Journal of Agricultural Education, 59(3), 185-197
https://doi.org/10.5032/jae.2018.03185

Instrument Validity in Manuscripts Published in the *Journal of Agricultural Education* between 2007 and 2016

Hailey R. Gates[1], Donald M. Johnson,[2] & Catherine W. Shoulders[3]

Abstract

We examined authors' treatment of instrument validity in a stratified random sample (n = 94) of quantitative studies published in the Journal of Agricultural Education from 2007 to 2016. A majority (78.7%) of studies reported use of either a researcher-developed instrument (41.5%) or an existing instrument modified by the researchers (37.2%). In 67.0% of articles, authors validated the instrument for the study reported; in 16.0%, the authors claimed instrument validity based on previous studies. Authors made no claim for instrument validity in 11.7% of articles and claimed an unspecified form of validity in another 11.7%. Among the 72 articles where specific validity claims were made, 67 (93.1%) claimed face and content validity, either alone or in combination with other forms of validity. Claims for content, concurrent, and discriminant validity, either alone or in combination with other forms of validity, were made in 11 (15.3%) of the 72 articles. Among 70 articles claiming face, content, or construct validity, 91.4% included a description of the validation panel; panelists were most often described as 'experts' (70.3%), although their area(s) of expertise were specified in only 29.7% of articles. We conclude with specific recommendations intended to shift the profession's subjective norms related to instrument validity.

Keywords: instrument, validity, manuscript, *Journal of Agricultural Education*

Introduction

The rigorous process of inquiry into social and behavioral questions via use of "systematic observation and measurement methods, standardized tests, sophisticated coding schemes for analyzing verbal and observational data, carefully constructed questionnaires, individual and group interview techniques, and large masses of data" (Krathwohl, 2009, p. 4) began only in the 20th century. While a relatively new field of research, progress within social science continues to accelerate. Growth within a knowledge base occurs through the creation of improved research methods and development of well-trained researchers (Krathwohl, 2009). Disciplinary faculty members are expected to contribute to that knowledge base through research and are evaluated on their ability to effectively do so (Knobloch, 2010).

When conducting research, faculty members contribute to the knowledge base through published works in peer-reviewed journals; within the agricultural education profession, the

[1] Hailey R. Gates is a Graduate Student in the Department of Agricultural Education, Communications and Technology at the University of Arkansas, E111A AFLS Building, Fayetteville, AR 72701. hrgates@uark.edu. 479-575-2035
[2] Donald M. Johnson is a Professor in the Department of Agricultural Education, Communications and Technology at the University of Arkansas, E111A AFLS Building, Fayetteville, AR 72701. dmjohnso@uark.edu. 479-575-2035
[3] Catherine W. Shoulders is an Associate Professor in the Department of Agricultural Education, Communications and Technology at the University of Arkansas, E111A AFLS Building, Fayetteville, AR 72701. cshoulde@uark.edu. 479-575-3799

premier peer-reviewed journal is the *Journal of Agricultural Education* (*JAE*) (Knobloch, 2010). The review process for manuscripts submitted to *JAE* includes an evaluation of research methods, requiring that reviewers critique with the following three questions in mind:

1. Are the essentials of methods/procedures reported?
2. Are the methods/procedures correct?
3. Are weaknesses in the methods/procedures accounted for and/or explained? (American Association for Agricultural Education, n.d.)

Reviewers are able to reject manuscripts based on this criteria, suggesting that researchers who are unable to carry out methodologically sound studies are less likely to publish or contribute to the knowledge base.

Many components contribute to the "correctness" of a study's methods. Among those is the quality of the instrument used to collect data; "the conclusions drawn and the recommendations made in such studies can be no better than the data on which they are based" (Huck, 2008, p. 75). In turn, the data collected can be no better than the instrument used to collect the data. According to Borg and Gall (1985), "the findings of research . . . can only be evaluated after the measurement tools that produced these findings have been carefully appraised" (p. 209). The two primary measures of instrument quality are reliability and validity (Huck, 2008).

