Investigation of Bivariate Grid-Type Items for Measuring Attitudes in Statistics Education: Preliminary Results

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada 2022 Annual Meeting

30 May 2022

Background & Motivation

- Noted difficulties with measuring change in attitudes during a single introductory statistics course (e.g., Whitaker, Unfried, et al., 2022)
  - *What if students enter introductory statistics courses without well-formed attitudes about statistics?*
  - *And then, what if a single course in statistics results in students forming stronger attitudes about statistics… that are still neutral?*

- Likert-type items (often used on attitude surveys) only allow for a single type of neutral attitude to be measured.

- So, let's try a different type of item!
  - Evaluative Space Grid (ESG)

ESG: Beyond Bipolar Scales

Cacioppo and Berntson (1994) articulate three assumptions for using bipolar scales (e.g., Likert-type items) for measuring attitudes:

1. An attitude is a joint function of positive (appetitive) and negative (aversive) affective/motivational reactions to a stimulus.
2. Positive and negative reactions to a stimulus have generally opposing effects on an attitude.
3. The positive and negative reactions that determine an attitude toward a stimulus are essentially reciprocally controlled. (Cacioppo, Gardner, & Berntson, 1997, pp. 5-6)

ESG: Beyond Bipolar Scales

Cacioppo and Berntson (1994) articulate three assumptions for using bipolar scales (e.g., Likert-type items) for measuring attitudes:

1. An attitude is a joint function of positive (appetitive) and negative (av
2. Po
opposing effects on an attitude.
3. ~~The positive and negative reactions that determine an attitude toward a stimulus are essentially reciprocally controlled.~~ (Cacioppo, Gardner, & Berntson, 1997, pp. 5-6)

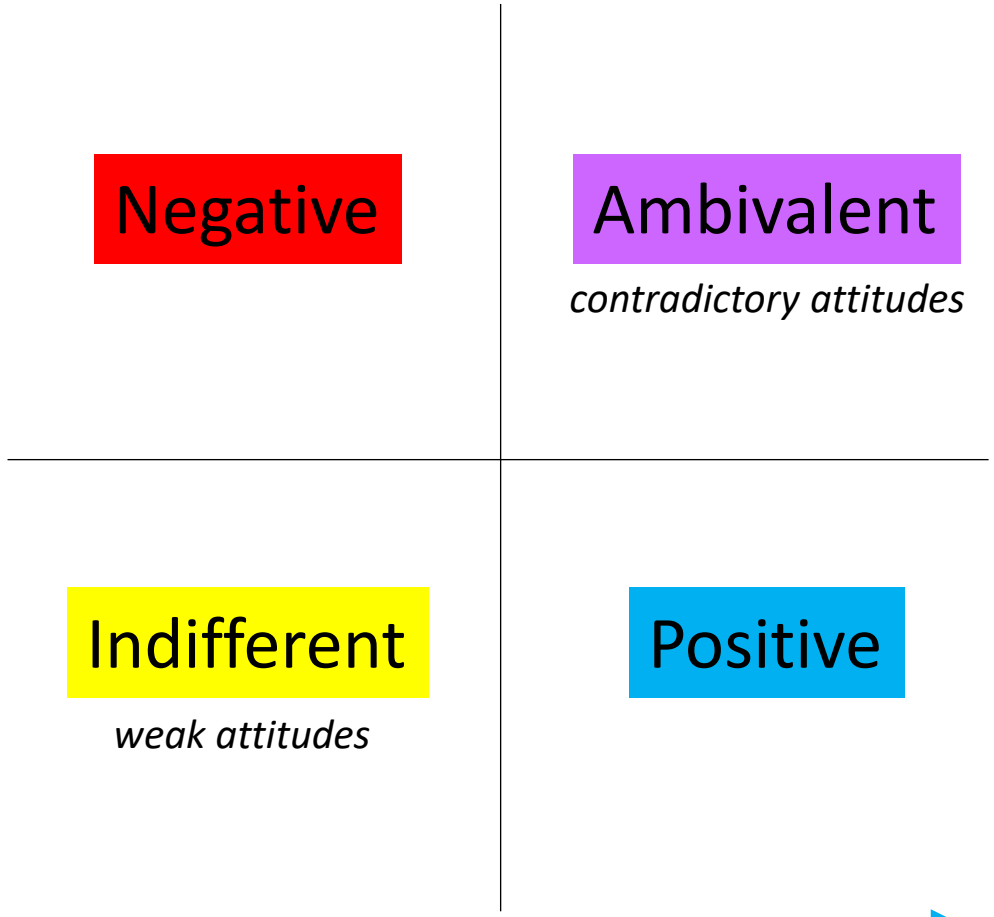
What if we do not require a reciprocal relationship between positive and negative reactions?

