Using Balanced Scales to Address Acquiescent Response Style

by

Maria Fernanda Alvarado Leiton

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey and Data Science)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Rachel E. Davis, Co-Chair, University of South Carolina
Research Associate Professor Sunghee Lee, Co-Chair
Research Associate Professor Brady T. West
Assistant Research Professor Ting Yan, University of Maryland

Maria Fernanda Alvarado Leiton

mleiton@umich.edu

ORCID iD: 0000-0002-8327-9320

Dedication

A mis papás, Luis y Felicia, por hacerme quien soy.

Acknowledgements

The completion of this dissertation would not have been possible without the extraordinary guidance of my co-chairs Sunghee Lee and Rachel Davis. They both have been truly devoted to this project and I am deeply grateful for their dedication and mentoring. They always made time for our weekly meetings, were prompt in providing feedback, and fostered a very positive work environment. I could not have asked for a more perfect team to guide me through this process. I would also like to thank the other members of my committee, Ting Yan and Brady T. West, for their support and valuable contributions, not only on this dissertation but throughout my time in the program.

I am also deeply grateful for the generous financial support I received from the Likert Fund, the Rackham Graduate Student Research Grant and the Universidad de Costa Rica to conduct the data collection I required for this dissertation. I would also like to thank the financial support I received from the Michigan Program in Survey and Data Science, the Daniel Katz Fellowship, and the Rackham Graduate school for the One-Term Dissertation Fellowship. Their generous support helped me focus on my dissertation work and bring it to an end.

I would also like to thank the program's faculty for sharing their knowledge with me, and for supporting me when I needed it. In particular, I would like to thank Fred Conrad for his unconditional support during my first year in the program. Many thanks to the program's staff too, who helped me successfully navigate many administrative procedures.

Finally, I would like to thank my friends and family. Thanks to my mom and dad who always believed in me and encouraged me to pursue my doctoral studies. To my husband Osvaldo, for being my rock, and the most patient and supportive person in the world. I love you. Many thanks to my brother, for his constant support, and for planning his wedding around my academic year so I could attend. And to my grandma, who’s wonderful cooking allowed me to focus for the past couple of months on writing and not cooking. To all my wonderful friends, thank you for all your support and for all the joy you bring into my life.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Measurement scales are widely used for collecting survey data about latent constructs in the social sciences. These scales are composed of multiple items that measure a single latent construct through rating response scales. Ideally, higher scores derived from ratings of these items indicate higher locations on a continuum of the latent construct. Nonetheless, errors stemming from how respondents choose their responses may complicate this measurement and lead to erroneous conclusions. Rating response scales are particularly vulnerable to Acquiescent Response Style (ARS), respondents' tendency to choose "agree" responses regardless of the content of the items.

Even though ARS has been studied for over half a century, there is still little agreement on how to address it. Balanced scales, formed by mixing items written in opposite directions of a given latent construct, are a well-known method used to measure and correct for ARS. However, concerns have been raised about the measurement properties of balanced scales, making their use controversial.

The goal of this dissertation was to provide an in-depth insight into the capability of balanced scales to not only measure ARS but also to correct for it. For this goal, this dissertation combined three studies. The first study investigated the effects of scale balancing under ARS on construct and convergent validity, reliability, and factor structure. The second study compared statistical methods to correct for ARS in computing scores of latent constructs using balanced scales. The third study empirically examined the differences in measurement properties of two wording strategies for drafting reverse-worded items for balanced scales.

Findings from this research suggest that scale balancing alone is insufficient to mitigate ARS-associated error and that statistical correction methods also need to be applied. However, these findings also imply that simple correction methods, such as Ordinary Least Squares regression and Confirmatory Factor Analysis, that use balanced scales may reduce the effects of ARS on scale scores. Furthermore, this study indicates that wording strategies used to generate balanced scales resulted in similar measurement properties. While the best practice for balanced scales design is still to be confirmed, this dissertation suggests that balanced scales may be a useful tool to control for ARS in surveys.

Chapter 1 Introduction

Measurement scales, such as Figure 1.1, are widely used for collecting data about latent variables in social science. Measurement scales are composed of multiple verbal statements or items that aim to measure a single latent construct. Respondents rate each item on a response scale with multiple categories, which are typically the same across all items. Responses to individual items are later combined to create an overall scale score that estimates the location of an individual on the latent construct continuum. In the scale in Figure 1.1, authored by Diener and colleagues (Diener, Emmons, Larsen, & Griffin, 1985), five items are aiming to measure different aspects of satisfaction with life; these items comprise the measurement scale. Each item is rated using a Disagree /Agree (D/A) response scale, and the combination of the item scores provide a measure of an individual's satisfaction with life.

Figure 1.1 Satisfaction with Life Scale (Diener et al, 1985)

Please say how much you agree or disagree with the following statements. (Mark (X) one box for each line.)

	Strongly disagree	Some what disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Some what agree	Strongly agree
In most ways my life is close to ideal.	☐	☐	☐	☐	☐	☐	☐
The conditions of my life are excellent.	☐	☐	☐	☐	☐	☐	☐
I am satisfied with my life.	☐	☐	☐	☐	☐	☐	☐
So far, I have gotten the important things I want in life.	☐	☐	☐	☐	☐	☐	☐
If I could live my life again, I would change almost nothing.	☐	☐	☐	☐	☐	☐	☐

Ideally, higher scale scores are an indication of a higher location on the latent variable continuum. Nonetheless, measurement errors can prevent this from happening. Rating response scales are particularly vulnerable to Acquiescent Response Style (ARS). ARS is the tendency of agreeing to items regardless of their content (Baumgartner & Steenkamp, 2001) and has been a vexing measurement problem for decades.

ARS first gained relevance during the second half of the twentieth century, as many researchers worried that commonly used measurement scales were being affected by this response style. One of the first scales to undergo this scrutiny was the California F scale of authoritarianism that was later determined to measure ARS instead of authoritarianism (Chapman & Campbell, 1959). Since then, researchers have cautioned about biases in score mean estimation (Van Vaerenbergh & Thomas, 2013), incoherence in associations with relevant variables (Danner, Aichholzer, & Rammstedt, 2015), incoherence in factor analysis structure (Rammstedt & Farmer, 2013; Rammstedt, Kemper, & Borg, 2013), subpopulation comparison inadequacy (Baron-Epel, Kaplan, Weinstein, & Green, 2010; Reynolds & Smith, 2010), and inadequacy of inferential statistical tests (Van Vaerenbergh & Thomas, 2013).

Even though the ARS problem has been studied for over half a century, there is still little agreement on how to address it. One controversial proposition has been the use of balanced scales to mitigate and/or measure ARS (Baumgartner & Steenkamp, 2001; Mirowsky & Ross, 1991; Weijters, Baumgartner, & Schillewaert, 2013). This type of scale is formed by writing items in opposite directions of the latent construct. The measurement scale in Figure 1.2 (Ryff, 1989) serves as an illustration of a balanced scale, as agreement to the first, third, and seventh items implies a higher sense of purpose in life, while agreement to the second, fourth, fifth, and sixth items implies the opposite.

Figure 1.2 Purpose in Life Scale (Ryff, 1989)

Please say how much you agree or disagree with each of the following statements. (Mark (X) one box for each line.)

	Strongly disagree	Some what disagree	Slightly disagree	Slightly agree	Some what agree	Strongly agree
I enjoy making plans for the future and working to make them a reality.	☐	☐	☐	☐	☐	☐
My daily activities often seem trivial and unimportant to me.	☐	☐	☐	☐	☐	☐
I am an active person in carrying out the plans I set for myself.	☐	☐	☐	☐	☐	☐
I don't have a good sense of what it is I'm trying to accomplish in life.	☐	☐	☐	☐	☐	☐
I sometimes feel as if I've done all there is to do in life.	☐	☐	☐	☐	☐	☐
I live life one day at a time and don't really think about the future.	☐	☐	☐	☐	☐	☐
I have a sense of direction and purpose in my life.	☐	☐	☐	☐	☐	☐

Balanced scales have become a popular way to measure and correct for ARS (e.g., Paulhus, 1991; Paulhus & Vazire, 2007; Rammstedt, Danner & Bosnjak, 2017). However, concerns have been raised about their suboptimal measurement properties. Specifically, concerns about scale reliability have been raised, as Cronbach's alpha tends to decline when balanced scales are used (Barnette, 2000; Stewart & Frye, 2004; Roszkowski & Soven, 2010; Solís-Salazar, 2015). Furthermore, for Confirmatory Factor Analysis (CFA), the factor structure is also threatened. It has been observed that one-dimensionality can be lost when using balanced scales (Pilotte & Gable, 1990; Williams, & Williams, 1996; Spector, Van Katwyk, Brannick, & Chen, 1997; DiStefano & Molt, 2006; Magazine, Marsh, 1996). Moreover, it has been reported that to achieve best fit, a two-factor model (where one factor is for the non-reversed and the other is for the reversed items) should be implemented (Solís-Salazar, 2015).

Because of the mixed views around balanced scales, the goal of this dissertation is to provide an in-depth insight into the capability of balanced scales to measure and correct for ARS.

To do this, the dissertation addresses the three following research questions:

1. Does the use of balanced scales mitigate the detrimental effects of ARS on construct measurement?

2. If balancing scales does not sufficiently remove ARS-associated measurement error, is there an optimal statistical method to correct for ARS when scales are balanced?

3. Is there a way to write items for balanced scales to reduce ARS while preserving measurement properties?

To answer these research questions, Chapter 2 begins with a simulation study to examine the effectiveness of balanced scales to counteract ARS. I use a Graded Response Model (GRM) to generate a 6-item-5-point scale unbalanced scale and its balanced counterpart. A total of 10,000 datasets of 5,000 cases each are generated. For each replicate sample, 20% of the cases were set to be affected by ARS by fixing the difficulty parameters of the GRM model, making responses on the high end of the rating scale more likely to occur, where a high response means an agreement. The ARS-free scores for this 20% of the sample are also created to serve as the true values of the study. The simulation contains nine ARS scenarios created by varying the correlation between ARS and a validation variable correlated to the latent construct used to generate the scale.

Once the data are generated, I examine four measurement properties: construct validity, convergent validity, reliability and factor structure of the measurement scale. To assess construct validity, I use the correlation of scores and the latent construct, and the mean of the sum scale scores. Because the items were generated from the latent variable, scores are a good indicator of the latent construct in this simulation. The scale scores are obtained by adding the scores of each item of the scale after reverse coding when needed. To assess convergent validity, I use the correlation between the sum scale scores and a validation variable for the latent construct. For reliability, I use Cronbach's alpha and the Greatest Lower Bound (GLB) as indicators. Lastly, I use fit indices for a one-dimensional Confirmatory Factor Analysis (CFA) model to assess factor structure. I compare the distribution of the aforementioned indicators of these measurement properties with their ARS-free distribution counterpart.

