Advancements in Agricultural Development
Volume 5, Issue 2, 2024
agdevresearch.org

# Interpreting Likert-type Scales, Summated Scales, Unidimensional Scales, and Attitudinal Scales: I neither Agree nor Disagree, Likert or Not

J. R. Lindner[1], N. J. Lindner[2]

## Abstract

This paper provides a rationale and convention for discussing the true limits and interpretation of data collected using unidimensional, summated, Likert-type, and attitudinal scales used in research investigating human behavior, sociology, education, psychology, and other related fields of study. All vague quantifiers must be described in methods and findings. The true limits of the scale and of each vague quantifier should be described. This information should be placed in the methods section. A five-point summated scale, for example, can be interpreted as follows: *Strongly Agree* = 5 - 4.51, *Agree* = 4.5 - 3.51, *Neither Agree nor Disagree* = 3.5 - 2.51, *Disagree* = 2.5 - 1.51, *Strongly Disagree* = 1.5 - 1. This paper also provides a rationale and convention for the use of nonstandardized effect size (*ES*) estimates to describe the magnitude and strength of the effect. This is accomplished by subtracting one summated *M* from another summated *M* and interpreted using the following convention: Small (*ES* = .19 and lower); Medium (*ES* = .20 - .49); and Large (*ES* = .50 and higher). The rationale for this is based on the intuitiveness of the measure, true limits of the scale, and scale intervals.

### Article History
Received: May 31, 2023
Accepted: July 31, 2023
Published: January 31, 2024

### Keywords
human behavioral research; behavioral scaling; educational research; psychological research; true limits; effect size; quality education

1. James R. Lindner, Professor, Auburn University, 5040 Haley Center, Auburn University, AL, 36849, jrl0039@auburn.edu, https://orcid.org/0000-0002-1448-3846
2. Nicholas J. Lindner, Research Technician III, Duke University, Viral Vector Core, Duke University School of Medicine, Durham, NC, 27716, nicholas.lindner@duke.edu, https://orcid.org/0000-0002-7140-6820

## Introduction and Problem Statement

The use of *unidimensional*, summated, Likert-type, or attitudinal scales are ubiquitous in research related to human behavior, sociology, education, psychology, and the field of agricultural leadership, education, and communications. Interpreting summated scores has mostly focused on the appropriateness of using single items versus summated scales. In a meta-analysis of the research reported in the *Journal of Agricultural Education*, Warmbrod (2014) found that the use of summated scores resulted in higher congruent interpretations than single items scores. In *A technique for the Measurement of Attitude*, Likert's (1932) work focused on the use of summated scales to measure attitudes, not single items. Because of the impact of Likert's (1932) work, summated scales are often referred to as Likert-type. The treatment of Likert-type scores as ordinal or interval has also been widely debated (Norman, 2010; Sullivan & Arino, 2013). Clason and Dormody (1994) noted the commonality of treating Likert-type scores as interval level data in agricultural education research. They write that it is "difficult to see how normally distributed data can arise in a single Likert-type item" (p. 34). Notwithstanding the criticism of how data are often treated, Clason and Dormody (1994) suggest that the focus of analysis should be on answering relevant research questions. The authors of this paper concur. Others have found that treating single item Likert-type scores as interval scores changes research conclusions (Gardner et al., 1998; Sisson & Stocker, 1989). The debate over the appropriateness of treating single or summated Likert-type scores as interval scores will continue and is not addressed in this paper. Warmbrod (2014) reminded us, however, that the use of the term "Likert-type" scale was synonymous with summated scale. Single items should be referred to appropriately; e.g., attitudinal scale, rating scale. Miller (1998) enunciated the appropriate analysis for the reporting of ordinal and interval and provides justification for such. Collectively Miller's (1998) and Warmbrod's (2014) advice is as sound today as when they first reported it. Building on their work, interpretation of scores may help us better understand and communicate findings.