Evaluative Space Grid:

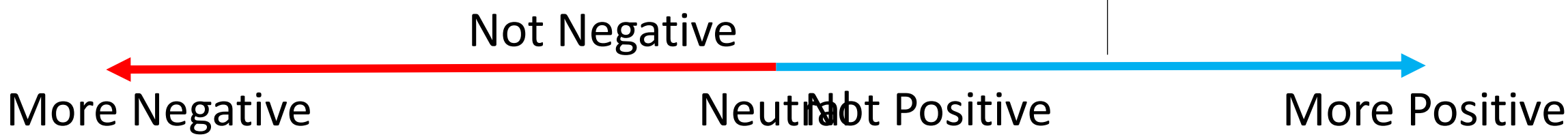
Not requiring a reciprocal relationship

Evaluative Space Grid:

Not requiring a reciprocal relationship



- **Labeling of quadrants**: (Audrezet, 2014)

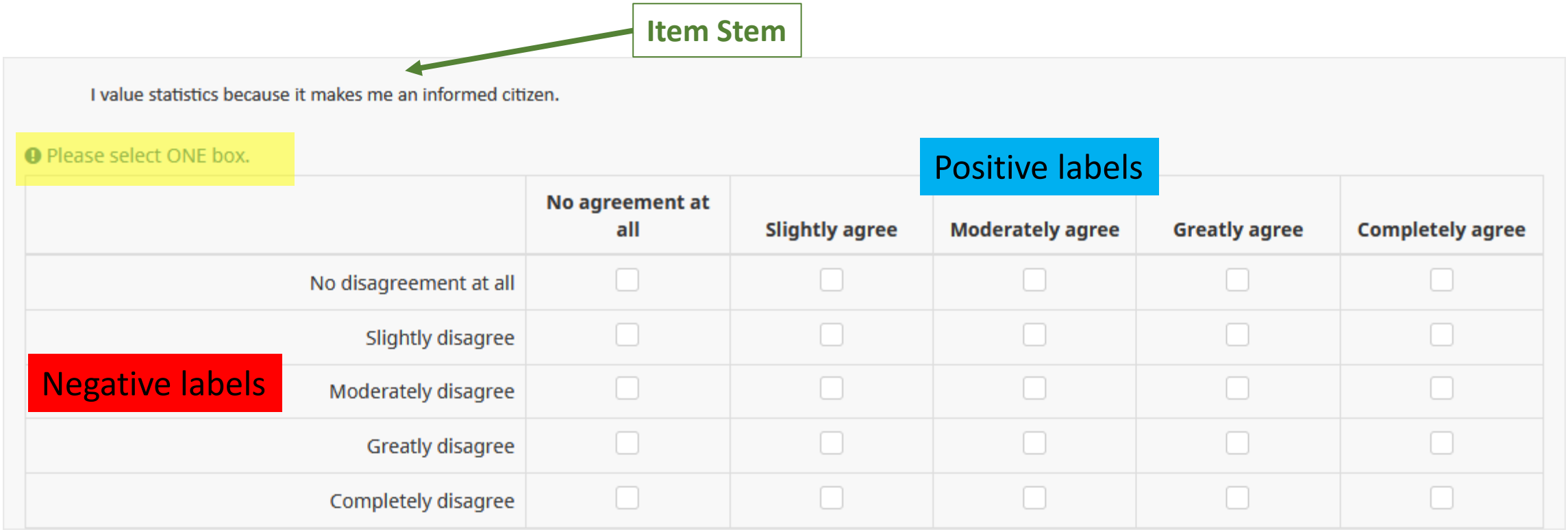


- **Evaluative Space Grid**: (Cacioppo & Berntson, 1994; Cacioppo et al., 1997; Larsen et al., 2009)

