measured. Consequently, reliable data is consistent – it reports what is being measured with dependable accuracy towards the intended response. Furthermore, it should be noted that data deemed reliable could be invalid. The scale could produce data deemed consistent but not valid relative to the underlying phenomena of interest. Data to be gathered must be valid in relation to the scope of the study, constructs, participants, and setting (McMillan & Schumaker, 2010).

Content validity in social sciences can be established in several ways. Examples include a panel of experts, literature review, or a Delphi method approach (Crocker & Algina, 1986; DeVellis, 2017; Messick, 1995; Williamson, 2007). However, the most robust approaches are those that use multiple sources to establish content validity. For example, literature review, while critical, may be insufficient to adequately define the content domain the scale is intended to cover. Therefore, the use of an expert panel may provide additional insights and triangulation of the concept that may otherwise be unconsidered. Given the context of international agricultural and extension education programs, the ability to convene an expert panel in person may be somewhat limited. Therefore, a Delphi method may provide a methodological approach to gather comprehensive viewpoints in an efficient manner, particularly when the process is facilitated online (Lamm, Lamm, Davis, & Swaroop, 2017). The focus on the Delphi method is also consistent with the international agricultural and extension education literature where previous research has recommended "to use the Delphi process to gather insights from RAS [rural advisory service] experts for future research" (Lamm et al., 2017, p. 103). The use of the Delphi method allows specifically identified experts to provide, review, consolidate, and work towards consensus on provided insights. Decidedly, the Delphi method has been used to gain experts' opinions regarding "content validation of constructs to be used in quantitative research" (Garson, 2014, Chapter 8, para. 1). A more comprehensive description of the Delphi method is provided as a set of operational guidelines; however, the underlying considerations are also applicable to establishing a more traditional expert panel.

To arrive at a consensus amongst experts, the Delphi method uses an iterative approach (Garson, 2014). Generally, the Delphi process includes a set of predictable steps. First, a questionnaire is created and sent to a panel of experts. Second, a panel of experts will complete the questionnaire and return it to the researcher. Third, the researcher analyzes data gathered from the expert panel. Finally, the researcher will modify the questionnaire based on their analysis and a second iteration of processing will be initiated (Stines, 2003). Based on a review of the literature, Gliddon (2006) found that the number of iterations included in a typical Delphi research study ranged from two to eight rounds. However, three rounds were identified as an optimal number of iterations for the majority of applications (e.g. Delbacq, Van de Ven, & Gustafson, 1975; Gliddon, 2006).

According to Czinkota and Ronkainen (1997), "the selection of the experts is critical to the success of a Delphic study" (p. 152). The Delphi method harnesses the expertise of individuals familiar with the issue of interest (Garson, 2014); however, one of the primary criticisms of the Delphi method has been, "if panelists are misinformed about a topic, the use of Delphi may only add confidence to their ignorance" (Roy & Garai, 2012, p. 39). Consequently, "the individuals comprising the expert panel should represent the research purpose in a way that legitimates the outcome of the Delphi process" (Garson, 2014, Chapter 6, para. 2).

In addition to selecting the correct experts to constitute the panel, identifying the correct number of panelists has also been explored within the literature. For example, Boulkedid Abdoul, Loustau, Sibony, and Alberti (2011) found that of 80 Delphi studies conducted between 1978 and 2009, the median size of the expert panel was 17. However, Skulmoski, Hartman, and Krahn (2007) found that across 41 doctoral dissertations that employed the Delphi method, the median was 28 within a range of 8 to 345. Overall, researchers have found that panels with larger numbers of experts tend to have lower agreement indexes (Meijering, Kampen, & Tobi, 2013); however, studies that desire a higher degree of precision in outcomes should have sufficiently large panels to address potential issues associated with unintended panel homogeneity (Garson, 2014).

Next, relevant organizations and individual experts are identified (Okoli & Pawlowski, 2004). A diversity of expertise within a common domain should provide sufficient coverage of the topic. Heterogeneity within the panel helps to ensure a variety of perspectives, while the Delphi method ensures all panelists have equal opportunity to contribute, mitigating the tendency to defer to the opinions of the most experienced experts or conformity to groupthink (Garson, 2014). The outcome of a robust and properly administered Delphi process may provide a foundation for establishing content validity. However, the Delphi outcomes should be employed as one source of content validity where further triangulation of content concepts through a literature review and other actions are utilized.