Shoulders, Johnson, and Flowers (2015), VanLeeuwen (1997), VanLeeuwen, Dormody, and Seevers (1999), and to a greater degree, Warmbrod (2014) addressed the concept of instrument reliability within agricultural education manuscripts. Equally or even more important (Ary, Jacobs, & Sorenson, 2010), and yet to be the focus of quantitative evaluation by the agricultural education profession, is validity. Validity has traditionally been defined as the extent to which an instrument measures what it purports to measure; however, the current definition of validity (Ary et al., 2010) no longer focuses on the instrument itself but, rather, "on the validity of the interpretations or inferences that are drawn from the instrument's scores" (p. 225). Thus, an instrument is valid only to the extent it allows accurate interpretations or inferences about the specific individuals completing the instrument.

According to Warmbrod (2014), researchers should provide evidence "in journal articles documenting the validity of test scores, including a description of item-generating strategies to establish content validity, judgements of experts attesting face validity, and empirical evidence documenting criterion and construct validity" (p. 30). As has been recommended by Johnson and Shoulders (2017), Lindner, Murphy and Briers (2001), Miller and Smith (1983), Roberts, Barrick, Dooley, Kelsey, Raven, and Wingenbach (2011), and Warmbrod (2014), this study sought to examine the profession's treatment of instrument validity as a means of improving the quality of research published within *JAE*.

Theoretical Framework

This study described the use of instrument validity within manuscripts published in JAE between 2007 and 2016 using the theory of planned behavior as the guiding framework (Ajzen, 1991). The theory of planned behavior posits that an individual's attitude toward a behavior, his or her perceived control over the behavior, and the subjective norms regarding the behavior interact to influence the individual's intention to perform the behavior, which in turn influences the individual's actual behavior (see Figure 1) (Ajzen, 1991).

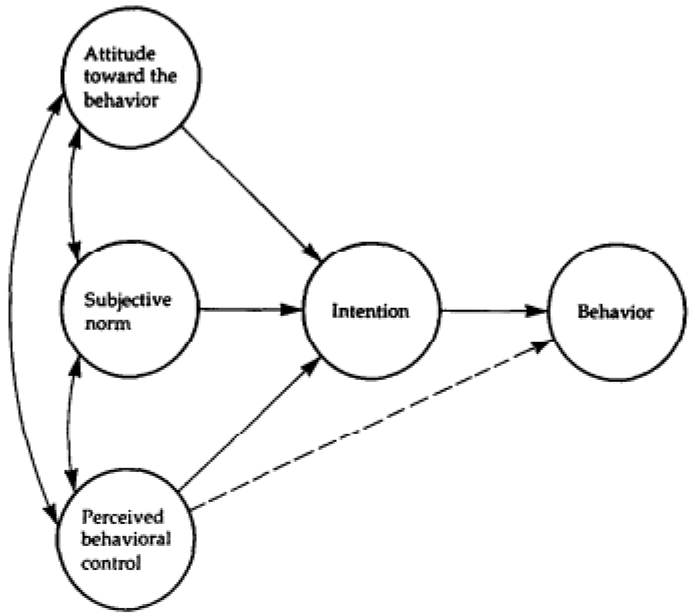

Figure 1. Theory of Planned Behavior (Ajzen, 1991).

Within this study, we assumed researchers publishing in *JAE* hold positive attitudes regarding instrument validity and the employment of methods used to establish instrument validity, as they continue to publish in a journal that includes validity as a criterion. Multiple publications within *JAE* have called for an examination of the use of validity within the journal's works, establishing a subjective norm that expects the employment of methods to establish validity of instruments within agricultural education research. By including within this manuscript a primer on establishing instrument validity (found within the Conceptual Framework), we aim to improve readers' perceived behavioral control by "reducing unfamiliar elements within the behavioral situation" (Johnson & Shoulders, 2017, p. 303). The theory of planned behavior states that successful efforts to improve individuals' perceived behavioral control can encourage individuals to engage in an intended behavior. Therefore, by bringing to light authors' practices as stated within *JAE* manuscripts and by providing a primer on methods to establish instrument validity, this manuscript enables researchers to examine, and potentially improve upon, their current behaviors regarding the establishment of instrument validity (see Figure 2).