Chapter 3 turns to the second research question, which grapples with which type of statistical adjustment is more effective to reduce ARS bias in scale scores and measures derived from these scores. To answer this question, I turn again to a simulation study. This simulation study has a similar set up to the study described for Chapter 2 but includes two measurement scales, each consisting of 6 items using a 5-point response scale in order to have a larger pool of items to compute a measure of ARS. I examine four methods of ARS correction: (1) Ordinary Least Squares regression (OLS; Bachman & O'Malley, 1984; Liu, Suzer-Gurtekin, Keusch & Lee, 2019), (2) CFA with two substantive factors and a response style factor (Billiet & McClendon, 2000), (3) CFA with one substantive factor and a response style factor (Savalei & Falk, 2014) and (4) Multidimensional Nominal Response Modeling (MNRM; Bolt & Johnson, 2009). I evaluate these methods through (1) the correlation of the corrected scores with the true latent scores, and (2) the correlation between the corrected scores and a validation variable versus the same correlation computed using the ARS-free scores.

Chapter 4 focuses on whether there is an optimal way to write reversed items to create a balanced scale for the purposes of mitigating ARS error. Data was collected for this dissertation and included a wording experiment in which respondents were randomly assigned to one of two types of balanced scales or a control group with unbalanced scales. The first type of balanced scale was formed by using negation particles (e.g., "no" or "not") in the original statement (e.g., “In most ways my life is not close to my ideal”). The second type was formed by using a polar opposite term to induce the reversal (e.g., “In most ways my life is far from my ideal”). Participants were recruited from three populations with different ARS tendencies: non-Hispanic White respondents in the United States, Hispanic respondents in Mexico interviewed in Spanish and Hispanic respondents in the United States interviewed in English. Four well-established scales, originally unbalanced, were included in the experiment: Satisfaction with Life (Diener et al, 1985), Sense of Control (Lachman & Weaver, 1998), Need for Affect (Maio & Esses, 2001) and Social Provisions (Ipachino et al., 2016). I compare three outcomes for the two types of balanced scales: (1) Cronbach’s alpha coefficient, (2) CFA fit measures, and (3) convergent validity. The first two of these measures assessed the measurement properties of the scales that have been heavily criticized about balanced scales. Because in this study there are no “true” scale scores to directly assess construct validity, convergent validity was assessed through correlations of scale scores with other constructs.

In the last chapter of this dissertation, I discuss the main results across the three studies and suggest directions for future work in this area.

References

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491-509.

Barnette, J. J. (2000). Reversals on Survey Internal Consistency. *Educational and Psychological Measurement*, *60*(3), 361–370.

Billiet, J. B., & McClendon, M. J. (2000). *Structural Equation Modeling : A Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items*. *5511*(September 2014), 37–41.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European Journal of Public Health*, *20*(5), 543–548.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Bolt, D. M., & Johnson, T. R. (2009). Due to Individual Differences in Response Style. *Applied Psychological Measurement*, (2001), 335–352.

Chapman, L., & Campbell, D. (1959). The Effect of Acquiescence Response-Set Upon Relationships Among the F Scale , Ethnocentrism , and Intelligence. *Sociometry*, *22*(2), 153–161.

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75.

DiStefano, C., & Molt, R. (2006). Program : Measuring General Versus Specific Gains. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(3), 440–464.

Iapichino, E., Rucci, P., Corbani, I. E., Apter, G., Bollani, M. Q., Cauli, G., ... & Bassi, M. (2016). Development and validation of an abridged version of the Social Provisions Scale (SPS-10) in Italian. *J Psychopathol*, *22*, 157-163.

Lachman, M. E., & Weaver, S. L. (1998). The sense of control as a moderator of social class differences in health and well-being. *Journal of personality and social psychology*, *74*(3), 763.

Liu, M., Suzer-Gurtekin, T., Keusch, F., & Lee, S. (2019). Response Styles in Cross-cultural Surveys. In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods* (pp. 477–500). New Jersey: Wiley.

Magazine, S., Williams, L., & Williams, M. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's affective and continuance commitment scales. *Educational and Psychological Measurement*, *56*(2), 241–250.

Maio, G. R., & Esses, V. M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of personality*, *69*(4), 583-614.

Marsh, H. W. (1996). Positive and Negative Global Self-Esteem: A Substantively Meaningful Distinction or Artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810–819.

Mirowsky, J., & Ross, C. E. (1991). Eliminating Defense and Agreement Bias from Measures of the Sense of Control : A 2 X 2 Index. *Social Psychology Quaterly*, *54*(2), 127–145.

Paulhus, D. L. (1991). *Measurement and control of response bias.* In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures of personality and social psychological attitudes* (p. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, *1*(2007), 224-239.

Pilotte, W., & Gable, R. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*, 603–610.

Rammstedt, B., Danner, D., & Bosnjak, M. (2017). Acquiescence response styles: A multilevel model explaining individual-level and country-level differences. *Personality and Individual Differences*, *107*, 190-194.

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145.

Rammstedt, B., Kemper, C., & Borg, I. (2010). Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study. *European Journal of Personality*, *27*, 71–81.

Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*(2), 230–243.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment and Evaluation in Higher Education*, *35*(1), 117–134.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, *57*(6), 1069.

Savalei, V., & Falk, C. F. (2014). Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches. *Multivariate Behavioral Research*, *49*(5), 407–424.

Solís-Salazar, M. (2015). El dilema de combinar ítems positivos y negativos en escalas. *Psicothema*, *27*(2), 192–199.

Spector, P. E., Van Katwyk, P. T., Brannick, M. T., & Chen, P. Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659–677.

Stewart, T. J., & Frye, A. W. (2004). Investigating the use of negatively phrased survey items in medical education settings: Common wisdom or common mistake? *Academic Medicine*, *79*(10 SUPPL.), 18–20.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195-217.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: An integrative model. *Psychological Methods*, *18*(3), 320–334.

Chapter 2 Measurement Properties of Balanced Scales in the Presence of ARS

2.1 Abstract

For measurement scales based on multiple items, creating balanced scales by mixing items worded in opposite directions of a latent construct has been recommended as a solution to reducing errors stemming from Acquiescent Response Style (ARS). This study explored the impact of ARS in balanced scales relative to unbalanced scales and the ability of balanced scales to mitigate ARS. To this end, a simulation study was conducted, in which two versions of 6-item measurement scale were generated through a Graded Response Model (GRM): (1) an unbalanced scale with all items measuring the construct in the same direction, and (2) a balanced scale reversing the direction of three out of the six items. For a sample of the generated cases, ARS was introduced to the scale items by modifying difficulty and discrimination parameters of the GRM. These cases were selected based on an ARS latent construct. This study examined the effects of ARS on four measurement properties: construct validity, convergent validity, reliability, and factor structure. To assess construct validity, scale scores that summed item scores were correlated with the true latent construct, and the score mean was also examined. For convergent validity, the correlation between the sum scale scores and a validation variable was examined. To assess reliability, two coefficients were used: Cronbach's alpha and the Greatest Lower Bound (GLB). Factor structure was assessed by the examination of fit indices of a one-dimensional CFA model. Results showed that reversing items was useful for reducing bias in scale score means but not sufficient to mitigate the effects of ARS on correlations. Furthermore,

the results indicated that unbalanced scales were not necessarily better than balanced scales in construct or convergent validity when ARS was present. Unbalanced scales had higher reliability and better fit for a one-dimensional CFA model than balanced scales, but these properties appeared to be artifacts of ARS, and not an indication of superior measurement. This study underlines the need for further steps, such as statistical correction models, to control for ARS in survey data obtained from balanced scales.

2.2 Introduction

Acquiescent Response Style (ARS) is the systematic tendency to endorse agreement categories in Agree/Disagree (A/D) rating scales. The impact of ARS can lead to a loss of scale validity (Cronbach, 1946), biases in score means (Van Vaerenbergh & Thomas, 2013), incoherent associations with relevant variables (Danner, Aichholzer, & Rammstedt, 2015), incoherent factor analysis structures (Rammstedt, Kemper, & Borg, 2012; Rammstedt & Farmer, 2013), inadequate population subgroup comparisons (Baron-Epel, Kaplan, Weinstein, & Green, 2010; Reynolds & Smith, 2010), and inadequate inferences (Van Vaerenbergh & Thomas, 2013). Therefore, for anyone using survey data, ARS poses a direct threat to data quality and limits the usability of measurement scales.

The cause of ARS is debated. Multiple theories explaining its occurrence have been proposed (Cronbach, 1946; Krosnick, 1991; Knowles & Condon, 1999). Recently, Lechner et al. (2019) unified these theories regarding the causes of ARS. Under this conceptualization of acquiescent behavior, differences in cognitive processing capacities and communication styles are identified as the mechanisms through which ARS operates, creating differences in acquiescence at the respondent, situational, and cultural levels (Lechner et al., 2019). At the respondent level, individuals who do not, or are unable to, allocate adequate mental resources to the survey responding process are believed to be more likely to exhibit acquiescent behaviors (Krosnick, 1991). At the situational level, survey characteristics such as administration by more experienced interviewers, telephone survey mode (compared to online and face-to-face surveys), and distractions during the interview (which produce cognitive burden to respondents) have shown to foster ARS (Knowles & Condon, 1999; Weijters, Schillewaert & Geuens, 2008; Olson & Bilgen, 2011). At the cultural level, social deference and cultural orientation towards collectivism may foster acquiescence (Johnson, Kulesa, Cho & Shavitt, 2005; Harzing, 2006; Rammstedt, Danner, Bonsjak; 2017), although at least one more recent study suggests that deference is not a determinant of acquiescent behavior (Davis, Johnson, Lee and Werner, 2019).