Based on the results of the content gathering process, scale development can then proceed into the item generation stages. It is the responsibility of the individual, or team, working on the scale to identify and decide on the best item type and format to ensure the underlying phenomenon is appropriately captured. Context and intended audience types should be considered during the item type selection process. Furthermore, careful consideration of context variables such as connectivity, reading level, available time, among other factors are considerations throughout this process (DeVellis, 2017). These items are particularly relevant when considering international contexts where infrastructure and other criteria may vary (Ganpat, Ramdwar, Stripling, & Roberts, 2013).

An additional consideration related to item type selection is the utility of item types to capture the underlying phenomena of interest. For example, phenomena that have binary outcomes (e.g., yes or no) may be effectively captured with binary item types. However, phenomena that have gradations of existence responses may be better captured through Likert-type items (Rossi, Lipsey, & Freeman, 2004). Regardless of item types, it is critical that the underlying content is appropriately captured through the scale development process; doing so will be necessary to establishing content validity. Based on the ubiquity of Likert-type scale items present within the existing international agricultural and extension education literature, the following areas for establishing validity are based on scales composed of Likert-type items; however, the specific guidelines, standards, and approaches should be amended to suit different item types as appropriate. Although beyond the scope of the present work, additional resources are available and should be employed specifically related to item development; for example, DeVellis (2017), among others.

Response Process Validity

Following the establishment of a scale, response process validity should be examined. The literature recommends a small sample of respondents or experts adequately suited to knowledgeably evaluate the face validity and interpretability of a scale. Any items or directions that are unclear should be revised and retested as appropriate (Crocker & Algina, 1986).

Establishing response process validity has both functional and strategic benefits. From a functional perspective, it is important to establish whether the individuals that the scale is intended for are able to complete it as expected. From an international perspective, considerations such as translation and localization are also important considerations (Radhakrishna, 2006). Many of the considerations associated with establishing content validity and developing the proposed scale are examined during the response process validity process. In addition to the functional importance of the step, establishing response process validity also has strategic value in the scale development process. Specifically, the time taken to intentionally examine the scale characteristics at the preliminary stages has the potential to mitigate time and expenses required to revise the instrument and recollect data during primary data collection. Additionally, scale items or directions that are unclear, or worse, incorrect, may bias all subsequent analyses. Therefore, it is important to establish response process validity prior to additional data collection or analysis (Crocker & Algina, 1986).

During an initial data collection associated with establishing response process validity, specific recommendations have been established in the literature (DeVellis, 2017). For example, if utilizing an online data collection instrument, periodic prompts are recommended for respondents to indicate whether they understood the requested response or if there is any confusion with the response process associated with the instrument. Responses to instrument items should be analyzed and any confusion over directions or response expectations should be examined. Expected outcomes from this process are to identify and/or modify items requiring updating or adjustment. If utilizing paper-based instruments, include a section at the end for respondents to identify items that were unclear or confusing. Alternatively, a facilitated administration of the instrument would provide the researcher real-time input and/or feedback from representative respondents.

Any modifications made to the scale based on establishing response process validity should be documented and considered relative to the previously established content validity. It is important to ensure any modifications don't have unintended consequences in the scale's ability to serve its intended purpose. An iterative process may be required to test, revise, and retest scale items, as well as the overall scale (DeVellis, 2017).

Internal Structure Validity

Establishing internal structure validity has been suggested within four primary domains (Clark & Watson, 1995; Crocker & Algina, 1986; Messick, 1989). First, individual item response distributions should be analyzed using descriptive statistics to ensure acceptable characteristics. Second, internal consistency amongst items should calculated using Cronbach's alpha. Third, validation of the hypothesized latent variables should be examined using exploratory factor analysis. Finally, latent variable structure should be further analyzed using confirmatory factor analysis (Crocker & Algina, 1986).