## Purpose

The primary purpose of this paper was to provide a rationale and convention for discussing the true limits and interpretation of data collected using unidimensional, summated, Likert-type, and attitudinal scales used in research investigating human behavior, sociology, education, psychology, and other related fields of study. A secondary purpose was to provide a rationale and acceptable convention for the use of nonstandardized effect size (ES) estimates.

## Discussion and Commentary

Accurately interpreting scores, regardless of whether they are obtained from single items or summated Likert-type scores that are reported as means (*M*) and standard deviations (*SD*), is critical. A very common summated scale would include five terms to describe a research participant's level of agreement to a particular statement: Strongly Agree, Agree, Neither Agree nor Disagree, Disagree, and Strongly Disagree. These terms are referred to by Dillman et al.

(2014) as vague quantifiers. They further explained "[t]he problem with vague quantifiers is that they are vague; that is, there is not a single and clear meaning for each of the labels in the same way there is with natural metrics" (p. 151).

When one individual responds to a single question, the response is directly tied to a specific vague quantifier (see Table 1).

**Table 1**

*Examples of How Single Items Questionnaires May be Developed*

| Statements | Vague Quantifiers and Associated Values | | | | |
|---|---|---|---|---|---|
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research. | Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research. | SA | A | NAD | D | SD |
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research. | 1 | 2 | 3 | 4 | 5 |
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research. | 5 | 4 | 3 | 2 | 1 |
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research. | ○ | ○ | ○ | ○ | ○ |

In the examples provided in this paper the following conventions were used: *Strongly Agree* (SA) = 5, *Agree* (A) = 4, *Neither Agree nor Disagree* (NAD) = 3, *Disagree* (D) = 2, and *Strongly Disagree* (SD) = 1. Whether the data are ordinal or interval is irrelevant. For example, individual one indicates that they *Agree* that "*Advancements in Agricultural Development* (*AAD*) has a positive influence on agricultural development practices worldwide through the rapid publication of research." In describing the individual's response, researchers should report that individual one agreed with the statement. Now consider two individuals, with one selecting *Agree* and one selecting *Neither Agree nor Disagree*. With two individuals responding, there are more options for describing the findings. One option is to note that one individual agreed and one individual neither agreed nor disagreed. Another option is to note that the summated score mean response was an $M = 3.5$, $SD = .71$. Here, the interpretation is more difficult as the mean lies exactly halfway between two vague quantifiers. This is where an *a priori* convention for describing findings is needed. It would be an appropriate convention to state that SA = 5 - 4.51, A = 4.5 - 3.51, NAD = 3.5 - 2.51, D = 2.5 - 1.51, SD = 1.5 - 1. Using this rule, researchers would note that although individual participants had different responses, on the whole participants neither agreed nor disagreed that "*Advancements in Agricultural Development* has a positive influence on agricultural development practices worldwide through the rapid publication of research." An assumption for this rule is that the five-point summated scales have four equal intervals (see Figure 1): SD—D, D—NAD, NAD—A, and A—SA.

**Figure 1**

*Intervals and Interpretations of a Five-Point Summated Scale*

| Interval One | Interval Two | Interval Three | Interval Four |
|---|---|---|---|

| Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|---|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |

| SA | A | NAE | D | SD |
|---|---|---|---|---|
| 5-4.51 | 4.50 - 3.51 | 3.50 - 2.51 | 2.50 - 1.51 | 1.50 - 1 |

*Note.* An *n* point summated scale will have *n*-1 equal intervals

Simply reporting a mean score would be meaningless without describing the vague quantifier. The true limits of a scale are the lowest and highest numerical values assigned to the vague quantifiers. Similarly, the true limits of an individual vague quantifier are the lowest and highest numerical values assigned to that quantifier.