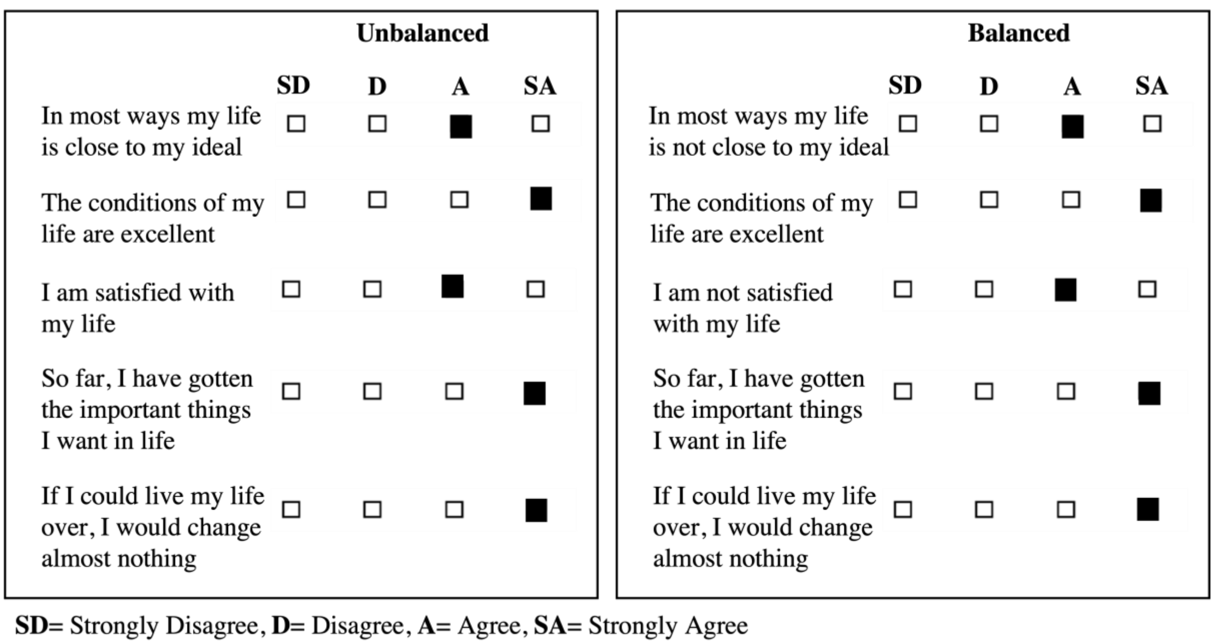
Balancing measurement scales by mixing items worded in opposite directions of the latent trait has been proposed as a correction for ARS in scale scores (Mirowsky & Ross, 1991; Baumgartner & Steenkamp, 2001; Weijters, Baumgartner, & Schillewaert, 2013). However, scale balancing remains controversial. Despite some reports of adequate measurement outcomes while using balanced scales (Winkler, Kanouse & Ware, 1981; Martinez-Molina & Arias, 2018), there is still resistance to their use (Benson & Hocevar, 1985; Eys, Carron, Bray & Brawley, 2007; Menold, 2020). Researchers have consistently reported the loss of one-dimensionality in factor analysis when using balanced scales, describing the emergence of a two-factor structure when there should be a theoretically logical one-factor solution. This rupture of one-dimensionality often occurs in one of two forms: (1) a two-factor solution, where each factor represents the direction of its corresponding items (e.g., Benson & Hocevar, 1985; Gnambs & Schroeders, 2017); or (2) a two-factor solution including a substantive factor and a method factor loading on all items (e.g., Menold, 2020). Another commonly reported effect of scale balancing is the shrinkage of reliability coefficients, usually Chronbach's alpha, and this result has been interpreted a direct impact on measurement reliability and measurement quality (e.g., Schriesheim, Eisenbach & Hill, 1991; Barnette, 2000; Eys, Carron, Bray & Brawley, 2007, Roszkowski & Soven, 2010). Even though unbalanced scales have consistently yielded more adequate measurement properties, it has been argued that this can easily be the result of response biases (Weijters & Baumgartner, 2012)

Furthermore, it is unclear how balanced scales help to correct for ARS. For example, it is unknown if balanced scales mitigate the effects of ARS on their own, or if further steps of correction, like statistical adjustments, are needed. There is mixed evidence around this issue, with some studies claiming that balanced scales fully mitigate ARS bias (Cloud & Vaugham, 1970; Primi, Haulk-Filho, Valentini & Santos, 2020), some claiming that they only correct bias for certain survey estimates (Savalei & Falk, 2014), and some developing methods to correct for ARS based on the use of balanced scales (Billiet & McClendon, 2000; Savalei & Falk, 2014). If balanced scales were to fully mitigate ARS, bias would be cancelled out when items are combined into a scale score, even though individual item responses would be influenced by ARS, because of the reverse-keying necessary to compute the scores. In this scenario, the reverse-keying of items is thought to reduce any excesses of agreeable responses, as exemplified in Figure 2.1, in which reverse-keying of the first and third items for the balanced scale decreases the number of agreeable responses from five to three.

In contrast, for the unbalanced scale in the figure, no reverse-keying is needed, and ARS remains confounded with the latent construct in the scale score. However, for balanced scales to cancel out ARS, two important assumptions need to be in place. First, ARS should impact all items in the scale equally; in particular, reversed and non-reversed items should be similarly affected. Second, the reversed versions of items should be equivalent in terms of measurement to their non-reversed counterpart. This means they should measure the same level of the latent construct (equal item difficulty) and linguistically carry the same meaning.

In survey practice, balanced scales could be a powerful tool to address ARS, if they were to mitigate the effects of ARS by themselves, while retaining adequate measurement. However, there is still lack of consensus regarding the measurement properties of balanced scales (relative to unbalanced scales), and their ability to mitigate ARS-related bias in surveys. To address this, this research compares the effects of ARS in balanced and unbalanced scales on four common measurement properties: construct validity, convergent validity, reliability, and factor structure. To evaluate these properties, a simulation study was conducted, allowing to isolate ARS, often confounded in real survey data with other measurement error sources (e.g., other response styles, social desirability, etc.), which limits the understanding of the functioning of balanced scales to address ARS.

Figure 2.1 Illustration of Acquiescent Response Style (ARS) for balanced and unbalanced scales



This exploratory simulation study addresses a research question that has not been previously addressed in the ARS literature: in comparison to unbalanced scales, do balanced scales mitigate the effects of ARS by improving construct and convergent validity without sacrificing other measurement properties like factor structure and reliability? Findings from this study will provide empirically based guidance to survey researchers and practitioners about the measurement properties of balanced scales and their ability to mitigate ARS.

2.3 Methods

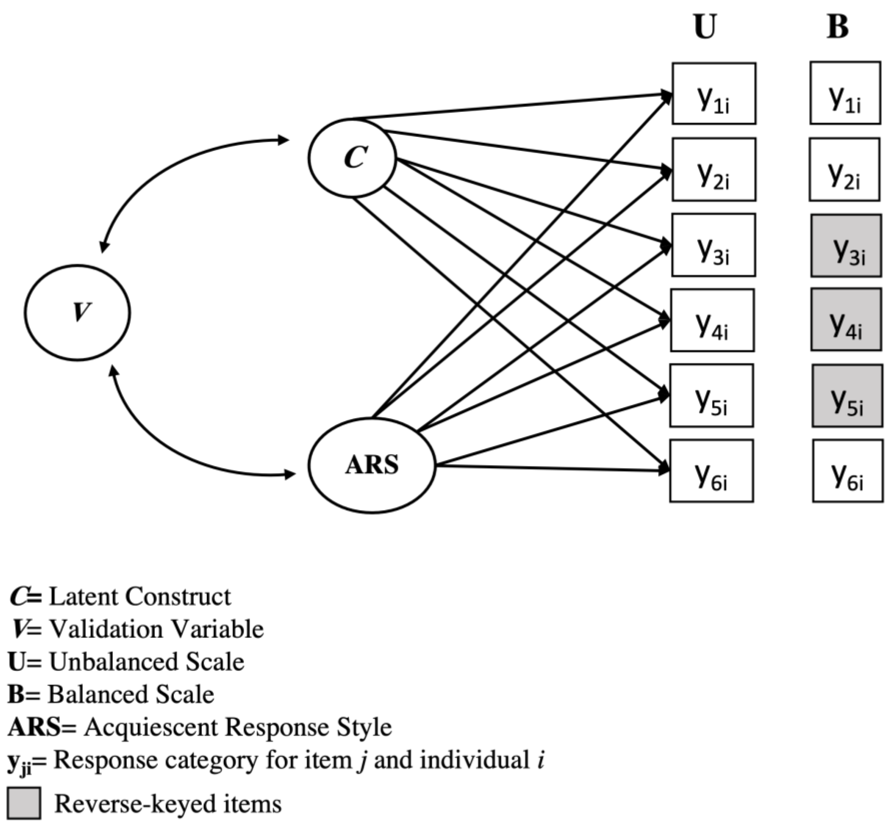
This research used a simulation study designed to compare balanced and unbalanced scales in the presence of ARS. This simulation setup was based on three parts: 1) a continuous latent construct of interest (C) measured through six items, where each item had five response options; 2) a continuous validation measure of C (V); and 3) an ARS construct. Under a structural equation modeling framework, the relationships between these constructs in the simulation setup were conceptualized as illustrated in Figure 2.2. The C and V latent constructs were drawn from a Multivariate Normal Distribution, $\binom{C}{V} \sim N_2 \left(\mu = \binom{0}{0}, \Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}\right)$. The ARS construct was drawn separately from a Standard Normal Distribution, $N(0,1)$, and later linked to the distribution of V using a Gaussian copula. This linking consisted of two steps. The first step was to adjust the target correlation between ARS and the validation variable (r) using the formula $r_{adjusted} = 2 \times \sin\left(\frac{r \times \pi}{6}\right)$. The second step was to multiply the normally distributed ARS construct by the Cholesky decomposition of a 2x2 triangular matrix of $r_{adjusted}$ with 1s in the diagonal. This step produced the normally distributed ARS construct correlated with V with a Pearson coefficient of r. Finally, the ARS construct was transformed into a Uniform Distribution, $U(0,1)$, using the *pnorm* command from the *stats* package in R.

The association structure between the three constructs in the simulation was an effort to emulate how ARS interacts with other variables in real surveys, as there are personal attributes, question features, and specific contexts that explain a person's tendency to acquiesce, like demographic variables such as education and age (e.g., Stukovsky, Palat, & Sedlakova, 1982; Weijters, Cabooter & Schillewaert, 2010). Therefore, in this simulation, it is assumed that the validation variable is, in fact, one of these ARS predictors (or another variable related to ARS).

Because V was the validation measure of C, these two variables were set to have a Pearson correlation of 0.7. Nine versions of the ARS construct were generated to evaluate different levels of the association between ARS and the validation variable V. These many and diverse scenarios of ARS were created in order to investigate how the interaction between ARS and other survey variables could impact its effects on the measurement properties of scales and the possibility to correct for ARS. The corresponding Pearson correlations between ARS and V were: 0.00, 0.05, 0.25, 0.50, 0.98, -0.05, -0.25, -0.50 and -0.98. In real surveys, these correlations come from the correlates of ARS, such as education and age (Stukovsky, Palat, & Sedlakova, 1982; Meisenberg & Williams, 2008; Weijters, Cabooter & Schillewaert, 2010). It is unlikely that ARS would be as strongly associated with any variable in real survey data, as in some of the most extreme scenarios studied here (e.g., r=0.98); however, this range of variation allows to better grasp the impacts of ARS in balanced and unbalanced scales.

The data generating process for the setup in Figure 2.2 comprised three steps. First, item-like data was generated using a Graded Response Model (GRM). Second, ARS was introduced to the items of a sample of simulated cases based on their position on the ARS continuum. Finally, scores were computed for each case, and the measurement properties of the scale were assessed. The details of each step are described next and the syntax for the data generation in R is in Appendix A1.

Figure 2.2 Illustration of the setup for the simulation study



2.3.1 *Step 1. Generating data using the GRM*

The GRM introduced by Samejima (1968) was utilized to generate the items measuring C. Following this model, C_i represents the latent trait of interest for subject i and $C_i \sim N(0,1)$. There are J items that measure C through a response scale of $k = 1, 2\ ,3, \ldots ,5$ ordered categories. For each item, y_{ij}, the realization of the measurement of item j for subject i, is observed. Then, the probability that a respondent with construct level C_i will endorse category k or higher for item j can be expressed as

$$P\big(y_{ji} \geq k \mid C_i\big) = \frac{\exp\big[a_j\big(C_i - b_{kj}\big)\big]}{1 + \exp\,[a_j(C_i - b_{kj})} \tag{2.1}$$

where a_j is the discrimination parameter for each item, and b_{kj} is the difficulty threshold for response category k for item j. These parameters were randomly selected following the approach of Jiang, Wang, & Weiss (2016), which prevents the occurrence of item distributions with too small or too large case counts for some of the categories. The realization for each item was generated using the *simIrt* function from the catIrt package (Nydick, 2014) in R, for a sample size of 5,000 cases. The *simIrt* function utilizes the GRM parameters $(a_{j,}\, b_{kj})$ and a vector for the latent construct (C) as input, and returns the discrete response categories. The function produces an unbalanced version of the scale, in which higher response categories indicate higher levels of the construct. A balanced version of the scale was generated by recoding the directionality of three of the items as shown in Figure 2.2.