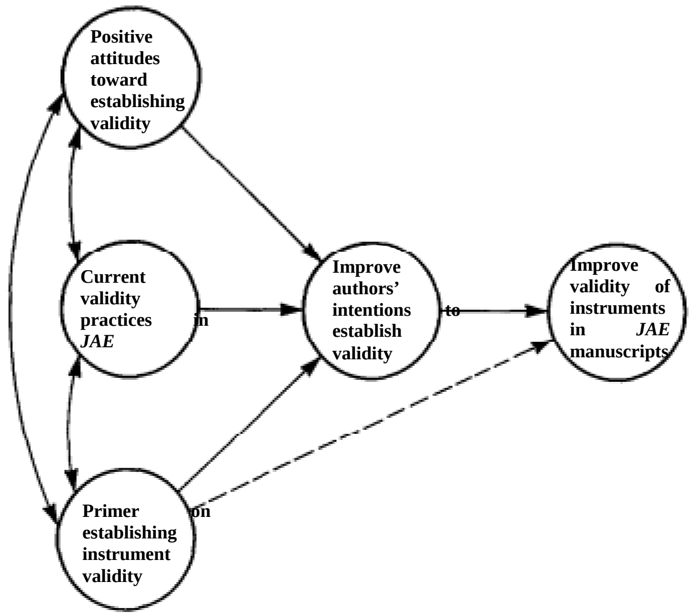

Figure 2. Adaptation of the Theory of Planned Behavior (Ajzen, 1991), displaying the design of the study.

Conceptual Framework

As is noted in the Theoretical Framework, altering one's perceived behavioral control by reducing the barriers that prevent him or her from carrying out an action can shift his or her intention to carry out a behavior (Ajzen, 1991). Lack of knowledge required to carry out a behavior is one such barrier that can be reduced (Johnson & Shoulders, 2017). Therefore, we offer a primer in instrument validity and methods to establish this criteria.

Defining Instrument Validity

Research in the social sciences often seeks to quantify attributes that cannot be measured directly (Kimberlin & Winterstein, 2008). Example attributes studied by agricultural education researchers include leadership, agricultural literacy, job satisfaction, burnout, creative thinking, and professionalism, among others (Radhakrishna & Wu, 1997). In studying an attribute, researchers must first develop a conceptual definition of the attribute and then translate this conceptual definition into an operational definition with associated instrument items and measurement scales (Kimberlin & Winterstein, 2008). According to Kimberlin and Winterstein (2008), "the ability to operationally define and quantify a construct is the core of measurement" (p. 278). Thus, fidelity between measurement items, the operational definition, and the original conceptual definition is an essential element of quality research (Trochim, 2006).

Instrument, or measurement, validity refers to the degree of accuracy achieved by the instruments used within a study (Huck, 2008), and the appropriateness of the inferences or decisions made from the results generated by the instrument (Ary et al., 2010; McMillan & Schumacher, 2010). While reliable data are consistent, it is possible for data to be simultaneously reliable and invalid – reliable data can consistently report something other than what the researcher intends to measure. The reverse, however, is not true; while reliable data is not necessarily valid, valid data must be reliable, as data cannot be accurate if not consistent. "Thus high reliability is a necessary but not sufficient condition for high validity" (Huck, 2008, p. 89). Instrument validity is a product of the specific situation and subjects in which the instrument is given; therefore, "in order to assure others that the procedures have validity in relation to the research problems, subjects, and setting of the study, it is incumbent on the investigator to describe validity in relation to the context in which data are collected" (McMillan & Schumacher, 2010, p. 173). Thus, instruments validated in different situations and with different subjects should not be assumed to be valid in different situations and with different subjects. According to Ary et al. (2010), "validity does not travel with the instrument. A test may be valid for use with one population or setting but not with another" (p. 225).