The data are not normally distributed in either of the two previously reported examples; let us now consider that option. Consider an example based on a single item scale score with 30 individuals responding. For this example, let us assume responses were as follows: SA($f$ = 5, 16.7%), A($f$ = 12, 40%), NAD($f$ = 12, 40%), and D($f$ = 1, 3.3%). It would be sufficient to describe the data in terms of the frequencies and percentages provided in the previous sentence. Additionally, because the data are normally distributed, reporting *M* and *SD* would also be appropriate (*M* = 3.67, *SD* = .88). A descriptive interpretation of this finding based on this rule is that participants tended to agree that "*AAD* had a positive influence on agricultural development practices worldwide through the rapid publication of research."

Now let us consider a summated unidimensional, attitudinal, summated scale, or Likert-type scale; see Table 2.

**Table 2**

*Examples of how Scaled Questions May be Written*

| Agricultural professionals' attitudes toward the influence of *Advancements in Agricultural Development* | Strongly Agree | Agree | Neither Agree nor Disagree | Disagree | Strongly Disagree |
|---|---|---|---|---|---|
| *Advancements in Agricultural Development* has a positive influence on agricultural development practices | 5 | 4 | 3 | 2 | 1 |
| *Advancements in Agricultural Development* reports research that have practical implications for agricultural professionals | 5 | 4 | 3 | 2 | 1 |
| *Advancements in Agricultural Development* presents research that reports practical implications in a timely manner | 5 | 4 | 3 | 2 | 1 |
| *Advancements in Agricultural Development* is a trusted source for agricultural professionals to publish their research | 5 | 4 | 3 | 2 | 1 |

Note that these are statements of behavior or attitude and that they have direction. In a summated scale, a total score or grand mean ($M^G$) score must be calculated. In our field of study, $M^G$ scores tend to be the preference, but the choice of mean statistic does not matter. Rather, the key to understanding the results is the interpretation of the scores. In this example, there are four statements used to define the construct being measured. Let us assume one respondent selected 4, 5, 4, and 4; $M$ = 4.25 and $SD$ = .50. Using the same convention described previously, the finding should be interpreted as the participant agreeing that "*AAD* had a positive influence on agricultural development practices worldwide through the rapid publication of research." With one respondent, it is impossible to estimate reliability. With two individuals responding there are more options for describing the findings. Let us assume a second responded select 4, 5, 5, and 5; $M$ = 4.75 and $SD$ = .50. In this example, one participant tended to agree, and one participant tended to strongly agree. The responsibility then falls on the investigator on how to interpret the finding. Based on the rule previously provided, the $M^G$ score would suggest that respondents collectively tended to agree ($M$ = 4.5, $SD$ = .53) that "*AAD* had a positive influence on agricultural development practices worldwide through the rapid publication of research." Estimating reliability again would be problematic given the small sample size and our understanding of the Central Limit Theorem (Field, 2018), which postulates that data tends to become normally distributed as $n$ approaches 30.

Continuing this example with an expanded sample size of $n$ = 30 assuming responses were as follows: R1|D($f$ = 1, 3.3%), NAD($f$ = 12, 40%), A($f$ = 12, 40%), and SA($f$ = 5, 16.7%); R2|NAD($f$ = 11, 36.7%), A($f$ = 14, 46.7%), and SA($f$ = 5, 16.7%); R3|D($f$ = 3, 10.0%), NAD($f$ = 11, 36.7%), A($f$ = 5, 16.7%), and SA($f$11, 36.7%); and R4|D($f$ = 4, 13.3%), NAD($f$ = 7, 23.3%), A($f$ = 11, 36.7%), and

SA($f$ = 8, 26.7%). Reliability is estimated for the summated unidimensional scale at a Cronbach alpha of .96. An $M^G$ and standard deviation for the summated scale are calculated ($M$ = 3.77, $SD$ = .85). The appropriate interpretation would be that participants agreed, "*AAD* had a positive influence on agricultural development practices worldwide through the rapid publication of research." The findings must be described using the same vague quantifiers used to collect the data and based on a rule set *a priori*.