2.3.2 *Step 2. Introducing ARS to items*

ARS was introduced to the items by favoring the occurrence of categories on the higher end of the rating scale. For 20% of the sample, items were generated a second time using difficulty and discrimination parameters (a_j and b_{kj}) that yielded more frequently the higher categories of the rating scale. For these cases, items were not reverse-keyed when creating the balanced version of the scale in order to simulate misresponse due to ARS.

Selecting 20% of the cases (and not a different percent) to introduce ARS was an arbitrary decision, however, it was informed by the estimated prevalence of ARS reported by Vannette and Krosnick (2014). The selection of the cases which would reflect ARS was made based on the position of each case on the uniformly distributed ARS construct, in order to preserve the correlation between the ARS construct and the validation variable. As a consequence, for each scenario of the simulation, the 20% of cases with the highest values on the ARS construct was the selected sample to which ARS was introduced.

2.3.3 *Step 3. Computing scale scores*

In this study, scale scores were generated using the methodology proposed by Likert (1932), in which scale scores are computed as the summation of the response codes for all items in the scale for each case, and using reverse coding when required (Equation 2.2). Using this methodology, two types of scores were created. The first used the ARS-free items from Step 1 of the simulation for the computation of the scores. These scores serve as the benchmark to assess the impact of ARS throughout the paper, and due to the lack of measurement errors, these are equal for the balanced and unbalanced scale. These items are referred to as the "No ARS" scores for the remainder of the paper. The second type of score was computed using the data from Step 2 of the simulation, for which 20% of cases included ARS.

$$S_i = \sum_{j=1}^{J} y_{ij} \tag{2.2}$$

The entire simulation process was replicated 10,000 times. This means that this study examines 10,000 data sets of 5,000 cases each.

2.3.4 *Evaluation criteria*

Construct validity

To measure construct validity, the correlation between scale scores and C, $Cor(S, C)$, and the score means ($\bar{S}$) were computed as the outcomes of interest across replications. In this study, due to the lack of other measurement errors, and because of how data was generated, unbiased scores are a good measure of the underlying latent construct *C*. The correlation $Cor(S, C)$

examined how ARS affected the degree to which the scores provided a good estimation of the latent construct at the case level, while the score means served to evaluate how ARS affected the description of the latent construct at the sample level. Summary statistics of the distribution of these measures for the unbalanced and balanced scales were compared, and in the case of $\bar{S}$, the "No ARS" mean scores were also included in the comparison to serve as benchmark.

Convergent validity

In this study, convergent validity was evaluated through the correlation between scale scores for the balanced and unbalanced scales under ARS and the validation variable V. Summary statistics of the distribution of this correlation were compared for unbalanced, balanced, and "No ARS" scores.

Reliability

Here, two indicators of reliability were used: Cronbach's alpha (Cronbach, 1951) and the Greatest Lower Bound (GLB; Jackson & Agunwamba, 1977). Unarguably, Cronbach's alpha is the most used and well-known measure of reliability. However, because of its numerous limitations in the measurement of reliability, which includes its underestimation (see Sijtsma, 2009, for a discussion on this topic) the GLB was included in the study. It has been shown that the GLB works better than Cronbach's alpha in scenarios in which the normality of the underlying latent construct and tau-equivalence, both assumptions of Cronbach's alpha, are not met (Chakraborty, 2017). The latter assumption requires that all items measure the same underlying latent construct, on the same scale, and are equally correlated to that latent construct (Peters, 2014). The violation of the tau-equivalence assumption is said to be the reason for Cronbach's alpha's underestimation of reliability (Graham, 2006).

Factor structure

Factor structure was included in this study as loss of one-dimensionality in CFA models is one of the main criticisms against balanced scales. Therefore, four fit measures of a one-dimensional CFA model were examined: the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). For CFI and RMSEA, commonly recommended bounds of good fit (CFI>0.95; RMSEA<0.05) were used. For AIC and BIC, smaller values were considered as indication of better model fit. Therefore, the model with the smallest AIC and BIC was considered the best fitting model.

2.4 Results

2.4.1 *Construct validity*

Figure 2.3 shows the effects of ARS on the correlation between C and the scale scores for the unbalanced and balanced scales. In this graph, high correlations indicate good construct validity as the scores closely represent the latent construct C. The median of the correlation between the scores and C was lower for balanced scales across simulation scenarios, which indicates that these scores provide worse representation of C when compared to unbalanced scales. This was particularly true for the (positive and negative) extreme cases of $Cor(ARS, V)$, for which construct validity was the worst for balanced scales. It is important to mention that the error bars in the graph show substantial overlapping of the distributions of $Cor(S, C)$ for most of the simulation scenarios. This implies that even though there are differences in construct validity between balanced and unbalanced scales, these differences appear to be small.

Figure 2.3 Median, 2.5th percentile and 97.5th percentile of the correlation between scale scores (S) and the latent construct C for unbalanced and balanced scales under ARS

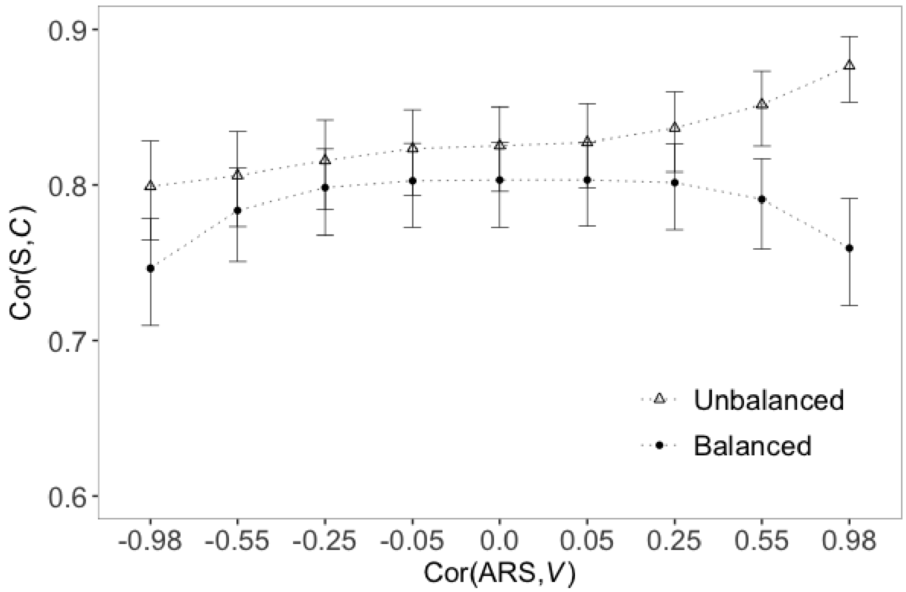
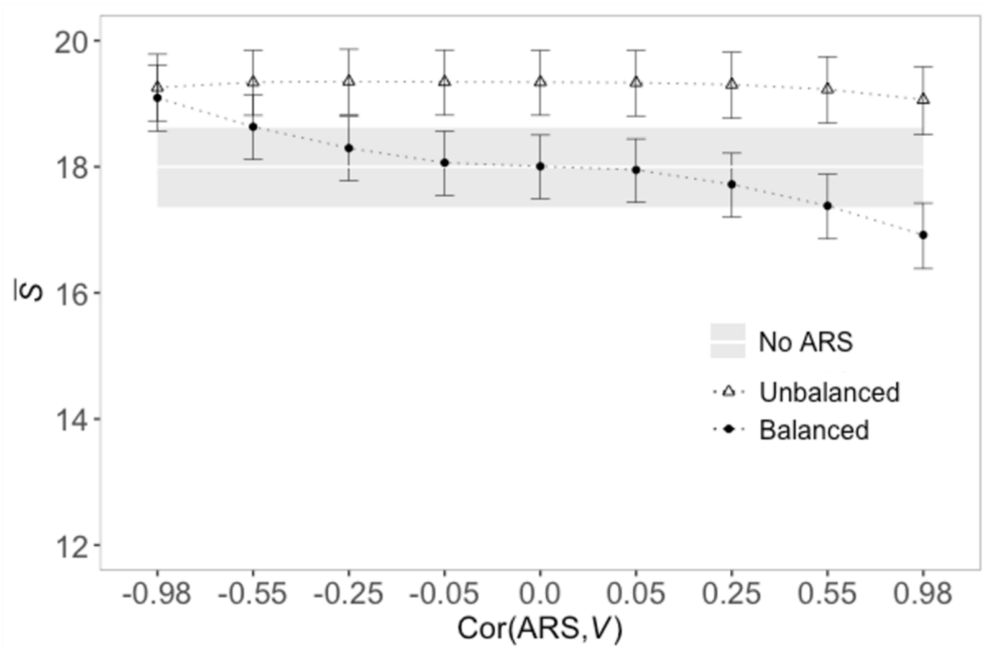


Figure 2.4 shows the median, 2.5th and 97.5th percentiles of scale score means for unbalanced and balanced scales and ARS-free items. In the figure, the "No ARS" line represents the median of the ARS-free score means and serves here as the benchmark. The results showed that bias affects balanced and unbalanced scales differently. For unbalanced scales, bias was somewhat constant and independent of the relationship between ARS and the validation variable *V*. Unbalanced scales produced overestimation of the mean of scale scores for all simulation scenarios. On the contrary, for unbalanced scales the correlation between *V* and ARS markedly changed the score means, their proximity to the ARS-free score means, and the direction of the bias (underestimation vs overestimation). However, for small and moderate correlations between *V* and ARS (ranging from -0.25 to 0.25), balanced scales mitigated most of the effects of ARS on score means, as these scenarios yielded the smallest shift in score means from the ARS-free benchmark.