Establishing Instrument Validity

Researchers can establish one or more of several types of instrument validity, as "there are different ways in which scores can be accurate" (Huck, 2008, p. 89). While works detailing research methods organize the same types of validity in various ways, we adhered to Huck's (2008) conceptual framework of validity, which aligns methods of establishing validity with three overall types: content validity, criterion-related validity, and construct validity. Content validity refers to the degree to which the content intended to be measured is covered by the instrument's items (Huck, 2008; McMillan & Schumacher, 2010). Content validity is established via a panel of experts who examine the instrument to determine whether alignment exists between the instrument's items and the content on which the instrument is intended to assess subjects (Huck, 2008; McMillan & Schumacher, 2010). Experts should be selected based on their technical expertise, as they should be capable of making sound judgements regarding the content. "When reporting on efforts made to assess content validity, researchers should describe in detail who examined the content, what they were asked to do, and how their evaluative comments turned out" (Huck, 2008, p. 95). While content validity requires a detailed description of the content to serve as a checklist when evaluating an instrument, face validity is a less systematic method of establishing whether an instrument's items align with the intended content (McMillan & Schumacher, 2010; Trochim, 2006). Face validity relies on the subjective opinions of experts to determine whether the instrument, "on its face" (Trochim, 2006, para. 6) appears to measure the intended content. Because of the subjectivity required to establish face validity, it is considered by some researchers to be the weakest way to demonstrate validity (Lund Research, 2012; Trochim, 2006).

Criterion-related validity is established by comparing scores on an instrument with scores on an instrument known to accurately measure a relevant criterion variable (Huck, 2008). Relevancy of variables with which the comparison is made is crucial; "if the other variables are illogical or if the validity of the scores associated with such variables is low, then the computed validity coefficients conceivably could make a truly good instrument look as if it is defective" (Huck, 2008, p. 95). Correlating subjects' scores between the two instruments results in a validity coefficient; higher *r* values indicate higher validity. Criterion-related validity can be established via concurrent instrument administrations, wherein subjects are given both instruments within a short time frame, or via a predictive format, wherein the criterion is measured years before or after the instrument is administered (Huck, 2008; McMillan & Schumacher, 2010).

Construct validity establishes the degree to which an instrument is able to measure "how much of a personality or psychological construct is possessed by the examinees to whom the instrument is administered" (Huck, 2008, p. 92). Researchers establish construct validity by employing one or more of the following three actions:

1. provide correlational evidence showing that the construct has a strong relationship with certain measured variables *and* a weak relationship with other variables, with the strong

and weak relationships conceptually tied to the new instrument's construct in a logical manner;

2. show that certain groups obtain higher mean scores on the new instrument than other groups, with the high- or low-scoring groups being determined on logical grounds *prior to* the administration of the new instrument; or

3. conduct a factor analysis on scores from the new instrument (Huck, 2008, p. 92).

When providing correlational evidence, strong relationships establish convergent validity, while weak relationships establish divergent, or discriminant, validity.

The multitrait-multimethod matrix (MTMM) method can also be used to establish construct validity in studies examining two or more constructs. MTMM uses correlations between multiple measures of each construct to examine construct validity through simultaneous consideration of the resulting matrix of convergent and divergent validity coefficients (Trochim, 2006).

Purpose and Objectives

The purpose of this study was to describe authors' practices in addressing instrument validity for selected articles published in the *Journal of Agricultural Education* from 2007 to 2016. Specific objectives were to:

1. Describe the primary purpose of research reported in selected articles published in *JAE* from 2007 to 2016;

2. Describe the nature of the dependent variables studied in selected articles published in *JAE* from 2007 to 2016;

3. Describe the source of research instruments used in selected articles published in *JAE* from 2007 to 2016;

4. Describe the nature of instrument validity claims made in selected articles published in *JAE* from 2007 to 2016; and

5. Describe the specific types of validity claimed in selected articles published in *JAE* from 2007 to 2016.

Methodology

This study utilized a content analysis strategy to make inferences from communication observed using replicable and reliable methods (Krippendorff, 1980). The articles identified for this study were observed from 559 manuscripts published by *JAE* between 2007 and 2016. One hundred thirty-five articles were randomly selected through stratified random sampling by volume and issue to ensure equal representation throughout the years. Ninety-four of the selected articles were deemed to be quantitative in nature; 41 articles were removed from the originally selected 135 because they conducted qualitative inquiry, for which instrument validity does not apply.

