Research Poster

An Examination of Delphi Consensus Levels Over 30 Years in the Journal of Agricultural Education

Keith J. Frost
Texas Tech University
Department of Agricultural Education and Communications
Box 42131
Lubbock, TX 79409-2131
(806)834-6526
keith.frost@ttu.edu

Dr. John Rayfield
Texas Tech University
Department of Agricultural Education and Communications
Box 42131
Lubbock, TX 79409-2131
(806)834-1956
john.rayfield@ttu.edu

An Examination of Delphi Consensus Levels Over 30 Years in the Journal of Agricultural Education

Introduction/Purpose

The Delphi method of research was developed by the Rand corporation for military use (Dalkey & Helmer, 1963) and has been described as a "method for structuring a group communication process so the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem" (Linstone & Turoff, 2002, p. 3). Since its military roots, the method has evolved into what some describe as a series of techniques (Rowe & Wright, 2011) used in business, program planning, nursing, and the social sciences (Hsu & Sanford, 2007).

Regardless of how it is described or used, Delphi research shares commonality in using anonymous response, controlled iterative feedback, and statistical group response aggregation to achieve consensus on a topic (Dalkey, 1969; Lintsone & Turoff, 2002; Hsu & Sanford, 2007). Merriam-Webster's online dictionary defines consensus as a general agreement, a judgement arrived at by most concerned, or unanimity (Merriam-Webster, 2019). Von der Gracht (2013) described consensus as a key component of Delphi study and noted a general lack of consistency in the studies he reviewed in the field of technological forecasting. A review of the literature published in the *Journal of Agricultural Education* showed numerous studies utilizing the Delphi method but no such review on the methods used to determine consensus in the field. This preliminary study on the topic aligns with the purpose of the American Association for Agricultural Education (AAAE) and the philosophy of the *Journal of Agricultural Education* (AAAE, 2017; AAAE, n.d.) The research question driving this study is: how has consensus been defined or calculated in articles published in the *Journal of Agricultural Education?*

Theoretical Framework.

Ajzen's Theory of Planned Behavior suggests that an individual's decision-making process and behavioral intentions are shaped by personal beliefs about the activity (including self-efficacy beliefs) and subjective normative attitudes from outside sources coupled with perceptions of behavioral or outcome control (Ajzen, 1991; 2006). In using this model, the researchers assume the published authors already hold positive beliefs regarding behavioral control and self-efficacy in using the Delphi methods. The evaluation of research practice aligns with the philosophy of the *Journal of Agricultural Education* that includes a focus on "theoretical considerations pertaining to research" (AAAE, n.d.). Further, the researchers posit an understanding of current practice and trends in the application of Delphi methods can lead to reflection with a potential positive shift in personal beliefs and/or normative attitudes towards the method in a way that will improve practice.

Methods

To assess the research question, summative content analysis procedures were used (Hsieh & Shannon, 2005). The search function of the online portal for the *Journal of Agricultural Education* was the only source used to find articles. A search query was run using the keyword "Delphi". The search results were reviewed for date of publish and any articles outside of the thirty-year window (1989-2019) were excluded. The remaining articles (n = 47) were downloaded, printed and given an initial read-through to confirm the use of Delphi or Modified Delphi methods. One article was found to be an article about the use of Delphi methods in

agricultural education and eliminated from further review. The research question was set *a priori* and a researcher generated evaluation form was developed with the following sections: article number, scale(s) utilized, and consensus level. When it was noted that some studies used different scales in different rounds of the study, columns were added to the original form. Documents were analyzed for content using handwritten information and, when complete, recorded data were entered into Microsoft Excel for organization and sorting.