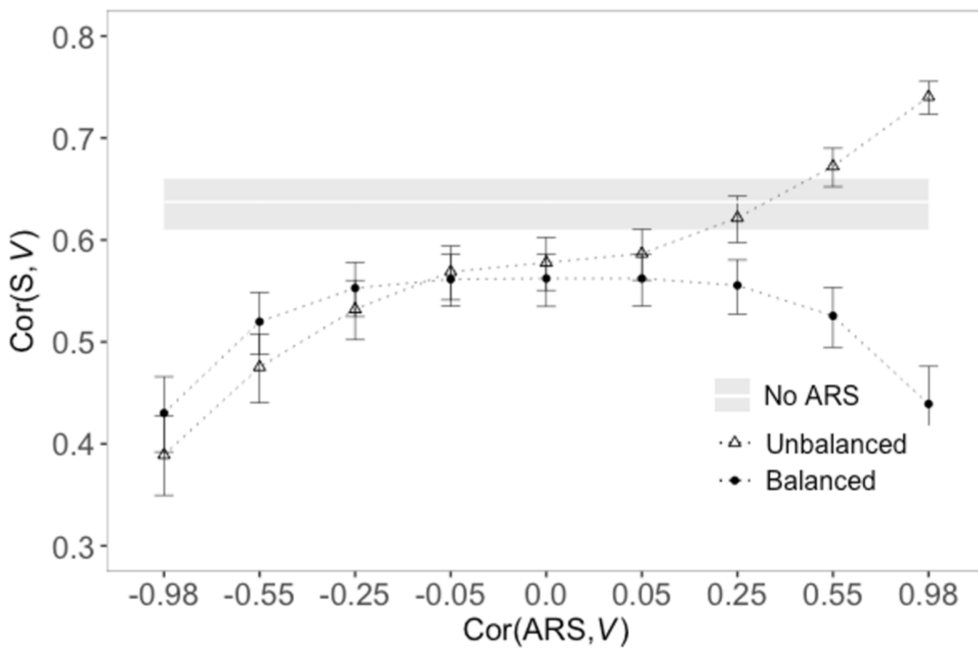
Figure 2.4. Median, 2.5th percentile and 97.5th percentile of scale score means ($\bar{S}$) for unbalanced and balanced scales under ARS and the ARS-free scales



2.4.2 *Convergent validity*

Figure 2.5 shows the median, 2.5th and 97.5th percentiles of the correlation between the validation variable *V* and the scale scores for balanced and unbalanced scales under ARS. The median of the ARS-free distribution of the correlation is represented in the figure under "No ARS" and serves as the benchmark. For both balanced and unbalanced scales, ARS led to the attenuation of the correlation between *V* and the scale scores for most of the simulation scenarios. Except for unbalanced scales under scenarios of moderate to high correlations between *V* and ARS, convergent validity decreased, regardless of scale balancing. However, loss of convergent validity was dependent on the correlation between *V* and ARS. Convergent validity deteriorated more for those scenarios where the correlation between *V* and ARS was high, regardless of its direction. The results also show that both unbalanced and balanced scales failed to preserve convergent validity, even for those scenarios for which the effects of ARS on score means were mitigated by scale balancing. In addition, Figure 2.5 indicates substantial overlap of the distributions of the correlation between scale scores and the validation variable for balanced and unbalanced scales in most simulation scenarios, implying small differences in convergent validity due to scale balancing.

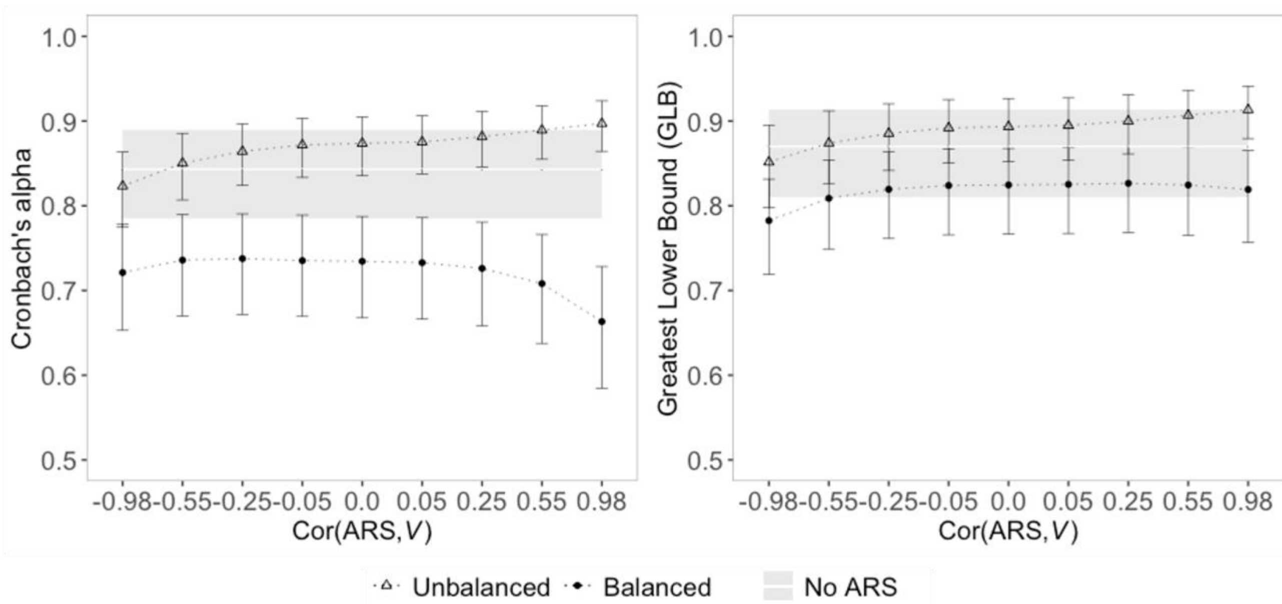
Figure 2.5 Median, 2.5th percentile and 97.5th percentile of the distribution of the correlation between validation variable (V) and scale scores (S) for unbalanced and balanced scales under ARS, and the ARS-free scale.



2.4.3 ***Reliability***

Figure 2.6 shows the results for Cronbach's alpha and Greatest Lower Bound (GLB) for balanced scales, unbalanced scales and ARS-free items. ARS produced changes in the distributions of both reliability indicators when compared to the ARS-free scenario; however, these changes were more evident for Cronbach's alpha. For unbalanced scales, reliability estimation using Cronbach's alpha and the GLB was inflated for most simulation scenarios. For balanced scales, ARS lowered reliability when compared to the ARS-free scale for both reliability indicators. Nonetheless, differences in the measurement of reliability between Cronbach's alpha and the GLB were evident. The GLB was more robust to the effects of ARS, as the distribution of the GLB overlapped considerably with the ARS-free values when compared to the ample decrease in reliability indicated by Cronbach's alpha for balanced scales.

Figure 2.6 Median, 2.5th percentile and 97.5th percentile of Cronbach's alpha and Greatest Lower Bound coefficients for unbalanced and balanced scales under ARS, and the ARS-free scale

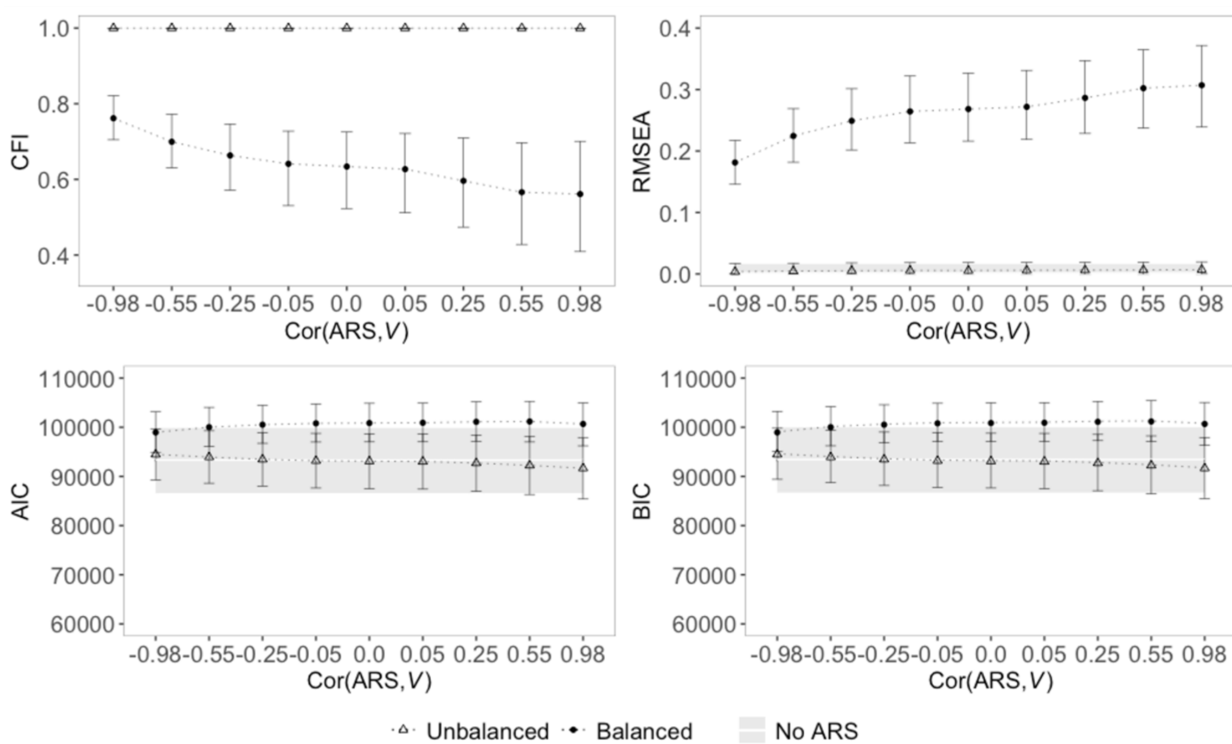


2.4.4 *Factor Structure*

Figure 2.7 presents the median, 2.5th and 97.5th percentiles of the distribution of fit indices of a one-dimensional CFA model for unbalanced and balanced scales under ARS. The same model specification was used to produce the fit indices of ARS-free items, and the median of these indices served as the benchmark values. None of the indices indicated better model fit for the balanced scales when compared to the unbalanced scales. It was clear that a model specification with only one factor did not fit the balanced scales adequately, as CFI and RMSEA values were appreciably far from their respective thresholds for good model fit. Although poor model fit is never desirable, in this case it was evidence of the effective identification of ARS as two latent factors (*C* and ARS) generated the data, making a one-factor solution unreasonable. This result also highlighted that the apparent good fit statistics of unbalanced scales were an artifact of ARS and denoted that they were unable to properly identify the two data generating constructs.

For balanced scales, CFI and RMSEA were sensitive to the correlation between the validation variable *V* and ARS, and the model fit was the worst for high, positive correlations. AIC and BIC also indicated that model fit worsened under ARS for balanced scales when compared to unbalanced scales, although the distributions of these indicators overlapped between balanced and unbalanced scales. These two fit indices were less sensitive to the correlation between *V* and ARS, as their values were relatively stable for all simulation scenarios.

Figure 2.7 Median, 2.5th percentile and 97.5th percentile of fit indices of the one-dimensional CFA model for unbalanced and balanced scales under ARS, and ARS-free scales



2.5 Discussion

This simulation study explored whether balanced scales mitigate ARS while yielding similar measurement properties to unbalanced scales. The study contributes to the discussion about the ability of balanced scales to mitigate the effects of ARS by isolating ARS (through a simulation study) and comparing its effects on measurement properties of balanced and unbalanced scales, which has been lacking in the extant literature about ARS. The results indicated that balanced scales reduced ARS-associated bias in scale score means; however, this mitigation of ARS did not extend to correlational analyses, where balanced scales yielded lower convergent validity than ARS-free items and worsened fit indices for a one-dimensional CFA model relative to unbalanced scales. ARS also impacted reliability estimation for balanced scales and resulted in reduced reliability coefficients when compared to the ARS-free scenario. Together, these results suggest that balanced scales are not sufficient to counteract the effects of ARS, particularly for correlational analyses.