Initially, descriptive statistics should be calculated to determine response distributions for each individual item to examine acceptability prior to conducting subsequent analyses (Clark & Watson, 1995). Result distributions should be further analyzed for skewness and kurtosis (Ferguson & Cox, 1993). Skewness values less than two and kurtosis values less than seven should be considered acceptable given established thresholds for factor analysis within psychological research (Fabrigar, Wegener, MacCallum, & Strahan, 1999; West, Finch, &

Curran, 1995). Following individual item analysis, an overall index score, generally the mean score, should be calculated in order to determine values of individual items. The resulting index score is then analyzed using descriptive statistics. Mean and standard deviation scale scores should be analyzed relative to individual item analysis and expectations.

To establish internal consistency, Cronbach's alpha should be calculated for each latent variable. Based on established social science standards, an observed value of 0.70 or above should be considered acceptable (Cortina, 1993; Schmitt, 1996; Streiner, 2003). If value(s) are observed below the acceptable range, additional individual item analysis may be required (DeVellis, 2017).

Although internal consistency has been identified as a necessary condition to establish internal structure validity, it has not been shown to be sufficient to analyze dimensionality of proposed constructs (Clark & Watson, 1995). Furthermore, internal consistency has no ability to investigate factor structure stability (Ferguson & Cox, 1993). Consequently, structural analysis of proposed scales conducted through exploratory factor analysis (EFA) are recommended as an appropriate next step to establish internal structure validity (Crocker & Algina, 1986). One of the necessary conditions to effectively analyze the factor structure of a proposed scale has been to ensure a sufficiently large number of responses. Specifically, a ratio of five respondents per item has been proposed (Ferguson & Cox, 1993).

After establishing the sufficiency of the respondent to item ratio, the Kaiser-Meyer-Olkin test of sampling adequacy and Bartlett's test of sphericity should be employed to ensure that observed factor structures are not found by chance. The Kaiser-Meyer-Olkin test observations should be evaluated according to established criteria; a minimum value of 0.5 might be considered acceptable with values above 0.8 considered very good (Dziuban & Shirkey, 1974). Additionally, results of the Bartlett's test of sphericity ($p < .05$) should be analyzed for significance as an indication of psychometric adequacy of the samples (Dziuban & Shirkey, 1974).

To investigate factor structure within the data, Kaiser's (1960) eigenvalue greater than 1 (K1) and Cattell's (1966) scree test are recommended (Hayton, Allen, & Scarpello, 2004). Based on the K1 analysis, the nature and dimensionality of the latent variable and associated factors should be analyzed. Confirmation of K1 analysis is done through an examination of a plot of eigenvalues known as a scree test (Cattell, 1966). Specifically, breaks or discontinuities within the eigenvalue plot will be used to identify stable factors followed by numerous smaller minor factors, "a few major factors account for the most variance, resulting in a steep 'cliff' as these factors are identified first, followed by a shallow 'scree' describing the small and relatively consistent variance accounted for by the numerous minor factors" (Hayton et al., 2004, p. 193). Additionally, a Promax rotation of the data is recommended to provide a validation of the dimensionality of the latent variable. Overall, the EFA analysis should identify the dimensions associated with a stable factor or factors associated with the latent variable. Based on the results of the EFA, the observed and hypothesized factor structure will be further analyzed through confirmatory factor analysis (CFA).

Confirmatory factor analysis has been established as an appropriate technique to validate observations against theoretical expectations (Henson & Roberts, 2006). From a CFA perspective, the hypothesized model should be analyzed using a statistically appropriate tool or software package capable of performing the necessary analysis. Model fit statistics should be calculated with recommended tests including: Chi-square test of model fit, root mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker Lewis Index (TLI), and standardized root mean square residual (SRMR).