Researchers developed and employed a standardized coding sheet to document the following information from each article: (a) purpose of research, (b) whether the instrument used existed previously or was researcher-developed, (c) whether instrument validity was addressed by researchers, (d) whether the instrument was validated for the study, (e) type(s) of validity addressed, (f) number of people on panel if face or content validity was used, (g) description of people on panel if face or content validity was used, (h) correlations presented if criterion-related validity was used, and (i) in-text citations used to support validity methods. The content coded for this study was considered to be manifest content. According to Potter and Levine-Donnerstein (1999), manifest content "is that [content] which is on the surface" (p. 259) and is easily observable. In coding manifest content, validity is achieved by defining content with "binary rules based on definitions" (p. 261). The coding sheet used in this study contained specific binary definitions and rules, based on the validity literature (Ary et al., 2010; Huck, 2008) that guided coding decisions.

Identification numbers were assigned to each article and the article's title page, purpose and objectives, and methods sections were printed and keyed to the identification number on the coding sheet to allow for data verification. All coding was completed by one researcher. A second researcher coded a randomly selected sample of 10 manuscripts (10.1%), with an overall agreement percentage of 88.6%. The resulting Cohen's *kappa* of .77 indicated "substantial" ($\geq$ .61) and approaching "almost perfect" (*kappa* $\geq$ .81) agreement between raters (Viera & Garrett, 2005). Data were analyzed using frequencies and percentages.

Results

Of the 559 articles published in *JAE* from 2007 to 2016, 135 (24.2%) articles were randomly selected to be included in the sample; of these selected articles, 94 (69.6%) articles met the criteria for inclusion and were analyzed as part of this study. The primary objective of the vast majority (90.1%) of all articles was to describe characteristics or phenomena. Relatively few articles sought to predict or to establish cause and effect relationships (see Table 1).

Table 1

Primary Objective(s) of Selected Articles Published in the Journal of Agricultural Education, 2007-2016

Purpose(s) of study	*f*	%
Describe	89	90.1
Predict	8	8.2
Establish cause and effect	1	1.0
Total	98	100.0

Note. Four of the 94 (4.2%) articles had primary objectives in two categories.

Approximately 86% of articles studied dependent variables measuring achievement, attitudes or perceptions, and/or behavior(s) (see Table 2). The remaining studies reported dependent variables categorized as 'other' and included measurements of adequacy, relationships, and skills. Over one-half of all articles studied variables in two or more categories.

Table 2

Nature of Dependent Variables in Selected Articles Published in the Journal of Agricultural Education, 2007-2016

Nature of dependent variable(s)	*f*	%
Achievement	42	31.3
Attitudes or perceptions	38	28.4
Behavior(s)	35	26.1
Other	19	14.2
Total	134	100.0

Note. Fifty-five of the 94 (58.5%) articles had dependent variables in two or more categories.

The largest single percentage (41.5%) of articles reported use of researcher-developed instruments; however, the combined categories of existing instruments and existing instruments modified by the researchers constituted over one-half (56.4%) of all reported instruments. Two studies (2.1%) reported use of both modified and researcher-developed instruments (see Table 3).