Although there are caveats to the use of balanced scales, it is important to note that unbalanced scales are not necessarily a better option for measuring latent constructs. The main arguments surrounding the utilization of unbalanced scales are the preservation of one-dimensionality of CFA models, and higher reliability when compared to balanced scales. (Schriesheim, Eisenbach & Hill, 1991; Barnette, 2000; Gnambs & Schroeders, 2017; Menold, 2020). Although these effects were also observed in the current study, the results suggested that these seemingly good measurement properties of unbalanced scales were artifacts of ARS. Reliability coefficients, and in particular Cronbach’s alpha, were considerably inflated for unbalanced scales. This occurs because Cronbach’s alpha is based on interitem correlations, which are artificially increased as ARS affects items systematically in one direction. Given that

Cronbach's alpha is so widely used in the social sciences, this artificial inflation of alpha is particularly worrisome, as it can result in a false sense of confidence in the measurement of key variables. In contrast, the GLB coefficient was more robust to the effects of ARS, and the inflation of reliability for unbalanced scales and reduction for balanced scales was of lesser importance. These findings are consistent with previous studies noting the limitations of using Cronbach's alpha to measure reliability (Komorita & Graham, 1965; Streiner, 2003; Revelle & Zinbarg, 2009; Vaske, Beaman, & Sponarski, 2017) and that other coefficients, including the GLB, are more suitable options for this purpose (Sijtsma, 2009; Peters, 2014).

For CFA, results from this study suggest that good fit indices of unbalanced scales are an artifact of ARS, as these scales yielded good fit for a one-dimensional model, when there are two latent constructs in place, showing the inability of the unbalanced scales to separate the substantive and style factors. In contrast, loss of one-dimensionality was evident for balanced scales, with CFI and RMSEA indices indicating poor fit, which was consistent with previous studies (Benson & Hocevar, 1985; Gnambs & Schroeders, 2017; Menold, 2020). This result is usually considered as evidence in support of the use of unbalanced scales, however, here it is interpreted as evidence in favor of the use of balanced scales. The poor model fit of balanced scales indicates that *there is* identification of the two latent constructs that are effectively influencing the items (ARS and the substantive construct), which are confounded when unbalanced scales are used. Thus, including a style factor when modeling balanced scales should be favored as a way to control for a measurement error that can be present but is often ignored.

Although this research was able to explore in depth the efficacy of balanced scales to correct for ARS relative to balanced scales, it is not without limitations. Most noticeably, the data generating process utilized here could be different from what occurs in real surveys, where the prevalence of ARS may be higher or lower, depending on the sample. Furthermore, only one source of measurement error is considered here. Although this was necessary to isolate the effects of ARS, it is unclear if other sources of error, for example, other response styles, impact the use of balanced scales. Finally, some of the simulation scenarios yielded better results for balanced scales than others. It is unknown how frequent each of these scenarios are in real life, and it is also challenging to assess which one survey data resembles.

In conclusion, this study indicates that the implementation of balanced scales as a method for addressing ARS was helpful but insufficient. Although there was correction of scale score means, correlational analyses were still affected by acquiescent responding. Nonetheless, the evidence presented here does not indicate that the use of balanced scales should be dismissed; it only shows how balanced scales (similarly to unbalanced scales) are affected by ARS. In fact, there was some indication of potential benefits from using balanced scales. A one-dimensional CFA model was not adequate, suggesting that ARS is, in fact, separated from the substantive construct. Further research is needed to explore possible paths that would allow taking advantage of this apparent separation of response style and substantive construct for ARS correction. For survey practitioners, the results from this study imply that it is advantageous to use balanced scales to mitigate the effects of ARS, particularly for reducing bias in scale score means. Furthermore, this research highlights how researchers should be wary when unbalanced scales produce an excellent model fit, as it might be an artifact of ARS.

References

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, *48*(2), 491-509.

Barnette, J. J. (2000). Reversals on Survey Internal Consistency. *Educational and Psychological Measurement*, *60*(3), 361–370.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European Journal of Public Health*, *20*(5), 543–548.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Baumgartner, H., Weijters, B., & Pieters, R. (2018). Misresponse to Survey Questions: A Conceptual Framework and Empirical Test of the Effects of Reversals, Negations, and Polar Opposite Core Concepts. *Journal of Marketing Research*, *55*(6), 869–883.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural equation modeling*, *7*(4), 608-628

Bolt, D. M., & Johnson, T. P. (2009). Addressing Score Bias and Differential Item Functioning Due to Individual Differences in Response Style. *Applied Psychological Measurement*, (2001), 335–352.

Chakraborty, A. P. R. (2017). Estimation of Greatest Lower Bound Reliability of Academic Delay of Gratification Scale. *IOSR Journal of Research & Method in Education (IOSRJRME)*, *07*(02), 75–79.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, *6*(4), 475–494.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130.

Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among Latino survey respondents. *Cross-Cultural Research*, *53*(1), 87-115.

Davis, R. E., Lee, S., Johnson, T. P., Conrad, F., Resnicow, K., Thrasher, J. F., … Peterson, K. E. (2020). The Influence of Item Characteristics on Acquiescence among Latino Survey Respondents. *Field Methods*, *32*(1), 3–22.

De Jonge, P., & Slaets, J. P. J. (2005). Response sets in self-report data and their associations with personality traits. *European Journal of Psychiatry*, *19*(4), 209–214.

Diamantopoulos, A., Reynolds, N. L., & Simintiras, A. C. (2006). The impact of response styles on the stability of cross-national comparisons. *Journal of Business Research*, *59*(8), 925–935.

Gnambs, T., & Schroeders, U. (2017). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment*, *27*(2), 404-418.

Graham, J. M. (2006). Congeneric and (Essentially) Tau-Equivalent Estimates of Score Reliability: What they are and how to use them. *Educational and Psychological Measurement*, *66*(6), 930–944.

Gudjonsson, G. H. (1990). The relationship of intellectual skills to suggestibility, compliance and acquiescence. *Personality and Individual Differences*, *11*(3), 227–231.

Gudjonsson, G. H., & Young, S. (2011). Personality and deception. Are suggestibility, compliance and acquiescence related to socially desirable responding? *Personality and Individual Differences*, *50*(2), 192–195.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567–578.

Javaras, K. N., & Ripley, B. D. (2007). An “unfolding” latent variable model for likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association*, *102*(478), 454–463.

Javeline, D. (1999). Response Effects in Polite Cultures : A Test of Acquiescence in Kazakhstan. *Public Opinion Quarterly*, *63*(1), 1–28.

Jiang, S., Wang, C., & Weiss, D. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, *107*, 109. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00109/full

Komorita, S. S., & Graham, W. (1965). *Number of Scale Points and the Reliability of Scales*. *25*(4), 987–995.
Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological Assessment*, *27*(4), 1301–1311.

Lechner, C., Partsch, M., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology*, 1–21.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140). Retrieved from https://psycnet.apa.org/record/1933-01885-001

Marin, G., Gamba, R., & Marin, B. (1992). Extreme Response Style and Acquiescence Among Hispanics: The Role of Acculturation and Education. *Journal of Cross-Cultural Psychology*, *4*(4), 498–509.

Marsh, H. W. (1996). Positive and Negative Global Self-Esteem: A Substantively Meaningful Distinction or Artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810–819.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education?. *Personality and Individual Differences*, *44*(7), 1539-1550.

Narayan, S., & Krosnick, J. A. (1996). Education Moderates Some Response Effects in Attitude Measurement. *Public Opinion Quarterly*, *60*(1), 58.

Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, *16*(2), 56–69.

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145.

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of Research in Personality*, *44*(1), 53–61.

Rammstedt, B., Kemper, C., & Borg, I. (2012). Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study. *European Journal of Personality*, *27*, 71–81. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/per.1894

Revelle, W., & Zinbarg, R. (2009). Coefficientes Alpha, Beta, Omega and the GLB. *Psychometrika*, *74*(1), 145–154.

Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*(2), 230–243.
Ross, C. E., & Mirowsky, J. (1984). Socially-Desirable Response and Acquiescence in a Cross-Cultural Survey of Mental Health. *Journal of Helath and Social Behavior*, *25*(2), 189–197.

Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association*, *96*(453), 20–31.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment and Evaluation in Higher Education*, *35*(1), 117–134.

Salinas-Jiménez, M. D. M., Artés, J., & Salinas-Jiménez, J. (2010). Income, Motivation, and Satisfaction with Life: An Empirical Analysis. *Journal of Happiness Studies*, *11*(6), 779–793.

Samejima, F. (1968). Estimation of Latent Ability Using a Response Pattern of Graded Scores1. *ETS Research Bulletin Series*, *1968*(1), i–169.

Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? *Conference on Human Factors in Computing Systems - Proceedings*, 2215–2223.

Savalei, V., & Falk, C. F. (2014). Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches. *Multivariate Behavioral Research*, *49*(5), 407–424

Schmitt, N., & Stuits, D. M. (1985). Factors Defined by Negatively Keyed Items: The Result of Careless Respondents? *Applied Psychological Measurement*, *9*(4), 367–373.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An Exploratory and Confirmatory Factor-Analytic Investigation of Item Wording Effects on the Obtained Factor Structures of Survey Questionnaire Measures. *Journal of Management*, *21*(6), 1177–1193.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *Psychometrika*, *74*(1), 107–120.

Sliter, K. A., & Zickar, M. J. (2014). An IRT Examination of the Psychometric Functioning of Negatively Worded Personality Items. *Educational and Psychological Measurement*, *74*(2), 214–226.

Solís-Salazar, M. (2015). El dilema de combinar ítems positivos y negativos en escalas. *Psicothema*, *27*(2), 192–199.

Stewart, T. J., & Frye, A. W. (2004). Investigating the use of negatively phrased survey items in medical education settings: Common wisdom or common mistake? *Academic Medicine*, *79*(10 SUPPL.), 18–20.
Streiner, D. L. (2003). An introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*, *80*(1), 99–103.

Stukovsky, R., Palat, M., & Sedlakova, A. (1982). Scoring position styles in the elderly. *Studia Psychologica*, *24*(2), 145–154.

van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of Reverse Wording of Questionnaire Items: Let's Learn from Cows in the Rain. *PLoS ONE*, *8*(7), 1–7.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217.

Vannette, D. L., & Krosnick, J. A. (2014). Answering Questions: A Comparison of Survey Satisficing and Mindlessness. *The Wiley Blackwell Handbook of Mindfulness*, *1–2*, 312–327.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking Internal Consistency in Cronbach's Alpha. *Leisure Sciences*, *39*(2), 163–173.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737-747.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236–247.

Wetzel, E., Böhnke, J. R., & Rose, N. (2016). A Simulation Study on Methods of Correcting for the Effects of Extreme Response Style. *Educational and Psychological Measurement*, *76*(2), 304–324.

Zhou, B., & Mcclendon, M. J. (1999). Cognitive Ability and Acquiescence. *American Statistical Association*, 1003–1012.