As an initial starting point for CFA analysis the chi-square test of model fit has been proposed with non-significant observations indicating strong model fit (Bollen, 1989). However, more recently scholars have questioned the appropriateness of this measure, “it is well documented that the chi-square test is very sensitive to sample size, and thus, very small differences between the observed and reproduced covariance matrices will result in a statistically significant chi-square value” (Vanderberg, 2006, p. 197). Therefore, alternative measures of model hit have been proposed to account for the practical considerations associated with theory based model development and testing. According to Hu and Bentler (1998) several benchmarks have been established to analyze model fit statistics and thus identify model misspecification. Specifically, the following thresholds have been proposed: RMSEA values less than 0.08 represent acceptable model fit with values less than 0.06 representing good fit; CFI and TLI values of 0.90 represent marginal fit, with values below 0.90 indicating poor fit and values 0.95 representing good fit; SRMR values less than 0.08 represent acceptable model fit with values less than 0.05 representing good fit.

A criticism of CFA has been that, due to the specified nature of the model, researchers have the ability to manipulate variable relationships to fit observed data, thus improving model fit without an *a priori* grounding for such manipulations (MacCallum, Roznowski, & Necowitz, 1992). However, specifying relationships that are expected due to theoretical rationale has been deemed as appropriate (McDonald & Ho, 2002). A recommendation is for CFA observations and model fit to be treated and analyzed from a theoretical perspective.

External Structure Validity

An examination of the hypothesized relationships between the construct of interest and previously established constructs has been proposed to establish external structure validity (Schwab, 1980). Theoretically implied relationships between the construct and other measures within a similar domain context should be related, but not redundant. Typically, the use of established measures within a nomological network of similar concepts is used to validate external structure validity (Messick, 1989).

During initial content validity and scale development stages it is recommended to complete an exhaustive literature review where identification of similar scales should be noted. To establish external structure validity of an instrument it is recommendation to administer previously established instruments during the data collection process. Once collected, data analysis should include correlations between the proposed scale latent variables and existing scale latent variables. Magnitude of correlations (Davis, 1971) between the proposed scale and previously established scales should inform and establish external structure validity if resulting values meet the desired threshold.

Consequential Validity

According to Messick (1995), consequential validity “appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice” (p. 745). This source of validity has been shown to be paramount when there is the potential for score inferences to be associated with scale scores (Blanton & Jaccard, 2006). Specifically, scale scores should be meaningful and, thus, not arbitrary (Greenwald, Nosek, &

Sriram, 2006). This type of validity is particularly relevant in international agricultural and extension education contexts where scale results may be related to numerous needs such as evaluations or capacity assessments (Davis, 2016).

To establish consequential validity, the recommendations of Blanton and Jaccard (2006) and Messick (1995) are presented. First, a recommendation has been to employ a progressive approach whereby a proposed scale validity evolves from evidence for the construct interpretation, to evidence of the basis for score use, to evaluating value consequences of scores, to evaluating functional worth of scores (Messick, 1995). Therefore, one of the primary methods for collecting consequential validity evidence has been to coordinate with those that are responsible for implementing the proposed instrument, collecting, and interpreting results and building consensus amongst this audience accordingly (Blanton & Jaccard, 2006). Additionally, contextual relevance has been noted as an important condition for score meaningfulness and evidentiary support (Blanton & Jaccard, 2006; Messick, 1995).

Conclusions, Recommendations and Implications

Instrument development and validation is critical to data gathering and the subsequent analysis. Although there is agreement that reliable and valid data is the foundation for social science research, challenges associated with scale development should be acknowledged (Emmerson & Neely, 1988; McKnight et al., 2007). Validation of scales, either researcher developed or those adapted from previous utilization in other settings, is paramount to sound research practices and reliable data. Within international agricultural and extension education contexts, these needs are amplified when additional considerations such as audience, culture, and intended use must also be considered (Davis & Sulaiman, 2014).

Initially, researchers should focus on the ability of the scale to appropriately measure the underlying phenomena of interest, thus establishing content validity. Researchers should conduct an exhaustive literature review towards the construct of the scale. Experts should then review the scale to confirm the validity of its purpose and intent. This step can be accomplished in several ways such as a panel of experts, literature review, or a Delphi process (Crocker & Algina, 1986; DeVellis, 2017; Messick, 1995; Williamson, 2007). Both content and face validity should be evaluated (Crocker & Algina, 1986). If a Delphi process is employed (Lamm et al., 2017; Lamm et al., 2018), an iterative approach (Garson, 2014) should be utilized. Response process validity should also be ensured by prompting respondents to indicate whether they understand requested response(s), or if any confusion is associated with the instrument. Within international contexts, translation (Radhakrishna, 2006) and audience characteristics (Davis & Sulaiman, 2014) are important considerations.