ESG: Beyond Bipolar Scales

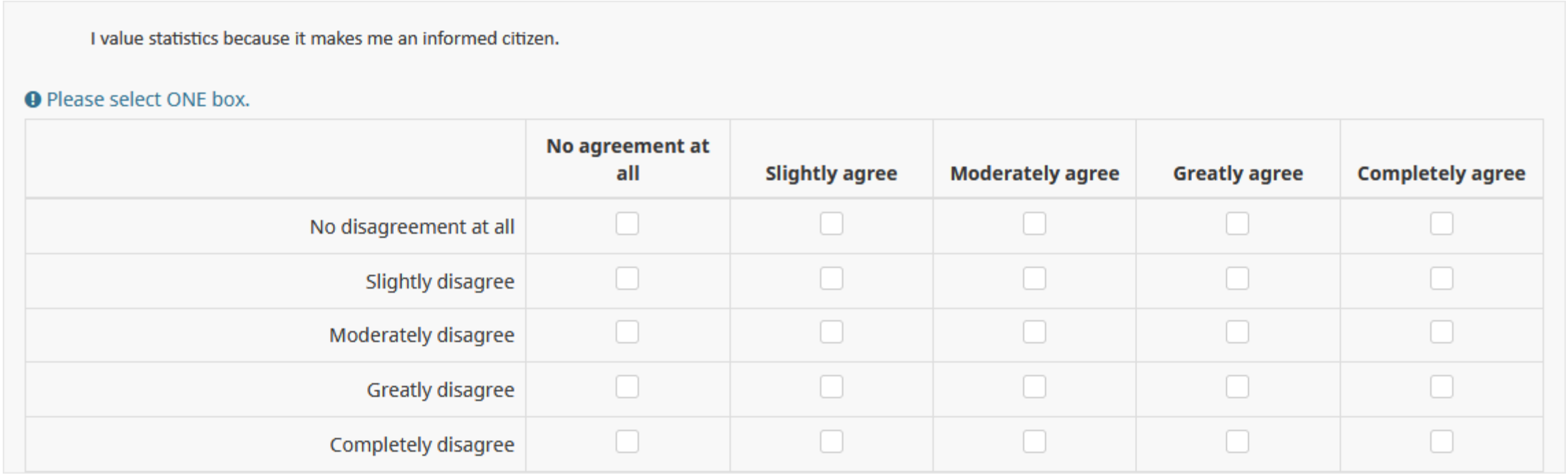
- Potential advantage: Better describe respondents' attitudes that would be ordinarily described as "neutral" (Cacioppo et al., 1997; Larsen et al., 2009)
- Potential challenges/disadvantages: many?
- Recent studies have been in the area of marketing/customer satisfaction
  - Restaurant or physician evaluation (Audrezet 2014; Audrezet et al., 2016; Audrezet & Parguel, 2018)
  - Attitudes of Swiss transit customers (Borriello, 2017)
- Perhaps useful for some constructs in Expectancy-Value Theory (EVT)? (EVT; e.g., Eccles, 1983, 2014; Eccles & Wigfield, 2002)
  - A theoretical framework for attitudes in statistics education
    - Widely-used SATS-36 (Schau, 2003) and in-development S-SOMAS (e.g., Unfried et al., 2021)
  - Limited research on some constructs (e.g., Cost/Negative Values)

Example Item (LimeSurvey)



Limitations around where labels can be placed in LimeSurvey resulted in the decision to reverse the direction of the vertical axis.

Example Item (LimeSurvey)



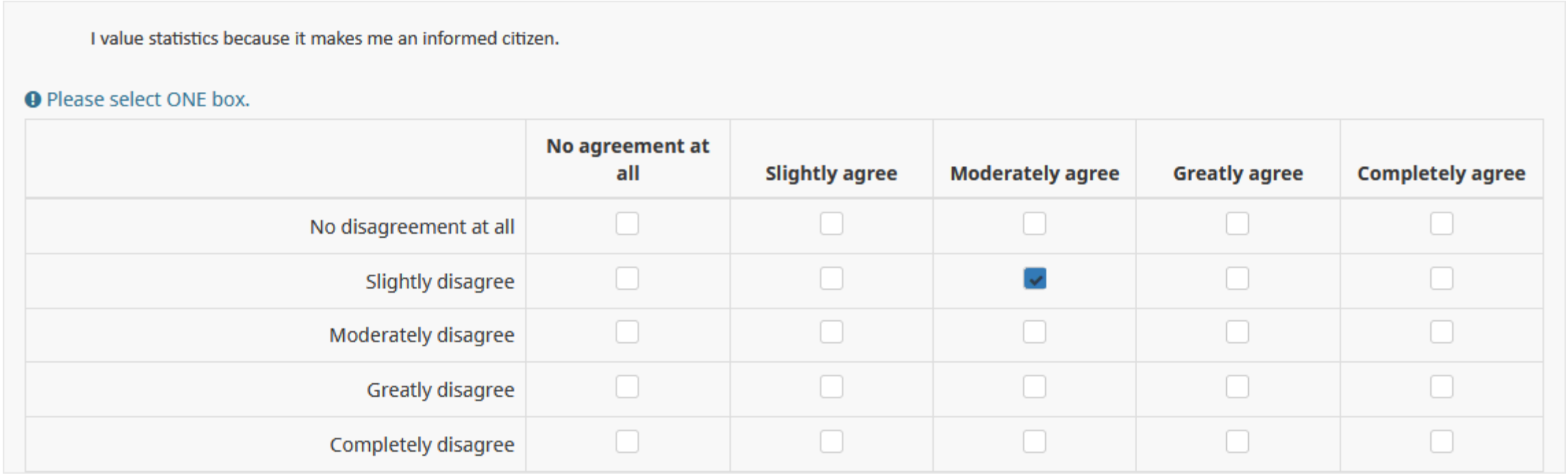
I value statistics because it makes me an informed citizen.

Please select ONE box.

	No agreement at all	Slightly agree	Moderately agree	Greatly agree	Completely agree
No disagreement at all	☐	☐	☐	☐	☐
Slightly disagree	☐	☐	☐	☐	☐
Moderately disagree	☐	☐	☐	☐	☐
Greatly disagree	☐	☐	☐	☐	☐
Completely disagree	☐	☐	☐	☐	☐

Limitations around where labels can be placed in LimeSurvey resulted in the decision to reverse the direction of the vertical axis.