Findings

Although this portion of a larger study is focused on consensus levels, it is necessary to include the different scales used as a grouping method. While most (n = 43) studies indicated using Likert-type scales to measure agreement level after the generative round, three studies did not. Of those, one asked the panel to rank the importance of generated topics, one indicated using content analysis methods, and the third also utilized content analysis methods, but indicated consensus had been met if 25% of the responses included the word or phrase. Six studies (13%) used a scale where 1=Strongest Disagreement and 4= Strongest Agreement. The majority (n = 19, 41.3%) used a similar scale (1-5) but added a neutral response (3) making five the strongest level of agreement. Eight studies (17.4%) used a 1-6 scale where one and six were the strongest levels of disagreement and agreement (respectively) and where three and four indicated slight dis/agreement. An additional three studies (6.7%) added a neutral response to create a 1-7 scale where 4 = neutral. Four studies (8.7%) used four different scales (1-5, 1-6, 0-6, 1-7, 1-9) where the lowest number indicated the lowest importance and the largest number the greatest importance. The remaining study indicated using a "Likert-type scale" and indicated consensus at 60% but did not provide the scale.

Of the six studies that used the 1-4 scale, one did not provide consensus level, one reported frequencies and percentages, one used a threshold to qualify for round three for consensus determination, and the remaining three used percent agreement levels at 3+, all using different thresholds (60, 75, 80%). Most studies (n = 19) indicated using a 1-5 scale. One did not provide consensus threshold, one reported means and standard deviations, five used four different thresholds for qualifying for round three consensus with five different consensus levels, the remaining 12 studies used eight different definitions of consensus. Studies using a 1-6 scale (n = *8)* were the most consistent with four different consensus definitions and one study using a qualifying level for round three consensus. Each of the studies using a 1-7 scale (n = 3) or a scale for determining importance (n = 4) used different levels to determine consensus.

Conclusions, Implications, and Recommendations

In the last 30 years, the *Journal of Agricultural Education* published 46 studies that used a version of the Delphi method. Of those, 12 did not define or use the idea of consensus to report findings. The remaining 38 studies used five different scales of measurement and 29 different operationalized definitions of consensus. This inconsistency could lead to a lack of clarity or understanding for those that read the studies and, perhaps diminished credibility in the mind of the reader regarding the individual studies or the Delphi techniques in general. Further, more detailed study of Delphi use in our profession is recommended. Additionally, it is recommended that the AAAE host a round-table discussion, or a professional development seminar on the Delphi method, its usage, and best practices for scale and consensus selection.

References

AAAE (2017). *Constitution and Bylaws of the American Association for Agricultural Education.* Retrieved from: http://aaaeonline.org/resources/Documents/National/AAAE%20National%20Constitution%20By%20Laws%20-%20May%202017%20-%20Final.pdf

AAAE (n.d.). *JAE Philosophy.* Retrieved from: http://jae-online.org/index.php/philosophy

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*, 179-211

Ajzen, I. (2006). *Behavioral Interventions Based on the Theory of Planned Behavior*. Retrieved from https://people.umass.edu/aizen/pdf/tpb.intervention.pdf

Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458-467.

Dalkey, N. C. (1969). *The Delphi method: An experimental study of group opinion.* Santa Monica, CA: The Rand Corporation

Hsu, C. C., & Sanford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research & Evaluation, 12*(10), 1-8.

Hsieh, H., & Shannon, S. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277-1288. Doi: 10.1177/1049732305276687

Linstone, H. A., & Turoff, M. (Eds.) (2002). *The Delphi method: Techniques and applications.* Retrieved from https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf

Merriam-Webster (2019). *Consensus*. Retrieved from: https://www.merriam-webster.com/dictionary/consensus

Rowe, G., & Wright, G. (2011). The Delphi technique: Pas, present, and future prospects – Introduction to the special issue. *Technological Forecasting & Social Change 78*(2011) 1487-1490. Doi: 10.1016/j.techfore.2011.09.002

Von der Gracht, H. (2013). Consensus measurement in Delphi studies: Review and implications for future quality assurance. *Technological Forecasting & Social Change, 79* (4), 1525-1536. Doi: 10.1016/j.techfore.2012.04.013