Evaluating Psychometric Properties to Advance Agricultural Education Scholarship

Robert Strong, *Texas A&M University*
979-845-1139
robert.strongjr@ag.tamu.edu

James R. Lindner, *Auburn University*
334-844-4434
jrl0039@auburn.edu

Clare Hancock, *Auburn University*
334-844-7475
cet0071@auburn.edu

Karissa Palmer, *Texas A&M University*
979-845-1139
karissa.palmer@ag.tamu.edu

Evaluating Psychometric Properties to Advance Agricultural Education Scholarship

Introduction/need for research

Instrumentation is a critical function in measuring social and behavioral science impacts on stakeholders, teachers, and change agents (Field, 2013). Inquiries on instrument quality offers researchers evidence of the extent measurement attributes were examined, and thereby, assisting the researcher select the best instrumentation tool to use (Dillman et al., 2014). Internal validity and reliability have long been considered the quality gatekeepers prior to collecting any social science data (Ary et al., 2019). Internal validity is as simply as the instrument measures what it is designed to measure (Field, 2013; Strong et al., 2022). Reliability is the internal consistency of the reproductive instrument results when utilized with a similar population (Cronbach, 1951). Reporting both validity and reliability in a study using instrumentation are cornerstones of disseminating social science research (Lindner et al., 2001). Warmbrod (2014) recommended agricultural education researchers document the empirical evidence proving an instrument's variables are both valid and reliable. Priority 2 of our *National Research Agenda* suggested examining research practices to better develop and implement agricultural teaching and learning processes for enhancing sustainable agricultural systems development (Lindner et al., 2016).

Theoretical Framework

The theory of planned behavior (Ajzen, 1991) framed this study by utilizing the constructs; subjective norms, attitudes, perceived behavioral control and resulting behavior. Subjective norms are beliefs individuals or groups will endorse and accomplish a specific behavior. Ajzen (1991) indicated attitudes are developed paradigms of thinking that result in one's behavior. Perceived behavioral control is the discernment of the complexity in carrying out a preferred behavior (Ajzen, 1991). Resulting behavior is predicted by one or each construct. The purpose of this study was to investigate data collection instrument development and reporting subjective norms and resulting behavior in agricultural education literature over the last three years. Specifically, research objectives were: 1) determine the number of articles reporting data collection instruments reporting construct reliability, and 2) determine the numerical range of statements or questions utilized to measure constructs.

Methodology

Wright et al. (2007) indicated a systematic review is a method using a comprehensive search based on explicit protocols to review existing literature with a synthesis of data focusing on key questions. Systematic reviews use five steps; identify the critical question, formulate search parameters, systematically search databases, analyze data, and lastly, summary and data interpretation (Lee et al., 2021). The authors systematically reviewed, using the five steps, all articles from *Advancements in Agricultural Development* (*AAD*), *Journal of Agricultural Education* (*JAE*), *Journal of Extension* (*JOE*), and *The Journal of Agricultural Education and Extension* (*TJAEE*) from 2020 to 2022 to answer the research objectives. Authors reviewed five hundred thirty-one (N = 531) articles from the four refereed publications.

Results/findings

JAE reported forty-seven (N = 47) articles in 2022 and seventeen (n = 17, 36.17%) reported data collection reliability coefficients. Of the seventeen articles, the numerical range of statements or questions utilized to measure constructs extended from 1 to 10. *JAE* reported in 2021 (N = 73)

published articles and thirty-two (*n* = 32, 43.83%) utilized data collection reliability coefficients. Of those thirty-two, the numerical range of statements or questions was 1 to 19. In 2020, *JAE* reported eighty-three (*N* = 83) articles published and forty-one (*n* = 41, 49.39%) tested reliability coefficients. Statements numerically ranged from 1 to 32. Thirty-two was the severe outliner.

Thirty-four (*N* = 34) articles were published in *JOE* in 2022 and five (*n* = 5, 14.70%) reported testing reliability coefficients. The range of statements or questions was 1 to 10. *JOE* had eighty-two (*N* = 82) articles published in 2021 and six (*n* = 6, 7.31%) reported data collection reliability coefficients. One to twelve was the range of statements or questions. In 2020, *JOE* reported sixty-eight (*N* = 68) articles and nine (*n* = 9, 13.23%) had data reliability coefficients. The numerical range of statements or questions utilized to measure constructs ranged from 1 to 7.