Example Item (LimeSurvey)



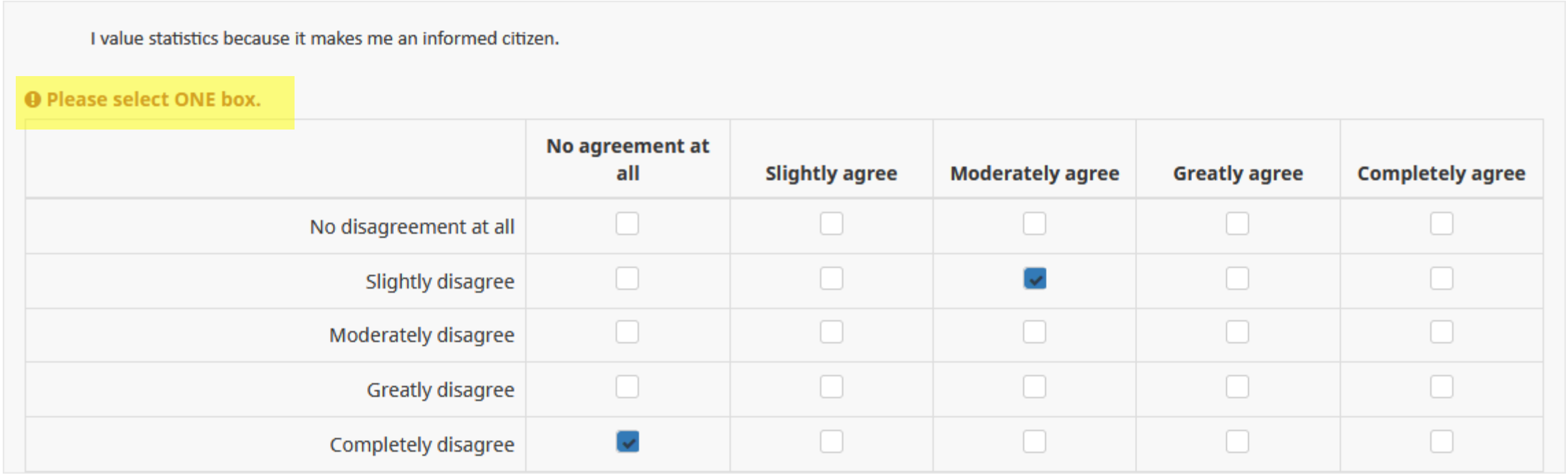
I value statistics because it makes me an informed citizen.

Please select ONE box.

	No agreement at all	Slightly agree	Moderately agree	Greatly agree	Completely agree
No disagreement at all	☐	☐	☐	☐	☐
Slightly disagree	☐	☐	☑	☐	☐
Moderately disagree	☐	☐	☐	☐	☐
Greatly disagree	☐	☐	☐	☐	☐
Completely disagree	☐	☐	☐	☐	☐

A single box should be checked to indicate a response to an ESG-type item.

Example Item (LimeSurvey)



LimeSurvey allows custom question validation code, but other survey platforms might not facilitate validation for ESG-type questions.

Description Current Studies

Study Name	Data Collected	Purpose	Population	Total Responses	Good Responses	Scales (Likert-type and ESG-type)
Study 0	Winter 2020	Proof of concept; basic properties	Introductory Statistics Students	42	33	• EVT Cost of Learning Statistics (S-SOMAS Pilot; Whitaker et al., 2019) • EVT Task Effort Cost (Flake et al., 2015) • EVT Emotional Cost (Flake et al., 2015)
Study A	Winter 2022	Proof of concept; basic properties	Amazon Mechanical Turk	133	51	• EVT Utility Value of Learning Statistics (S-SOMAS Pilot; Whitaker et al., 2019) • Dependent Learning (TOOLS; Kerr et al., 2006) • Conscientiousness (Big Five; Goldberg, 1992)
Study B	Winter 2022	Invalid Response Rate Comparison	Amazon Mechanical Turk	301	164	Big Five (Goldberg, 1992) – *ESG-type only* • Extraversion • Agreeableness • Conscientiousness • Emotional Stability • Intellect/Imagination

Scales are all from existing instruments that use Likert-type items.

Notes:

- All data collected using LimeSurvey
- Some results from Study 0 published (Whitaker, Barss, et al., 2022)
- Further data collection planned with introductory statistics students
- ESG-type items were created by using existing item stems with the ESG response scale
- Amazon Mechanical Turk participants were paid US$2.56 for participating

Description Current Studies

- There are three sets of results we will examine:
  - **Invalid Response Rates**: Determine the rate of invalid responses to ESG-type items when no question validation is used

    Study B
  - **Response Times**: Compare the average time per item for Likert-type and ESG-type items

    Study A
  - **Internal Consistency**: Examine the internal consistency of scales (and compare to reference values)
    - Especially Expectancy-Value Theory (EVT) scales (e.g., Eccles, 1983; Eccles & Wigfield, 2002)

    Study 0 Study A
- All trying to address this big question: *Are ESG-type items appropriate for use in statistics education attitude research?*

Invalid Response Rates

Determine the rate of invalid responses to ESG-type items when no question validation is used

- Participants randomly assigned to groups:
  - Group 1: "Long Instructions" ($N_1 = 81$)
  - Group 2: "Short Instructions" ($N_2 = 83$)
  - Both groups received "Long Instructions" for the *first* page of ESG-type items
- 52-items on survey (50 Big Five plus 2 validation)
- Aim: Determine the number of instances where an invalid response is entered (i.e., 2 or more boxes selected on one item)
  - Participants who provided invalid responses for more than *half* of the items are excluded
  - A valid response is selecting *1 or 0* boxes

Aggregating responses across all items and participants, we have the following counts:

	Invalid Responses	Valid Responses
Group 1	33	3659
Group 2	22	3878

The error rate is likely to differ for each participant (or different types of participants).