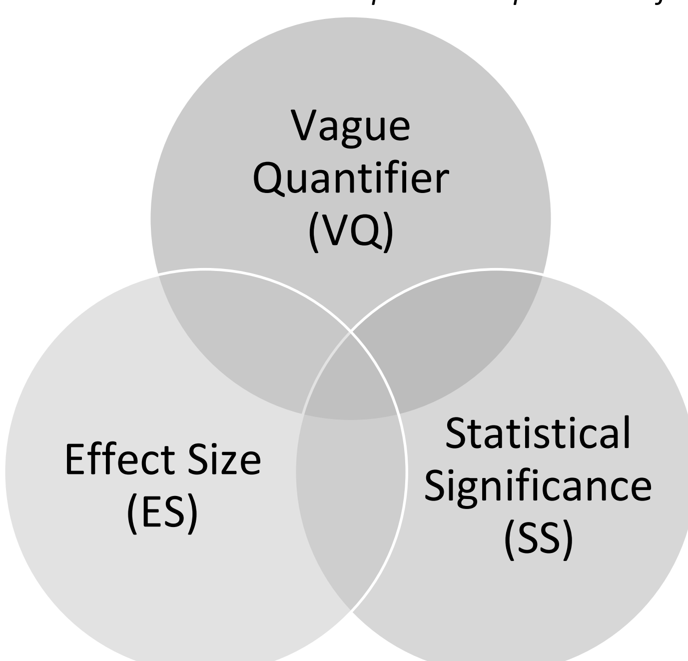
As noted previously, a five-point scale will have four intervals and five interpretations; see Table 3 in the *Recommendations*. So, while the intervals are at equal distances, the interpretations are not. A four-point scale would have three intervals and a six-point scale would have five intervals. When dealing with the type of scales described herein this *cannot* be overcome. Hence, the need to discuss statistical significances and effect sizes to better understand the findings of a study. This same problem occurs whether the data are treated as ordinal or interval. Stated as a caution, extending the numerical values of the beginning and ending of a scale to create equal intervals would also be problematic in that it would over or under inflate mean scores.

### Interpretations of findings

When analyzing findings of data collected using these scales, it is important to collectively consider the interpretation of the Vague Quantifiers (VQ), Statistical Significance (SS), and Effect Size (ES). Each interpretation provides one incomplete understanding of the findings. It is at the intersection of all three interpretations alone that presents a complete picture or maximized interpretation of the findings; this can and should be presented (see Figure 2).

**Figure 2**

*Maximizing Interpretation and the Relationship and Interpretation of Findings*



### Making and Understanding Interpretations

As described previously, all findings should be interpreted using, at a minimum, an appropriate VQ. If comparisons between groups or variables are made, then SS should be provided. Subsequently, an ES should be calculated to describe the magnitude and strength of the effect (Field, 2018). It is the reporting and analysis of these three interpretations that provides the most meaningful information related to the research and its consumer.

According to Field (2018), common measures of ES include Cohen's *d* and Pearson's correlation *r* (Cohen, 1988, 1992). The calculation and reporting of ES will help the researchers determine if the effect was small, medium, or large, and ultimately will add strength to data interpretations, discussions, and recommendations. Cohen (1988) operationally defined effect sizes as: small ($d = .2$) differences that are difficult to detect; medium effect sizes ($d = .5$) as differences that are evident by looking at them; and large ($d = .8$) differences that are clearly different. Cohen (1988) then cautions researchers by writing "the reader is counseled to avoid the use of these conventions, if he can, in favor of exact values provided by theory or experience in the specific area in which he is working" (p. 184). Cohen (1992) also noted that calculating and reporting ES and the value of power analysis in the establishment of sample sizes is often overlooked. The use of *nonstandardized effect sizes* could also be provided as means of describing the magnitude and strength of the effect. In our example above this would be achieved simply by subtracting one summated *M* from another summated *M* and interpreted using the convention provided in the recommendations below. The rationale for this is based on the intuitiveness of the measure, true limits of the scale, and scale intervals. Problems with logical interpretations exist when findings result in illogical descriptions. A couple of many possible examples are provided below.