AAD had fourteen (*N* = 14) articles published in 2022 and seven (*n* = 7, 50%) articles reported data collection reliability coefficients. The numerical range of statements ranged from 1 to 7. Twenty-seven (*N* = 27) articles were published in 2021 and twelve (*n* = 12, 44.44%) utilized data collection reliability coefficients. The numerical range of statements was from 1 to 7. Twenty-three articles (*N* = 23) were published in 2020 and eight (*n* = 8, 34.78%) reported data collection reliability coefficients. The numerical range of statements or questions ranged from 1 to 5.

TJAEE had (*N* = 22) articles published in 2022 and three (*n* = 3, 13.63%) articles had construct reliability coefficients. Statements ranged from 1 to 36. Thirty-six was an outlier given the small number of instrumentation studies in 2022. In 2021, thirty-three (*N* = 33) articles were published and seven (*n* = 7, 21.21%) reported construct reliability coefficients. The numerical range of statements was 1 to 10. There were twenty-five (*N* = 25) articles and three (*n* = 3; 12%) articles tested construct reliability coefficients in 2020. Statements or questions ranged from 1 to 7.

Conclusions

Authors reported fewer construct items produced lower construct reliability coefficients and thus, producing the potential of higher levels of error (Cronbach, 1951). Results indicated the majority of our published scholarship has not utilized data collection instruments over the last three years. If the researchers who have, chose to implement smaller numbers of items to measure constructs.

Implications/recommendations/impact on profession

There are two competing and acceptable, in our professoriate, indicators of construct reliability. Likert's (1932) convention in his quintessential work on measuring social variables suggested that for measurements to be reliable an *alpha* of .9 should be achieved. While Cronbach's (1951) convention postulates that a construct reliability of .7 be achieved. It is easy to see why many would select Cronbach's convention in that the amount of time to increase reliability to the threshold suggested by Likert may be inhibiting to researchers. Besides, what difference does .2 make anyway? With a threshold of .7, a potential variance of up to 30% exists; subsequently with a threshold of .9, a potential variance of only up to 10% exists (Field, 2013). A difference of 20% variance can be a substantial difference in the power of analysis and interpretation of effect size (Ary et al., 2019). As a profession we should seek the highest level of reliability as possible, when possible. When developing an instrument, researchers should include a maximum number of statements and questions and eliminate those that do not contribute to reliability and add additional questions when acceptable levels of reliability are not achieved.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Ary, D., Jacobs, L. C., Sorensen Irvine, C. K., & Walker, D. A. (2019). *Introduction to research in education* (10th ed.). Cengage.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method* (4th ed.). John Wiley & Sons.

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics* (4th ed.). SAGE Publications.

Lee, C. L., Strong, R., & Dooley, K. E. (2021). Analyzing precision agriculture adoption across the globe: A systematic review of scholarship from 1999–2020. *Sustainability, 13*(18), 10295. https://doi.org/10.3390/su131810295

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Lindner, J. R., Rodriguez, M. T., Strong, R., Jones, D., & Layfield, D. (2016). Research priority area 2: New technologies, practices, and products adoption decisions. In Roberts, T. G., Harder, A., & Brashears, M. T. (Eds). *American Association for Agricultural Education national research agenda: 2016-2020*. Gainesville, FL: Department of Agricultural Education and Communication.

Lindner, J., Murphy, T., & Briers, G. (2001). The handling of nonresponse error in agricultural education. *Journal of Agricultural Education, 42*(4), 43–53. https://doi.org/10.5032/jae.2001.04043

Strong, R., Wynn, J. T., Lindner, J. R., & Palmer, K. (2022). Evaluating Brazilian agriculturalists' IoT smart agriculture adoption barriers: Understanding stakeholder salience prior to launching an innovation. *Sensors*, *22*(18), 6833. https://doi.org/10.3390/s22186833

Warmbrod, J. R. (2014). Reporting and interpreting scores derived from Likert-type scales. *Journal of Agricultural Education, 55*(5), 30–47. https://doi.org/10.5032/jae.2014.05030

Wright, R. W., Brand, R., Dunn, W., & Spindler, K. (2007). How to write a systematic review. *Clinical Orthopedics and Related Research, 455*, 23–29. https://doi.org/10.1097/BLO.0b013e31802c9098