No clear evidence of a difference in the overall rates between Groups 1 and 2.

Baseline error rate estimate: 0.72%

Study A

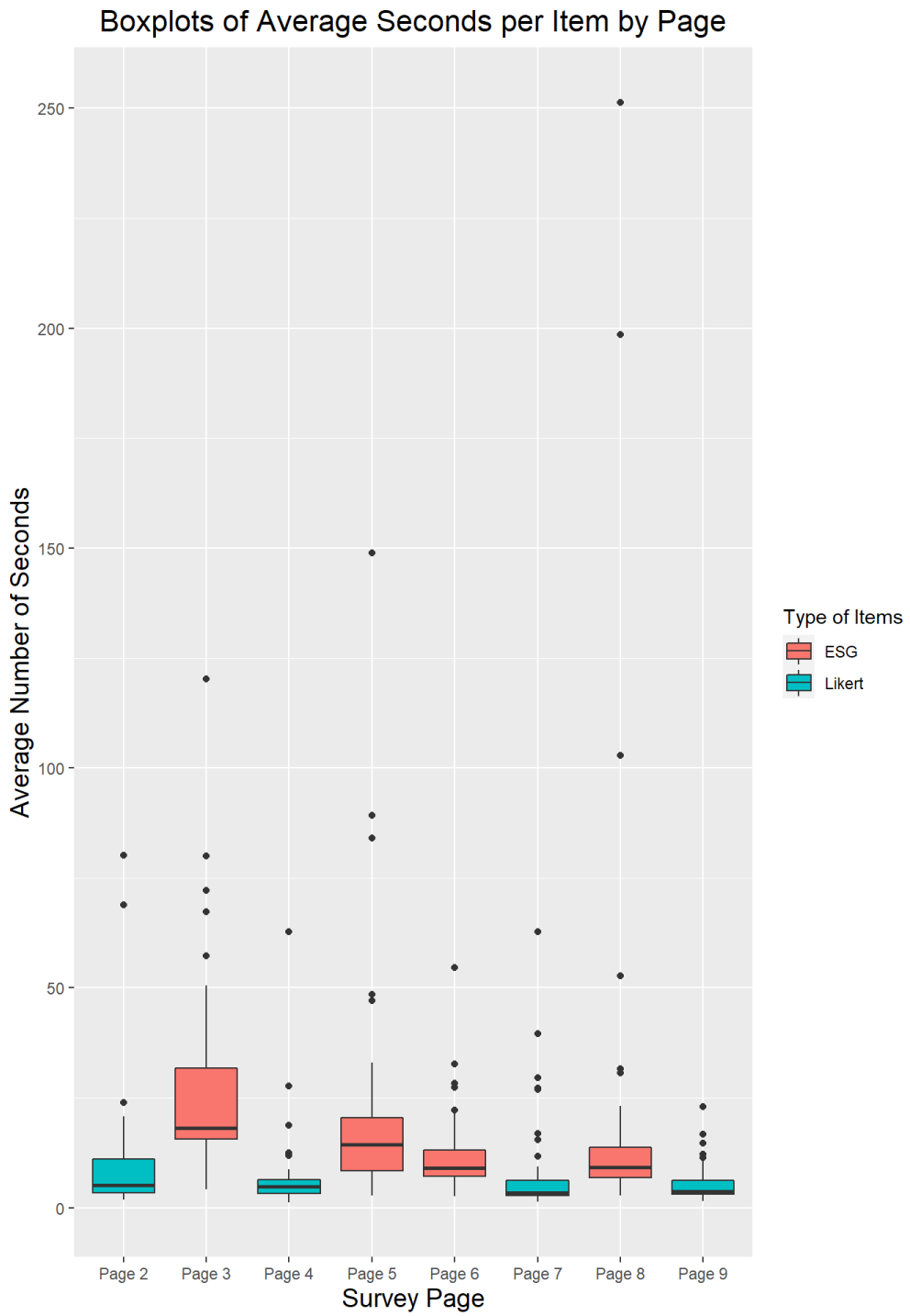
Response Times

Compare the average time per item for Likert-type and ESG-type items

- Friedman's Test then Conover's post-hoc
  - Likert-type
    - Page 9 vs. Page 2 [First Likert-type page] - p-value: 0.37049
    - Page 9 vs. Page 4 - p-value: 0.99634
    - Page 9 vs. Page 7 - p-value: 0.99974
  - ESG-type
    - Page 8 vs. Page 3 [First ESG-type page] - p-value: 0.00077
    - Page 8 vs. Page 5 - p-value: 0.29906
    - Page 8 vs. Page 6 - p-value: 0.99996
- ***ESG-type items seem to take about 2-3 times as long to complete as Likert-type items after familiarity***