Next, internal structure of the scale should be validated with descriptive analysis. Distributions from analysis should be analyzed for skewness and kurtosis to provide normalcy of the data. Following descriptive analysis, internal consistency analysis should calculate a Cronbach's alpha. Cronbach's alpha values greater than .70 are generally deemed acceptable. Subsequent analysis will examine the hypothesized variable structure through exploratory factor analysis followed by latent variable structure examination through confirmatory factor analysis (Clark & Watson, 1995). A minimum of five responses per item is advised (Ferguson & Cox, 1993). Establishing external structure validity is facilitated by collecting data from within the nomological network of conceptually related scale. Similarities without redundancy are an indication of external structure validity. Lastly, consequential validity ensures the usefulness of the latent variable information as represented by the scale result. From an international perspective, consequential validity is very important as scale results can have implications beyond research, such as evaluations or capacity assessments (Davis, 2016).

Scale development and validation are foundational to appropriate data gathering procedures and validity of results based on research questions and/or objectives of study in social sciences. Proper implementation of scale measures will allow researchers to gather pertinent and reliable data to draw conclusions and recommendations. Researchers should ensure that scales to be developed follow proper guidelines and procedures. A robust framework and set of actionable methodological recommendations should provide international agricultural and extension education researchers and practitioners a robust foundation upon which to construct valid and reliable scales and directly addresses identified needs within the literature (Lamm et al., 2018).

References

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley. https://doi.org/10.1002/9781118619179

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *Plos One, 6*(6), e20476. https://doi.org/10.1371/journal.pone.0020476

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. https://doi.org/10.1037/1040-3590.7.3.309

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104. https://doi.org/10.1037/0021-9010.78.1.98

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Mason, OH: Cengage Learning.

Czinkota, M. R., & Ronkainen, I. A. (1997). International business and trade in the next decade: Report from a Delphi study. *Journal of International Business Studies, 28*(4), 827-844. https://doi.org/10.1057/palgrave.jibs.8490121

Davis, J. A. (1971). *Elementary survey analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Davis, K., & Sulaiman, R.V. (2014). The “new extensionist”: Roles and capacities to strengthen extension and advisory services. *Journal of International Agricultural and Extension Education, 21*(3). 6-18. https://doi.org/10.5191/jiaee.2014.21301

Davis, K. (2016). How will extension contribute to the sustainable development goals? A global strategy and operational plan. *Journal of International Agricultural and Extension Education, 23*(1). 7-13. https://doi.org/10.5191/jiaee.2016.23101

Delbecq, A. L., Van de Ven, Andrew H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Glenview, IL: Scott, Foresman.

DeVellis, R.F. (2017). *Scale development: Theory and applications*. Los Angeles, CA: Sage Publications.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*(6), 358-361. https://doi.org/10.1037/h0036316

Emmerson, G. J., & Neely, M. A. (1988). Two adaptable, valid, and reliable data-collection measures: Goal attainment Scaling and the semantic differential. *The Counseling Psychologist. 16*(2), 261-271. https://doi.org/10.1177/0011000088162007

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users’ guide. *International Journal of Selection and Assessment, 1*(2), 84-94. https://doi.org/10.1111/j.1468-2389.1993.tb00092.x

Ganpat, W., Ramdwar, M., Stripling, C. T., & Roberts, T. G. (2013). Information and communication technologies use by agriscience teachers in Trinidad and Tobago. *Journal of International Agricultural and Extension Education 20*(2), 19-33.

https://doi.org/10.5191/jiaee.2013.20202

Garson, G. D. (2014). *The Delphi method in quantitative research.* Statistical Associated Publishing.

Gliddon, D. G. (2006). *Forecasting a competency model for innovation leaders using a modified Delphi technique* (Doctoral dissertation). Available from ProQuest Dissertations & Theses Full Text. (3292523).