Table 3

Source of Research Instruments Used in Selected Articles Published in the Journal of Agricultural Education, 2007-1016

Instrument source	*f*	%
Existing - no modifications	18	19.2
Existing - modified by researcher(s)	35	37.2
Researcher-developed	39	41.5
Existing - modified by researcher(s) and researcher-developed	2	2.1
Total	94	100.0

Instrument validity was addressed in some manner by authors in 83 of the 94 (88.3%) selected articles published in *JAE* between 2007 and 2016 (see Table 4). Of those 83 articles, over three-fourths (75.9%) reported the instrument(s) were validated for the current study, while 18.1% reported validation based on previous research and the remaining 6% reported validation based on both current and previous research.

Table 4

Nature of Validity Claims Made in Selected Articles Published in the Journal of Agricultural Education, 2007-1016

Nature of validity claim	*f*	%
None made	11	11.7
Based on validation from previous research	15	16.0
Based on validation for current research	63	67.0
Based on both validation from previous research and validation for current research	5	5.3
Total	94	100.0

As previously indicated 11.7% of articles made no validity claims while an additional 11.7% claimed some unspecified form of validity, generally using wording similar to, 'the instrument was examined by a panel of experts and judged to be valid." A combination of face and content validity was claimed in 42 (58.3%) of the 72 articles where specific validity claims were made. In addition, single claims for either face (11.1%) or content (15.3%) validity were made in slightly over one-fourth of the 72 articles. Claims for face and content validity, either alone or in combination with other forms of validity, were claimed in 67 (93.1%) of the 72 articles. Claims for content, concurrent, and discriminant validity, either alone or in combination with other forms of validity, were made in 11 (15.3%) of the 72 articles.

Table 5

Type(s) of Validity Claimed in Selected Articles Published in the Journal of Agricultural Education, 2007-2016

Type(s) of validity claimed	*f*	%
None	11	11.7
Not specified	11	11.7
Face	8	8.5
Content	11	11.7
Construct	2	2.1
Discriminant	2	2.1
Face and content	42	44.7
Content and construct	4	4.3
Face, content and construct	2	2.1
Construct, concurrent and discriminant	1	1.1
Total	94	100.0

Among the 70 studies reporting some combination of face, content, or construct validity, 64 (91.4%) described the composition of the instrument validation panel. Panel members were most frequently described as ‘experts’ (*f* = 45, 70.3%), although the specific area of expertise was specified in only 22 cases (29.7%). The most frequently specified area of expertise was agricultural and extension education (*f* = 10, 45.5%). Panelists were described as ‘faculty experts’ or ‘expert faculty’ in 11 of the 64 (17.2%) studies with no additional information provided. The number of individuals serving on the validation panel was reported for 27 (38.6%) of 70 studies and ranged from two to 13, with a mean of 5.22 (*SD* = 2.59) members.

Conclusions/Implications/Recommendations

Previous scholars have examined the methods agricultural education researchers use to review the literature (Kildow, Zimmerman, Shoulders, & Johnson, 2014; Swafford & Anderson, 2007), develop theoretical frameworks (Camp, 2001), control for nonresponse error (Johnson & Shoulders, 2017; Linder et al., 2001; Miller & Smith, 1983), analyze and report results based on Likert-type scales (Warmbrod, 2014), describe effect sizes (Kotrlik, Williams, & Jabor, 2011), and select papers for presentation at scholarly conferences (Shoulders et al., 2015). The purpose of these efforts has been to improve the quality and impact of the research conducted and reported by the profession. As indicated by the number of manuscripts published on the topic, the agricultural education profession holds value in examining the quality of its research, and is dedicated to data-driven improvement of that research. To date, instrument validity has remained an unexamined aspect of the profession’s research quality, yet holds great influence on the ability of researchers to report accurate data and make appropriate generalizations (Ary et al., 2010; McMillan & Schumacher, 2010). To close this gap, the current study addresses the profession’s approach to instrument validity in a continued effort to improve research in agricultural education.