Statistic	Page 2	Page 3	Page 4	Page 5	Page 6	Page 7	Page 8	Page 9
Minimum	1.8	4.3	1.3	2.8	2.7	1.4	2.8	1.5
1st Quartile	3.4	15.6	3.3	8.4	7.2	2.7	6.8	3.1
Median	5.1	18.2	4.8	14.3	9.1	3.4	9.2	3.7
Mean	9.8	26.4	6.7	21.0	11.5	7.7	21.2	5.5
3rd Quartile	11.1	31.7	6.3	20.5	13.1	6.2	13.7	6.3
Maximum	80.2	120.2	62.8	148.9	54.6	62.7	251.4	22.9

Summary of respondents' average time per item by page (in seconds)



Study 0 | Study A

Internal Consistency

Examine the internal consistency of scales (and compare to reference values)

Study	Scale	Coefficient Alpha: Likert-type	Coefficient Alpha: ESG-type	*Reference Value*
Study 0	EVT Cost of Learning Statistics			*(S-SOMAS Pilot 0; N=1175; unpublished)*
Study 0	EVT Task Effort Cost			*(Flake et al., 2015)*
Study 0	EVT Emotional Cost			*(Flake et al., 2015)*
Study A	EVT Utility Value of Learning Statistics			*(S-SOMAS Pilot 1; N=500; unpublished)*
Study A	Dependent Learning			*(TOOLS; Kerr et al., 2006)*
Study A	Conscientiousness			*(Big Five; International Personality Item Pool, n.d.)*

Notes:

- Bivariate responses from ESG-type items were converted to a unidimensional value using an extension of the mapping proposed by Audrezet et al. (2016).
- There are limitations to using coefficient alpha (e.g., Schmitt, 1996) but values are readily available in the literature for comparison.

Coefficient Alpha	Interpretation (Henson, 2001)
0.90	Common threshold for higher-stakes uses of instruments (e.g., educational/clinical decisions)
0.80	Common threshold for standard research uses of instruments
0.70	Common threshold for low-stakes, exploratory research uses of instruments
0.60	Older threshold for low-stakes, exploratory research uses of instruments

Study 0 Study A

Internal Consistency

Examine the internal consistency of scales (and compare to reference values)

Study	Scale	Coefficient Alpha: Likert-type	Coefficient Alpha: ESG-type	Coefficient Alpha: *Reference Value*	
Study 0	EVT Cost of Learning Statistics	0.85	0.80	*0.84*	*(S-SOMAS Pilot 0; N=1175; unpublished)*
Study 0	EVT Task Effort Cost	0.95	0.93	*0.95*	*(Flake et al., 2015)*
Study 0	EVT Emotional Cost	0.88	0.80	*0.94*	*(Flake et al., 2015)*
Study A	EVT Utility Value of Learning Statistics	0.83	0.80	*0.88*	*(S-SOMAS Pilot 1; N=500; unpublished)*
Study A	Dependent Learning	0.88	0.85	*0.70*	*(TOOLS; Kerr et al., 2006)*
Study A	Conscientiousness	0.82	0.65	*0.79*	*(Big Five; International Personality Item Pool, n.d.)*

Notes:

- Bivariate responses from ESG-type items were converted to a unidimensional value using an extension of the mapping proposed by Audrezet et al. (2016).
- There are limitations to using coefficient alpha (e.g., Schmitt, 1996) but values are readily available in the literature for comparison.

Coefficient Alpha	Interpretation (Henson, 2001)
0.90	Common threshold for higher-stakes uses of instruments (e.g., educational/clinical decisions)
0.80	Common threshold for standard research uses of instruments
0.70	Common threshold for low-stakes, exploratory research uses of instruments
0.60	Older threshold for low-stakes, exploratory research uses of instruments

Takeaways

- Low invalid response rate (less than 1%) may mean that other survey platforms would be reasonable even without question validation Study B
- ESG-type items seem to take about 2-3 times as long to complete as Likert-type items after familiarity Study A
- Coefficient alpha levels generally seem similar when comparing scales with Likert-type items and scales with ESG-type items… Study 0 Study A
  - … to each other
  - … to reference values
  - Some notable exceptions (maybe based on response scale)?

Limitations and Future Work

- Disinterested participants – low quality responses? Need to examine further.
- ESG-type items are created by naïvely adapting existing items – may not work well in all cases (especially Big Five?)
- Need to determine which types of constructs (if any) benefit from ESG-type items (chosen constructs might not)
- High barrier for other researchers to use – *GridItemTools* R package being created (currently only supports data from LimeSurvey)
- Need to collect longitudinal data!