Chapter 3 A Comparison of Methods for Correcting for Acquiescent Response Style

3.1 Abstract

The previous chapter in this dissertation showed that balanced scales, which include items representing opposite directions of a latent construct, are limited in mitigating the effects of Acquiescent Response Style (ARS), calling for statistical adjustments. Many ARS adjustment methods have been developed, but there is little evidence on how they compare. Understanding how ARS adjustment methods work for scale scores is necessary to better use survey data, as scale scores are commonly used by researchers to describe respondents' opinions, attitudes and beliefs. The aim of this research was to evaluate four ARS adjustment methods while using balanced scales: Ordinary Least Squares (OLS) regression, Confirmatory Factor Analysis (CFA) as proposed by Billiet and McClendon (2000), CFA as proposed by Savalei and Falk (2014), and Multidimensional Nominal Response Modeling (MNRM). These methods were selected as they stem from well-known modelling frameworks and do not require any special research designs (e.g., contentless scales; multi-trait, multi-method designs). Two separate studies were conducted to assess the efficacy of the adjustment methods. The first was a simulation study in which a six-item balanced scale was generated using a Graded Response Model (GRM). The results showed that the OLS adjustment and Billiet and McClendon's CFA adjustment were more effective than MNRM and Savalei and Falk's CFA adjustment in correcting scale scores, particularly when correlating scales with validation variables. The second study used Web survey data with n=2,363 participants in the U.S. and Mexico. Corrected and uncorrected scores of a Satisfaction

with Life scale were correlated to scores of an Emotional Expressivity scale and a Depression Symptoms scale. In this second study, Billiet and McClendon's CFA adjustment was the most effective in increasing the magnitude of the correlations when compared to the rest of the methods, which was interpreted as improved convergent validity. MNRM yielded the poorest results among all methods, considerably weakening the magnitude of the correlation between Satisfaction with Life and Depression Symptoms. The two studies in this research consistently showed that Billiet and McClendon's CFA approach is a simple yet powerful method to combat the effects of ARS.

3.2 Introduction

Acquiescent response style (ARS) in measurement scales has been a source of concern for decades. This response style is characterized by the selection of agreeable response options, regardless of item content (Baumgartner & Steenkamp, 2001). The effects of ARS on measurement have been well documented and include incoherent correlations and factor structures, limitations in subgroup comparisons and biased score estimations (Baron-Epel, Kaplan, Weinstein, & Green, 2010; Reynolds & Smith, 2010; Rammstedt, Kemper, & Borg, 2012; Rammstedt & Farmer, 2013; Van Vaerenbergh & Thomas, 2013; Danner, Aichholzer, & Rammstedt, 2015).

For some time now, researchers have recommended the use of balanced scales as a solution for addressing ARS (Mirowsky & Ross, 1991; Baumgartner & Steenkamp, 2001; Weijters, Baumgartner, & Schillewaert, 2013). While some state that balanced scales mitigate the effects of ARS on their own (Cloud & Vaugham, 1970; Billiet & Davidov, 2008; Primi, Haulk-Filho, Valentini & Santos, 2020), others argue they are a good source of information to correct for acquiescent behavior (e.g., Billiet & McClendon, 2000; Hutton, 2017). Following this second line of reasoning, extensive research has been done to develop adjustment methods aimed at reducing the ill effects of acquiescence (and response styles in general). These methodologies differ in the information they require, their scope, and vary from simple tallies of agreeable responses to more sophisticated modelling approaches including Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT, Webster, 1958; Billiet & McClendon, 2000, Javaras & Ripley, 2007, Bolt & Johnson, 2009).

Because of the large collection of methods to correct for ARS, some efforts have been made to compare the available approaches (De Beuckelaer, Weijters, Rutten, 2010; Savalei & Falk, 2014; Fan, 2019; Liu, Suzer-Gurtekin, Keusch & Lee, 2019; de la Fuente & Abad, 2020). However, these previous studies have focused heavily on parameter recovery of CFA models and the identification of the response style or of groups of acquiescent respondents. Therefore, little is known about the correction of scale scores, which is relevant as users of survey data are normally interested in using scale scores to describe respondents and their attitudes. To address this gap in understanding, this study analyzes the effectiveness of adjustment methods in eliminating ARS from scale scores of balanced scales.

Because the list of available methods for ARS correction is extensive, this study selected a subset of four methods. These methods were selected as they do not require special scales (e.g., content-free scales) or special designs (e.g., multi-trait multi-method design), which pragmatically makes them a more suitable option for most survey practitioners, particularly for those engaging in secondary analysis. The selected methods are: (a) Ordinary Least Squares (OLS) regression, (b) Billiet and McClendon (2000) CFA model specification, (c) Savalei and Falk (2014) CFA model specification, and (d) Multidimensional Nominal Response Modeling

(MNRM; Falk & Cai, 2016). Next, the rationale for each method and its application to ARS are reviewed.

3.3 Methods of Correcting for Acquiescent Response Style

3.3.1 *Ordinary Least Squares Regression*

For this correction method, the first step is to create a measure of agreeableness, which is considered a proxy of ARS. The agreeableness measure is typically computed by tallying the number of agreements within a set of items for each respondent (e.g., Bachman & O'Malley , 1984; Liu et al, 2019). For example, assume there are *J* items measured using a disagree/agree rating scale (D/A). Each item has *K* response categories where *m* is the last category in the scale that does not express agreement[1]. Therefore, response categories greater than *m* indicate agreement. The response to item *j* for respondent *i* is y_{ij}, a number between 1 and *K*. Then, an indicator variable of agreement can be expressed as follows:

$$A_{ij} = \begin{cases} 0 \ if \ y_{ij} \leq m \\ 1 \ if \ y_{ij} > m \end{cases} \tag{3.1}$$

and the measure of agreeable responses for respondent *i* is computed as:

$$agreement_i = \sum_{j=1}^{J} A_{ij} \tag{3.2}$$

It is desirable that the pool of items used to create this agreeableness score is heterogeneous in nature, meaning that items are not strongly correlated to each other, so that content (the construct) is effectively separated from acquiescence (Baumgartner & Steenkamp,

[1] If there is a midpoint in the scale (e.g., "Neither agree nor disagree") then the midpoint is category *m*; if there is no midpoint, then the last category expressing disagreement is used as *m* (e.g., "Slightly Disagree").

2001, Weijters, 2006, Van Vaerenbergh & Thomas, 2013, De Beuckelaer, Weijters, Rutten, 2010). In theory, continuously endorsing a large number of heterogeneous items is unlikely to occur unless a measurement error such as ARS is present (Liu et al, 2019).

The next step is to compute scale scores, which, for this correction method, is typically done by summing responses to items after reverse-coding (z_{ij}) as

$$\theta_i^{SUM} = \sum_{j=1}^{J} z_{ij} \tag{3.3}$$

Note that z_{ij} is equal to y_{ij} for non-reversed items. Once the scores are calculated, they are modeled as a function of the agreeableness measure as follows, and the model parameters are estimated using OLS regression:

$$\theta_i^{SUM} = \beta_0 + \beta_1 agreement_i + \varepsilon_i \tag{3.4}$$

The final step is to obtain the ARS-corrected scores (θ_i^C) as follows:

$$\theta_i^C = \theta_i^{SUM} - \hat{\beta}_1 agreement_i \tag{3.5}$$

This correction method comes from Webster (1958) and Wetzel, Böhnke, and Rose (2016). Equation 3.5 corresponds exactly to Webster's approach, while Wetzel and colleagues' approach simply uses the residuals in Equation 3.4 as the corrected score, i.e., $\theta_i^C = \varepsilon_i$. These two corrections are equivalent, but the latter one is centered around 0.

3.3.2 *Confirmatory Factor Analysis with a Response Style Factor*

Including a response style factor in CFA models has become a common method to address ARS (e.g., Mirowsky & Ross, 1991; Billiet & McClendon, 2000; Savalei & Falk, 2014). The rationale of this methodology is that construct and response style are separated by constraining the sign of factor loadings of each construct in opposite directions for the reverse-keyed items. While the items loading on the substantive latent construct have positive and negative loadings (according to their wording directions), the response style factor has only positive loadings, which allows model identification. Here, the modeling strategies of Billiet and McClendon (2000) and Savalei and Falk (2014) are assessed. Although these two approaches follow a similar rationale, they differ in the number of balanced scales used, and, therefore, the available information to capture ARS.

Billiet and McClendon's (2000) model is represented in Figure 3.1a. This model specifies two correlated substantive latent constructs (F1 and F2) measured through two balanced scales. A third latent construct (A) measured through all items is specified to capture ARS with all factor loadings fixed to 1. This constraint in factor loadings assumes that ARS affects all items equally. Note that the ARS factor essentially represents a *general* response factor and not precisely an acquiescence factor. Therefore, Billiet and McClendon (2000) recommend correlating the style factor to proxies of acquiescent responding, such as the one in Equation 3.2, to check whether it measures ARS. With this model specification, the factor scores from the two substantive factors are assumed to be free from the effect of ARS.

Salvalei and Falk (2014) proposed a CFA model similar to Billiet and McClendon's (2000) model but different in that, as represented in Figure 3.1b, it uses only one balanced scale to measure both the substantive factor and the response style factor. This model assumes independence of the substantive and response style factor, and that all items are affected equally by ARS (i.e., factor loadings of the response style are fixed to 1). Just like with Billiet and McClendon's model, the ARS factor is a *general* response style factor; therefore, it is necessary to check that it represents acquiescent responding. Substantive scores stemming from this model are considered to be corrected for ARS. It is unclear how this adjustment method compares to

Billiet and McClendon's method; however, if these two methods yield similar results, Savalei and Falk's model has the advantage of being simpler to implement, as it requires less items.

Figure 3.1 Illustration of Billiet and McClendon (2000) and Savalei and Falk (2014) CFA methods for ARS correction

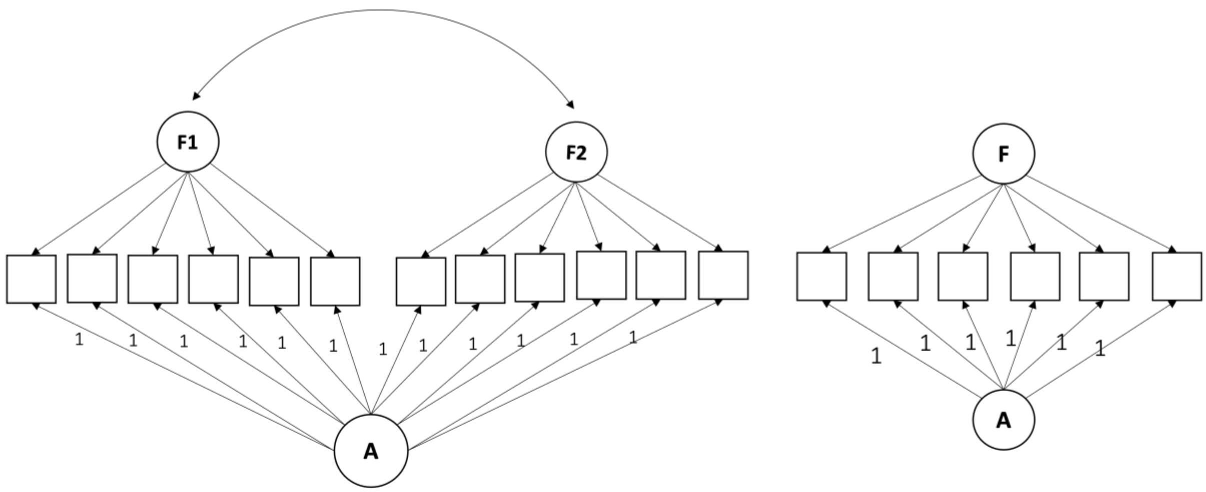


Figure 3.1a. Billiet & McClendon (2000) model. Residual terms are not shown but are assumed to exist.