Consistent with Fuhrman and Ladewig (2008), a majority (90.1%) of the quantitative studies examined were primarily descriptive in nature. The primary dependent variables studied were achievement (31.3%), attitudes or perceptions (28.4%), and behaviors (26.1%). The use of existing instruments, which is commonly used to ensure accuracy (Huck, 2008), was seen most frequently (56.4%) with and without modifications. Despite the criteria established for acceptance in *JAE*, 22 (23.4%) articles reviewed for this study either made no claim toward the nature of their instruments’ validity, or did not specify what type of validity was employed. These findings suggest researchers did not attempt to establish instrument validity, and imply that *JAE* reviewers either failed to take note of the omission of instrument validity, or deemed the omission to be inconsequential to the accuracy of the manuscripts’ findings and conclusions. Huck (2008) identified a lack of validity as a threat to the overall correctness of the study itself. Further, an additional 16% of the articles claimed instrument validity via the validity established and cited in previously published research. However, according to Ary et al. (2010), “validity does not travel with the instrument” (p. 25). McMillan and Schumacher (2010) stated that instrument validity should be established within the context in which data are collected in order to ensure the instrument yields valid interpretations or inferences about the specific group completing the instrument. The high value placed on quality research within *JAE* is not represented through these practices; therefore, we recommend researchers take efforts to establish and thoroughly report test validity within manuscripts. Additionally, we recommend that reviewers for *JAE* evaluate studies attempting to establish validity via previous studies accordingly, and consider offering feedback to authors that encourages them to establish validity within their own studies and adequately report the process and subsequent results.

The use of face validity, which has been considered by researchers to be the weakest measurement of validity due to its subjective nature (Lund Research, 2012; Trochim, 2006), was reported by over half (53.2%) of the 72 articles. The use of an instrument validation panel was described in the majority (91.4%) of articles reporting face, content, and construct validity. Huck (2008) recommended the use of an instrument validation panel be accompanied with a description of who the experts are, what they do, and how their comments were evaluated. Panel members were most frequently described as 'experts' (70.3%), with information pertaining to what they do or who they are limited to 'faculty experts' or 'expert faculty' (17.2%). Fewer than half (45.5%) indicated these panel members were employed within the field of agricultural and extension education, while the remaining articles did not offer further information regarding the fields in which these experts worked. We recommend manuscript space be devoted by researchers to provide thorough descriptions of the merits of members of expert panels responsible for establishing face, content, and construct validity.

Fewer than 10% of the articles included efforts to establish concurrent or discriminant validity, considered to be most rigorous among validity methods (Huck, 2008). We posit this low rate of inclusion may be attributed to researchers' lack of awareness or confidence in performing these tests. Alternately, the low rate of inclusion of these types of validity may be a reflection of authors' understanding and acceptance of the subjective norms related to validity within the agricultural education profession. Therefore, adhering to the theory of planned behavior, we recommend the findings of this study be disseminated with researchers in order to stimulate conversation related to expectations of establishing instrument validity. Through these discussions, we can shift subjective norms of describing phenomena based on face-validated instruments, thereby stimulating progress within agricultural education through quality research (Krathwohl, 2009).

References

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

American Association for Agricultural Education. (n.d.). JAE review process. Retrieved from http://jae-online.org/index.php/review-process.

Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to Research in Education* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Borg, W. R, & Gall, M. D. (1985). *Educational Research: An Introduction* (4th ed.). New York: Longman.

Camp, W. G. (2001). Formulating and evaluating theoretical frameworks for career and technical education research. *Journal of Vocational Education Research*, *26*(1), 4-25.

Fuhrman, N. E., & Ladewig, H. (2008). Characteristics of animals used in zoo interpretation: A synthesis of research. *Journal of Interpretation Research, 13*(2), 31-42.

Huck, S. W. (2008). *Reading statistics and research.* (5th ed.). Boston, MA: Pearson Education.