Acknowledgements & Funding

- Research Assistants:
  - Nerissa Gailey, Joseph Barss, Bailey Drew, and D'Vaughn Rolle-Johnson
- Jessica Flake (for permission to use her EVT Cost scales)
- The MASDER Team (for the S-SOMAS scales)
  - http://sdsattitudes.com
- Funding:
  - Studies A & B were supported by an internal grant from Mount Saint Vincent University.
  - The S-SOMAS Utility Value items from Study A and the S-SOMAS Pilot 1 data used to compute the reference value of coefficient alpha are based upon work supported by the National Science Foundation under Grant No. 2013392. (The ESG item studies are not a component of the NSF-funded work.)

Future Updates

- The Surveys of Motivational Attitudes toward Statistics/Data Science (SOMAS / SOMADS)
  - Student & Instructor versions
  - Student SOMAS should be finalized in AY 2022-2023 and then made freely available for broad use
  - http://sdsattitudes.com

- Updates on my ESG research:
  - R package: https://github.com/douglaswhitaker/GridItemTools
  - My website: http://douglaswhitaker.ca

References

Audrezet, A. (2014). *L'ambivalence des consommateurs: Proposition d'un nouvel outil de mesure* [Thèse de doctorat, Université Paris-Dauphine]. https://tel.archives-ouvertes.fr/tel-01084068/document

Audrezet, A., Olsen, S. O., & Tudoran, A. A. (2016). The GRID scale: A new tool for measuring service mixed satisfaction. *Journal of Services Marketing*, *30*(1), 29–47. https://doi.org/10.1108/JSM-01-2015-0060

Audrezet, A., & Parguel, B. (2018). Using the Evaluative Space Grid to better capture manifest ambivalence in customer satisfaction surveys. *Journal of Retailing and Consumer Services*, *43*, 285–295. https://doi.org/10.1016/j.jretconser.2018.04.008

Borriello, A. (2017). *The role of attitudes in determining individual behavior in transportation—From psychology to discrete choice modeling* [Dissertation, Università della Svizzera Italiana]. http://doc.rero.ch/record/306643/files/2018ECO001.pdf

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, *115*(3), 401–423. https://doi.org/10.1037/0033-2909.115.3.401

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond Bipolar Conceptualizations and Measures: The Case of Attitudes and Evaluative Space. Personality and Social Psychology Review, 1(1), 3–25. https://doi.org/10.1207/s15327957pspr0101_2

Eccles, J. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), Achievement and achievement motives: Psychological and sociological approaches (pp. 75–145). San Francisco: W.H. Freeman.

Eccles, J. S. (2014). Expectancy-Value Theory. In R. Eklund & G. Tenenbaum (Eds.), Encyclopedia of Sport and Exercise Psychology. https://doi.org/10.4135/9781483332222.n110

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. Annual Review of Psychology, 53, 109–132.

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. Contemporary Educational Psychology, 41, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Henson, R. K. (2001). Understanding Internal Consistency Reliability Estimates: A Conceptual Primer on Coefficient Alpha. *Measurement and Evaluation in Counseling and Development, 34*(3), 177–189.

International Personality Item Pool. (n.d.). *Characteristics of the Preliminary IPIP Scales Measuring the Big-Five Domains*. Retrieved May 24, 2022, from https://ipip.ori.org/newBigFive5broadTable.htm

Kerr, M. S., Rynearson, K., & Kerr, M. C. (2006). Student characteristics for online learning success. *The Internet and Higher Education*, *9*(2), 91–105. https://doi.org/10.1016/j.iheduc.2006.03.002

Larsen, J. T., Norris, C. J., McGraw, A. P., Hawkley, L. C., & Cacioppo, J. T. (2009). The evaluative space grid: A single-item measure of positivity and negativity. *Cognition & Emotion*, *23*(3), 453–480. https://doi.org/10.1080/02699930801994054

Schau, C. (2003). *Survey of Attitudes Toward Statistics (SATS-36)*. http://evaluationandstatistics.com/

Schmitt, N. (1996). Uses and Abuses of Coefficient Alpha. *Psychological Assessment, 8*(4), 350–353. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.615.4053&rep=rep1&type=pdf

Unfried, A., Posner, M., Bond, M., Kerby-Helm, A., Bolon, W., Whitaker, D., & Batakci, L. (2021, August). *Why Do We Need Yet ANOTHER Instrument Measuring Student Attitudes?* Presentation at the 2021 Joint Statistical Meetings (JSM), Virtual. http://sdsattitudes.com/wp/jsm2021/

Whitaker, D., Barss, J., & Drew, B. (2022). Measuring Opportunity Cost in Statistics Using Evaluative Space Grid Items: Results from a Pilot Study. *Journal of Research in Science, Mathematics and Technology Education*, *5*(1), 11–46. https://doi.org/10.31756/jrsmte.511

Whitaker, D., Drew, B., & Barss, J. (2021) *GridItemTools: Grid item tools*. R package version 0.0.12. https://github.com/douglaswhitaker/GridItemTools

Whitaker, D., Unfried, A., & Bond, M. E. (2022). Challenges associated with measuring attitudes using the SATS family of instruments. *Statistics Education Research Journal*, *21*(1), Article 4. https://doi.org/10.52041/serj.v21i1.88

Questions?

Thank you!