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the implicit association test: Comment on Blanton and Jaccard (2006). *American Psychologist, 61*, 56–61. https://doi.org/10.1037/0003-066X.61.1.56

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205. https://doi.org/10.1177/1094428104263675

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement,10*(1), 141-151. https://doi.org/10.1177/001316446002000116

Lamm, K. W., Lamm, A. J., Davis, K., & Swaroop, B. J. (2017). Identifying knowledge management capacity needs of rural advisory service networks. *Journal of International Agricultural and Extension Education 24*(2), 93-106. https://doi.org/10.5191/jiaee.2017.24207

Lamm, K. W., Lamm, A. J., Davis, K., & Swaroop, B. J. (2018). Effective advocacy for extension networks: An evaluation of critical capacities. *Journal of International Agricultural and Extension Education 25*(2), 43-56. https://doi.org/10.5191/jiaee.2018.25204

Lordo, R. A. (2001) Learning from data: Concepts, theory, and methods, *Technometrics*, *43*(1), 105-106. https://doi.org/10.1198/tech.2001.s558

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504. https://doi.org/10.1037/0033-2909.111.3.490

McDonald, R. P., & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82. https://doi.org/10.1037/1082-989X.7.1.64

McMillan, J. H., & Schumaker, S. (2010). *Research in education: Evidence-based inquiry*. (7th ed.) Upper Saddle River, NJ: Pearson Education.

McKnight, P. E., McKnight, K. M., Sidani, S. & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. New York, NY: The Guildford Press.

Meijering, J. V., Kampen, J. K., & Tobi, H. (2013). Quantifying the development of agreement among experts in Delphi studies. *Technological Forecasting and Social Change, 80*(8), 1607-1614. https://doi.org/10.1016/j.techfore.2013.01.003

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13-103). New York, NY: Macmillan Publishing Company.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. https://doi.org/10.1037/0003-066X.50.9.741

Nunnally, J. C., & Bernstein, I. H. (1967). *Psychometric theory*. New York: McGraw-Hill.

Okoli, C., & Pawlowski, S. D. (2004). The delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15-29. https://doi.org/10.1016/j.im.2003.11.002

Radhakrishna, R. (2006). Teaching an extension program development course in an international setting. *Proceedings of Association of International Agricultural and Extension Education, 22*, 517-524.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach*. Thousand Oaks, CA: Sage.

Roy, T. K., & Garai, A. (2012). Intuitionistic fuzzy Delphi method: More realistic and interactive forecasting tool. *Notes on Intuitionistic Fuzzy Sets, 18*(2), 37-50.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350-353. https://doi.org/10.1037/1040-3590.8.4.350

Schwab, D. P. (1980). Construct validity in organizational behavior. In B. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (2nd ed., pp. 3-43). Greenwich, CT: JAI.

Skulmoski, G., Hartman, F., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research, 6*(1), 1-21. https://doi.org/10.28945/199

Stines, A. C. (2003). *Forecasting the competencies that will define "best-in-class" business-to-business market managers: An emergent Delphi-hybrid competency forecasting model* (Ph.D.). Available from ProQuest Dissertations & Theses Full Text. (305306481). Retrieved from http://search.proquest.com/docview/305306481?accountid=10920

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*(1), 99-103. https://doi.org/10.1207/S15327752JPA8001_18

Taras, V., Rowney, J., & Steel, P. (2009). Half a century of measuring culture: Review of approaches, challenges, and limitations based on the analysis of 121 instruments for quantifying culture, *Journal of International Management, (15)*4, 357-373. https://doi.org/10.1016/j.intman.2008.08.005

Vandenberg, R. J. (2006), Statistical and methodological myths and urban legends, *Organizational Research Methods, 9*(2), 194-201. https://doi.org/10.1177/1094428105285506

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56−75). Thousand Oaks, CA: Sage.

Williamson, S. N. (2007). Development of a self-rating scale of self-directed learning. *Nurse Researcher*, *14*(2), 66-83. https://doi.org/10.7748/nr2007.01.14.2.66.c6022