Johnson, D. M., & Shoulders, C. W. (2017). Power of statistical tests used to address nonresponse error in the Journal of Agricultural Education. *Journal of Agricultural Education, 58*(1), 300-312. doi:10.5032/jae.2017.01300

Kildow, H., Zimmerman, B., Shoulders, C. L., & Johnson, D. M. (2014). Use of citations within manuscripts published by the Journal of Agricultural Education. *Journal of Agricultural Education, 55*(5), 78-92. doi:10.5032/jae.2014.05078

Kimberlin, C. L., & Winterstein, A.G. (2008). Validity and reliability of measurement instruments used in research. *American Journal of Health System Pharmacists, 65*: 2276-2284. doi:10.2146/ajhp070364

Knobloch, N. A. (2010). Becoming a scholar as a teacher educator. In R. M. Torres, T. Kitchell, & A. L. Ball (Eds.), *Preparing and advancing teachers in agricultural education* (pp. 42-85). Columbus, OH: The Ohio State University.

Kotrlik, J. W., Williams, H. A., & Jabor, M. K. (2011). Reporting and interpreting effect size in quantitative agricultural education research. *Journal of Agricultural Education, 52*(1), 132-142. doi:10.5032/jae.2011.01132

Krathwohl, D. R. (2009). *Methods of educational and Social Science Research.* (3rd ed.). Long Grove, IL: Waveland Press.

Krippendorff, K. (1980). Validity in content analysis. In E. Mochmann (Ed.), *Computer strategies for communication analysis* (pp. 69-112). Frankfurt, Germany: Campus.

Lindner, J. R., Murphy, T. H., & Briers, G. E. (2001). Handling nonresponse in social science research. *Journal of Agricultural Education, 42*(4), 43-53. doi:10.5032/jae.2001.04043

Lund Research. (2012). Face validity. Retrieved from http://dissertation.laerd.com/facevalidity.php

McMillan, J. H., & Schumacher, S. (2010). *Research in Education.* (7th ed.). Upper Saddle River, NJ: Pearson Education.

Miller, L. E., & Smith, K. L. (1983). Handling nonresponse issues. *Journal of Extension*, *21*(5), 45-50.

Potter, W. J., & Levine-Donnerstein, L. (1999) Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research*, *27*(3), 258-284, doi:10.1080/00909889909365539

Radhakrishna, R. B., & Xu, W. (1997). A review of subject matter topics researched in agricultural and extension education. *Journal of Agricultural Education, 38*(3), 59-69. doi: 10.5032/jae.1997.03059

Roberts, T. G., Barrick, R. K., Dooley, K. E., Kelsey, K. D., Raven, M. R., & Wingenbach, G. J. (2011). Enhancing the quality of manuscripts submitted to the *Journal of Agricultural Education*: Perceptions of experienced reviewers. *Journal of Agricultural Education, 52*(3), 1-5. doi: 10.5032/jae.2011.03001

Shoulders, C. W., Johnson, D. M., and Flowers, J. (2015). Peer review in agricultural education: Interrater reliability of manuscript reviews for the 2014 National Agricultural Education Research Conference. *Journal of Agricultural Education, 56*(3), 121-133. doi:10.5032/jae.2015.03121

Swafford, M., & Anderson, R. (2007). A two decade (1981-2000) analysis of the literature cited in the Journal of Agricultural Education. *Journal of Agricultural Education,* 48(3), 37-45. doi:10.5032/jae.2007.03037

Trochim, W. M. (2006). Research methods knowledge base. Retrieved from https://www.socialresearchmethods.net/kb/measval.php

VanLeeuwen, D. M. (1997). Assessing reliability of measurements with generalizability theory: An application to inter-rater reliability. *Journal of Agricultural Education, 38*(3), 36-42. doi: 10.5032/jae.1997.03036

VanLeeuwen, D. M., Dormody, T. J., & Seevers, B. S. (1999). Assessing the reliability of student evaluations of teaching (sets) with generalizability theory. *Journal of Agricultural Education, 40*(4), 1-9. doi: 10.5032/jae.1999.04001

Viera, A. J., & Garrett, J. M. Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360-363.

Warmbrod, R. J. (2014). Reporting and interpreting scores derived from Likert-type scales. *Journal of Agricultural Education, 55*(5), 30-47. doi:0.5032/jae.2014.05030