Group A, $M^G = 3.50$ (*Neither Agree* nor *Disagree*) and Group B, $M^G = 3.51$ (*Agree*).

Group C, $M^G = 1.51$ (*Disagree*) and Group D, $M^G = 2.5$ (*Disagree*)

These problems provide a strong rationale for further descriptions using both statistical significance and effect size. For illustrative purposes, assumptions about statistical significance are made. For the type of research discussed in this paper, statistical significance is normally set a priori at an alpha of .05. Setting a statistical significance level a priori limits concerns about p - hacking or searching for significant differences after the research is completed (Head et al., 2015). Group A neither agreed nor disagreed that "*AAD* has a positive influence on agricultural development practices worldwide through the rapid publication of research." Group B agreed that "*AAD* has a positive influence on agricultural development practices worldwide through the rapid publication of research." There was no statistical significant difference between the two groups and the nonstandardized effect size was small. So, while the interpretations of the vague quantifiers indicate one conclusion, that there is a difference, a "so-what," or "meaningful" difference does not exist. This is referred to as a difference without distinction, a formal logical fallacy. Implications and recommendations based on this difference would be minimal to non-existent. Groups C and Group D disagreed that "*AAD* has a positive influence on agricultural development practices worldwide through the rapid publication of research." There was statistically significant differences between the two groups and the nonstandardized effect size was large. So, while the interpretations of the vague qualifiers indicate one conclusion, that there was no difference, a "so-what", or "meaningful" difference does exist. Implications and recommendations based on this difference should be maximized and highlighted despite the lack of difference in vague quantifiers.

## Alternate Scales and Other Considerations

If an investigator chooses to use a semantic differential scale (Osgood et al., 1957) or a Thurstone scale (1928), similar rules should be adopted *a priori* and interpretations made based on those rules. Semantic differential scales allow researchers to measure attitudes using a variety of adjectives with bipolar meanings, such as *good* and *bad*, or *interesting* and *boring*. Interpretations for a semantic differential scale would focus on participants' positive or negative attitudes toward a construct. Thurstone scales allow researchers to measure attitudes towards favorable and unfavorable statements regarding a particular issue using a two-point scale: *agree* and *disagree*. Interpretations for a Thurstone scale would focus on participants' tendency to *agree* or *disagree* to a particular issue.

How vague quantifiers are defined determines how they can be interpreted. Consider a scale using vague quantifiers related to importance. If you used a unidimensional four-point scale and labeled the vague quantifiers *important*, *moderately important*, *slightly importan*t, and *not important*, those same terms must be used to describe the findings. If an investigator only used *important* and *not important*, then those are the only terms that should be used to describe the findings. If they used a unidimensional scale and reported *M* and *SD*, then the findings should be reported in terms of the rules set *a priori* using the one appropriate vague quantifier. Discussion of where in the interval *M* falls cannot be reflected in changing the language of the *a priori* defined VQs. This is the purpose of discussing the ES along with the SS. When reporting *M* and *SD*, it is likely not appropriate to discuss data not presented unless a logical justification can be provided. For example, an *M* of 4 should be described as participants having agreed; noting a majority or plurality of respondents neither agreed nor disagreed would not be appropriate without reporting all the data. Within that same example, a majority could also include those that agreed and those that strongly disagree. If the researcher is going to calculate and report *M*, then *M* should be used for interpretation.