Investigation of Bivariate Grid-Type Items for Measuring Attitudes in Statistics Education: Preliminary Results

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada 2022 Annual Meeting

30 May 2022

MOUNT SAINT VINCENT UNIVERSITY

Supplemental Slides

Investigation of Bivariate Grid-Type Items for Measuring Attitudes in Statistics Education: Preliminary Results

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada 2022 Annual Meeting

30 May 2022

MOUNT SAINT VINCENT UNIVERSITY

Mapping ESG-type responses to unidimensional scores

- Proposed by Audrezet et al. (2016) because it satisfies a set of six constraints:

$$S(i,j) = (b+2)i + bj - 1 - 6b$$

  - with $-1 < b < 0$; usually $b = -0.5$ is chosen.
  - This form is limited to mapping to a 9-point response scale (with 1 = lowest and 9 = highest).

- Extension for arbitrary end-points:

$$S(i,j) = (i-1)\left(\frac{RespScale_{Upper} - RespScale_{Lower}}{4}\right) + (i+j-6)b + RespScale_{Lower}$$

  - with $\frac{RespScale_{Lower} - RespScale_{Upper}}{8} < b < 0$; usually $b = \frac{RespScale_{Lower} - RespScale_{Upper}}{16}$ is chosen
  - When $RespScale_{Upper} = 9$ and $RespScale_{Lower} = 1$, this form simplifies to the above.

- In both, i represents the positive response and j represents the negative response (1 = low, 5 = high)

Study B: Long Instructions

- Instructions: These items all use a grid to record your responses. Using the grid, you will be able to record the extent to which you agree and the extent to which you disagree with each statement. Please select the ONE box that best describes your OVERALL feeling about each statement.
- Example 1: Taylor is responding to the item “I like eating kale.” Taylor really dislikes the taste of kale, but also knows that kale has a lot of nutrients. Taylor chooses the box that corresponds to “Greatly disagree” (because Taylor does not like the taste) and “Moderately agree” (because Taylor appreciates the nutritional value of kale).
- The highlighting represents Taylor’s thinking – there will be no highlighting on the survey. Taylor selects the box in the corresponding row and column.
- Example 2: Drew is responding to the item “I dislike driving in Halifax.” Drew really hates the traffic during rush hour, but also finds driving to be more convenient than other transportation options. Drew chooses the box that corresponds to “Completely agree” (because Drew hates driving in heavy traffic) and “Moderately disagree” (because Drew appreciates the convenience driving).
- The highlighting represents Drew’s thinking – there will be no highlighting on the survey. Drew selects the box in the corresponding row and column.

Study B: Short Instructions

- Instructions: These items all use a grid to record your responses. Using the grid, you will be able to record the extent to which you agree and the extent to which you disagree with each statement. Please select the ONE box that best describes your OVERALL feeling about each statement.

Investigation of Bivariate Grid-Type Items for Measuring Attitudes in Statistics Education: Preliminary Results

Douglas Whitaker
Mount Saint Vincent University

Statistics Society of Canada 2022 Annual Meeting

30 May 2022

Abstract

Likert-type items are ubiquitous in attitude research in statistics education but imply a reciprocal relationship between positivity and negativity in the construct being measured. Based on historical challenges measuring some constructs in a widely used framework in statistics education (Eccles's Expectancy-Value Theory [EVT]), we speculate that the reciprocal relationship implied by the Likert-type items may not be appropriate. Evaluative Space Grid (ESG) items have been proposed as an alternative: respondents indicate their positivity and negativity on a grid that does not impose a reciprocal relationship. However, there have been relatively few studies that focus on ESG items. This presentation reports on a set of preliminary studies that seek to describe the psychometric properties of ESG items and document evidence of their appropriateness (or lack thereof) for measuring EVT constructs. Data have been collected from introductory statistics students and a general participant pool.

Items de type grille bivariée pour mesurer les attitudes en enseignement de la statistique : résultats préliminaires

Douglas Whitaker
Université Mont Saint-Vincent

Congrès annuel 2022 de la Société statistique du Canada

30 mai 2022

MOUNT SAINT VINCENT UNIVERSITY

Résumé

Les items de type Likert sont souvent utilisés dans la recherche sur les attitudes en enseignement de la statistique, mais ils impliquent une relation réciproque entre la positivité et la négativité dans le concept mesuré. Puisqu'il semble difficile de mesurer certains concepts dans un cadre largement utilisé en enseignement de la statistique (théorie de l'espérance-valeur d'Eccles, ou EVT), nous supposons que la relation réciproque impliquée par les items de type Likert n'est peut-être pas appropriée. Les items de type Evaluative Space Grid (ESG) ont été proposés comme alternative : les répondants indiquent leur positivité et leur négativité sur une grille qui n'impose pas de relation réciproque. Cependant, relativement peu d'études se sont concentrées sur les items ESG. Cette présentation rend compte d'un ensemble d'études préliminaires qui cherchent à décrire les propriétés psychométriques des items ESG et à documenter les preuves de leur adéquation (ou non) à la mesure des concepts EVT. Les données ont été recueillies auprès d'étudiants en introduction aux statistiques et d'un groupe de participants général.