To this point, this discussion has focused on interpretations of means and standard deviations from *unidimensional*, summated, Likert-type, or attitudinal scales; this is regardless of whether the data are single item or summated scores and regardless of whether data are treated as ordinal or interval. This paper has not addressed the presentation of data in table format and whether *M* and *SD* or frequencies (*f*) or percentages (*%*) are most appropriate. While the reporting of *f* and *%* certainly provides more description than simply reporting *M* and *SD*, the latter is acceptable. In some instances, both *f* and *%* and *M* and *SD* are reported. The authors leave this discussion for another investigation, as ultimately it is the $M^G$ and $SD^G$ that are the focal point of this paper and discussion.

## Recommendations

Based on the previous discussions, the following recommendations are provided for interpreting scales described herein. Merely reporting means is insufficient for interpreting the types of scales described in this paper.

1. The scale used must be appropriately described in the methods. All vague quantifiers must be described in methods and findings. The true limits of the scale and the true limits of each vague quantifier should be described as shown in Table 3. This information should be placed in the methods section and may follow this general format. A five-point summated scale was used to collect data for this study: *Strongly Agree* = 5 – 4.51, *Agree* = 4.5 – 3.51, *Neither Agree nor Disagree* = 3.5 – 2.51, *Disagree* = 2.5 – 1.51, *Strongly Disagree* = 1.5 – 1.

**Table 3**

*Examples of Vague Quantifiers and Associated Values*

| **Five Point Scale** | | **Four Point Scale** | |
|---|---|---|---|
| *Vague Quantifier* | *Values* | *Vague Quantifier* | *Values* |
| Strongly Agree | 5.0 – 4.51 | Strongly Agree | 4.0 – 3.51 |
| Agree | 4.5 – 3.51 | Agree | 3.5 – 2.51 |
| Neither Agree nor Disagree | 3.5 – 2.51 | Disagree | 2.5 – 1.51 |
| Disagree | 2.5 – 1.51 | Strongly Disagree | 1.5 – 1.00 |
| Strongly Disagree | 1.5 – 1.00 | | |

*Note*. The scale convention should be adjusted based on the number and order of vague quantifiers.

Calculate, report, and interpret the *M* and *SD* of scales based on rules set *a priori*. If data does not have a purposeful and meaningful order (other than the order the questions were asked), then report by highest mean or percentage first. When using *M* and *SD* in a table also add a note at the bottom of the table describing all vague quantifiers. Discuss those items with the highest and lowest means or those items of interest if data are dichotomized.

2. Report statistical analysis and highlight differences. *Alpha* for statistical analysis should be set *a priori*. In the type of research being discussed herein *alpha* is typical set *a priori* at .05. Caution is warranted against data dredging or *p*-hacking (Head, et al., 2015).

3. Calculate and report effect sizes using an appropriate standardized or nonstandardized convention. This will aid in the interpretation of the data and provide a strong response to the “so what” question of differences detected. In our field, medium and large *ES* have the

most value when making implications and recommendations. Small *ES* often do not reach the “so what” level of value.

4. A nonstandardized effect size calculation can help in describing the magnitude and strength of the effect. For a summated, unidimensional, attitudinal, and Likert-type scale, this can be accomplished by simply subtracting mean scores. A convention for interpreting nonstandardized effect sizes should be reported.

   The convention for a *nonstandardized effect size* scales described in this paper could be interpreted as shown below.

   Small (*ES* = .19 and lower)

   Medium (*ES* = .20 - .49)

   Large (*ES* = .50 and higher)

   Using the type of scaling described in this paper, a nonstandardized effect size calculation may be the most appropriate, but as Cohen cautioned such calculations and interpretations should be ultimately determined by what researchers in a given field of study are attempting to accomplish. The convention provided above is based on the minimum and maximum values of the vague quantifiers provided above. Referring to Figure 1, note that a difference of .5 at the lowest and highest end of a scale would move the interpretation from one vague quantifier category to another, and from the middle of the other vague quantifier categories to another. For a standardized effect size calculation, Cohen’s *d* and Pearson’s correlation *r* are widely recognized as robust and useful. Their use, however, is limited for a variety of reasons including complications in calculations. Hence, the recommendation to consider a nonstandardized convention is provided.

5. From our example earlier in the paper, what follows is a sample of how findings can be described; this is a general format of which infinite other formats can be generated. Overall participants tended to *Agree* (*M* = 3.77, *SD* .85) that “*AAD* had a positive influence on agricultural development practices worldwide through the rapid publication of research.” Faculty (*M* = 4.1, *SD* = .76) tended to *Agree* with the statement while Graduate Students (*M* = 3.43, *SD* = .83) tended to *Neither Agree nor Disagree*. Comparing attitudes between faculty and graduate students using a *t-test* (alpha set a *priori* at .05), t (28) = 2.30, $p < .05$, a statistical difference existed. Faculty were more likely to agree, and this finding was statistically significant, and the *ES* (.67) was large.

6. When illogical findings occur in statistical testing, additional discussion is warranted. For example, let us assume that Master’s students had a mean score of 3.5 (*Neither Agree nor Disagree*) on the construct discussed in Table 2. Second, assume doctoral students had a mean score of 3.51 (*Agree*). While differences between master’s students and doctoral students are interpreted as being different, there is no statistical difference, no practical difference, and the effect size is minimal. There is no “so what” here and caution is

warranted against conclusions being drawn from the differences in interpretation. Now consider augmenting the example. Let us assume master's students had a mean score of 3.51 (*Agree*) on the construct discussed in Table 2; doctoral students had a mean score of 4.5 (*Agree*). Data suggests the interpretation is the same for both. Let us assume there is a statistical difference between the groups and that the effect size is large. In this example, additional discussion should focus on the true differences despite having the same vague quantifiers; statistical significance, and a large effect size.

The attributes and viability of the recommendations provided above will be borne out over time (Lindner et al., 2016). If the recommendations prove to be appropriate, they will be adopted, if not they will be discarded, perhaps because they were not useful or a better alternative was available (Lindner et al., 2001).

## Acknowledgements

**James Lindner** – conceptualization, methodology, writing original draft, review, and editing; **Nicholas Lindner** – writing original draft, review, and editing.

## References

Clason, D. L., & Dormody, T. J. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education, 35*(4), 31-35. https://doi.org/10.5032/jae.1994.04031

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.) Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

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. (2018). *Discovering statistics using IBM SPSS statistics: North American edition* (5th ed.). Sage.

Gardner, D. G., Cummings, L. L., Dunham, R. B., & Pierce, J. L. (1998). Single-item versus multiple-item measurement scales: An empirical comparison. *Educational and Psychological Measurement, 58*(6), 898–915. https://doi.org/10.1177/0013164498058006003

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, *13*(3), e1002106. https://doi.org/10.1371/journal.pbio.1002106

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

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

Lindner, J. R., Rodriguez, M. T., Strong, R., Jones, D., & Layfield, D. (2016). Research priority area 2: New technologies, practices, and products adoption decisions. *American Association for Agricultural Education National Research Agenda*, *2020*, 19-27. https://aaea.wildapricot.org/resources/Documents/AAAE_National_Research_Agenda_2016-2020.pdf

Miller, L. E. (1998). Appropriate analysis. *Journal of Agricultural Education, 39*(2), 1-10. https://doi.org/10.5032/jae.1998.02001

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education 15*, 625–632. https://doi.org/10.1007/s10459-010-9222-y

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. University of Illinois Press.

Sisson, D. V., & Stocker, H. R. (1989). Research corner: analyzing and interpreting Likert-type survey data. *Delta Pi Epsilon Journal, 31*(2), 81-85.

Sullivan, G. M., & Arino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Educating, 5*(4), 541-542. https://doi.org/10.4300%2FJGME-5-4-18

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*(4), 529-554. https://doi.org/10.1086/214483

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