Figure 3.1b. Savalei & Falk (2014) bidimensional model. Residual terms are not shown but are assumed to exist.

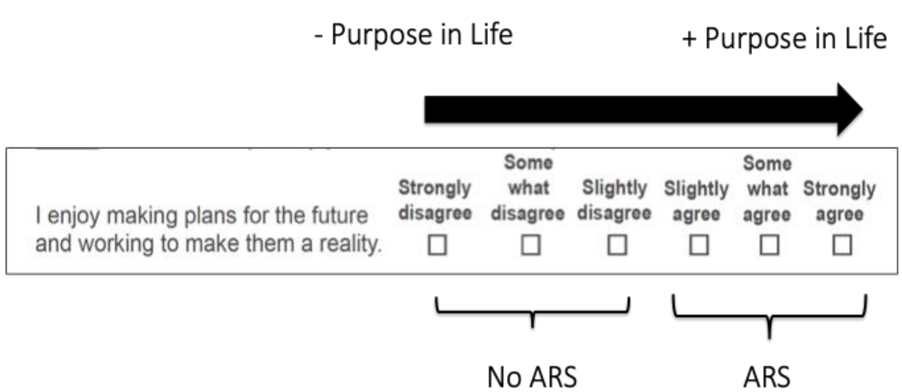
3.3.3 *Multidimensional Nominal Response Model*

The Multidimensional Nominal Response Model was first proposed by Bolt and Johnson (2009) as an extension of the Nominal Response Model by Bock (1972). Bolt and Johnson's model stems from the IRT framework but allows for the modeling of nominal variables and multiple latent constructs simultaneously. Assuming nominal variables means that response categories are not thought to be ordered. However, through the use of scoring functions (described below), users can indicate how categories are ordered in reference to the latent construct. In addition, because the model allows for multiple latent constructs at once, the substantive construct and ARS (or even more response styles) can be modeled simultaneously.

A crucial concept in the use of MNRM is the scoring function, which is a vector that indicates the order of the response categories. For example, for the item from the Purpose in Life scale (Ryff, 1989) with six response categories in Figure 3.2, the ordering of categories can be specified as [0,1,2,3,4,5], and this vector is the scoring function. Using this ordering, the first categories in the rating scale (i.e., “Strongly Disagree” and “Somewhat Disagree”) will be considered in the model as expressing lower Purpose in Life than the later categories, such as “Somewhat Agree” and “Strongly Agree”.

To specify response styles in the model, the scoring function indicates how categories are associated with the response style, typically in a binary fashion. In the case of ARS, the response categories expressing agreement indicate acquiescence, whereas the categories expressing disagreement or neutral opinions (i.e., the midpoint) indicate the absence of acquiescence. Then the scoring function associating the item in Figure 3.2 with ARS would be [0,0,0,1,1,1]. Note that this specification of the ARS scoring function is conceptually similar to the operationalization of ARS in the OLS regression adjustment.

Figure 3.2 Illustration of the differential ordering of categories for an item of the Purpose in Life scale through the Multidimensional Nominal Response Model



This study followed the parametrization from Falk and Cai (2016) for MNRM as follows. Equation 3.6 illustrates the probability of selecting the category k, given two latent constructs, θ_1 and θ_2, that influence category selection. The symbol $\circ$ represents the entrywise product. The parameters α_{djk} and c_{jk} denote the category slope and intercept parameters, respectively, for category k, item j, and construct d. The term s_{dk} is a vector containing a scoring function that links the categories of the scale with the latent construct d. For example, considering the item in Figure 3.2, if θ_1 represents the latent score for Purpose in Life, then $s_{1k} = [0, 1\ , 2, 3, 4, 5]$. Thus, θ_2 represents ARS and $s_{2k} = [0, 0\ , 0, 1, 1, 1]$.

$$\Pr\left(Y_{ij} = k \mid \theta_1, \theta_2\right) = \frac{\exp\left(\left[\alpha_{1jk} \circ s_{1k}\right]\theta_{1i} + \left[\alpha_{2jk} \circ s_{1k}\right]\theta_{2i} + c_{jk}\right)}{\sum_{k=1}^{K} \exp\left(\left[\alpha_{1jk} \circ s_{1k}\right]\theta_{1i} + \left[\alpha_{2jk} \circ s_{1k}\right]\theta_{2i} + c_{jk}\right)} \quad (3.6)$$

From this, the estimate of θ_1 for each individual i is considered the score for the substantive latent construct after correcting for the response style (Falk & Cai, 2016). There are multiple ways in which these scores can be estimated; here the expected a posteriori (EAP) method is used, following the recommendation of Falk and Ju (2020). EAP is computationally advantageous in a setting with multiple latent constructs and has good properties in terms of precision and construct validity of the score (Falk & Ju, 2020).

To compare the aforementioned methods, this research was comprised of two studies. The first was a simulation study in which ARS was the sole source of measurement error and the correction methods were evaluated in their efficacy to remove said error from scale scores. The second study used data from a Web survey to evaluate the impacts of the application of the correction methods to real survey data.

3.4 Study 1: Simulation Study

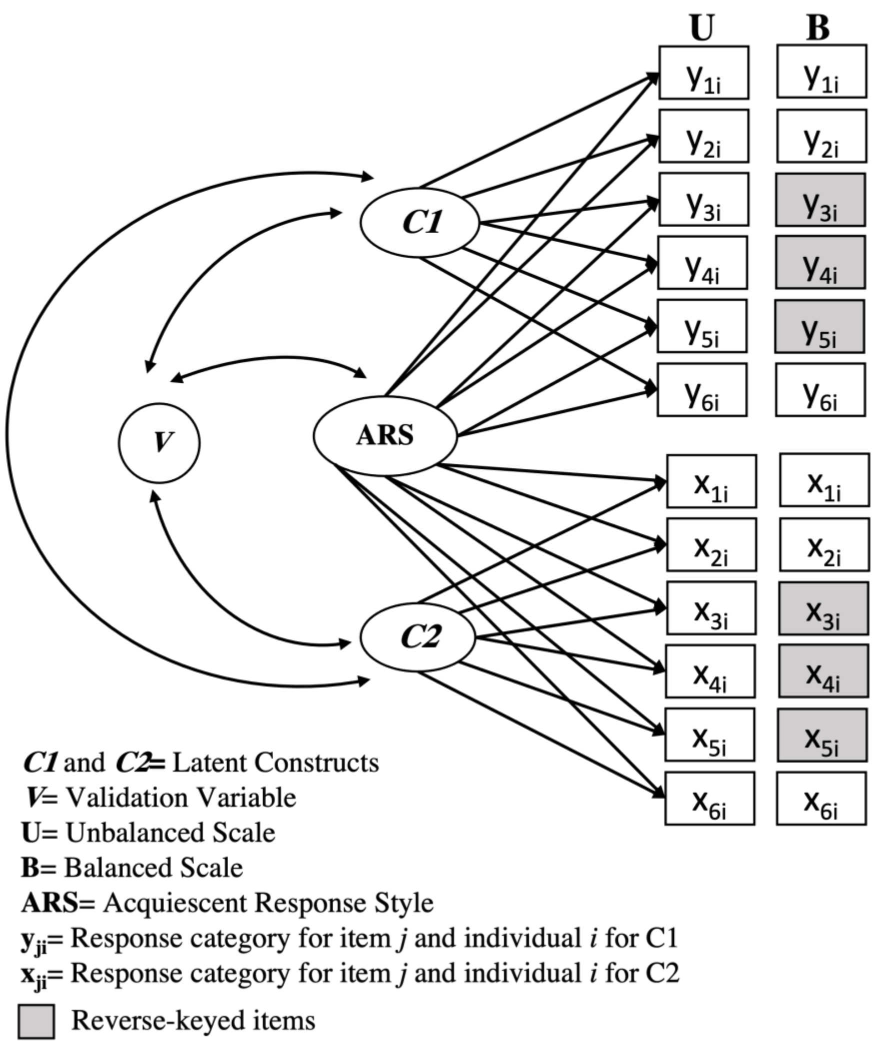
The goal of this first simulation study was to examine in depth how the selected correction methods work on scale scores when ARS is the sole source of measurement error. Because previous studies have found consistent results for the identification of acquiescent groups in real and simulated data for some of the correction methods examined here (e.g., Fan, 2019; Liu et al, 2019), the main hypothesis for this simulation is that the four methods are similar in their efficacy to reduce ARS-related bias in scale scores.

3.4.1 *Data generating process*

The design of this simulation study, illustrated in Figure 3.3, had four main components: two correlated latent constructs (*C1* and *C2*), a continuous ARS construct influencing the items measuring *C1* and *C2*, and a continuous validation variable (*V*) correlated with ARS. *C1* was the primary substantive construct of interest and the one for which corrected scores were computed. *C2* was included to provide the extra items recommended for the OLS regression adjustment and required for Billiet and McClendon's (2000) CFA model specification. The two substantive constructs were defined to be weakly correlated (r=0.10) to introduce some heterogeneity in their substantive contents, as recommended in the literature (e.g., Weijters, 2006; Liu et al, 2019). Both latent constructs were also correlated with the validation variable to different degrees ($r_{C1,V}$ =0.7 and $r_{C2,V}$=0.5).

To generate the data, *C1*, *C2* and V were drawn from a Multivariate Normal Distribution, $\begin{pmatrix} C1 \\ C2 \\ V \end{pmatrix} \sim N_3 \left(\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.1 & 0.7 \\ 0.1 & 1 & 0.5 \\ 0.7 & 0.5 & 1 \end{pmatrix} \right)$, and the ARS construct was drawn separately from a Standard Normal Distribution, $N(0,1)$. To link the ARS construct to V, a Gaussian copula was used and implemented in two steps. First, the target correlation between ARS and the validation variable (r) was adjusted as follows: $r_{adjusted} = 2 \times \sin\left(\frac{r \times \pi}{6}\right)$. Second, the matrix product of the normally distributed ARS construct and the Cholesky decomposition of a 2x2 triangular matrix of $r_{adjusted}$ with 1s in the diagonal was used to produce the desired correlation (r) between ARS and V. Finally, the ARS construct was transformed into a Uniform Distribution, $U(0,1)$, using the *pnorm* command from the *stats* package in R.

Figure 3.3 Illustration of the setup for the simulation study



Multiple levels of the correlation between ARS and V were included in the simulation (r= 0.00, r=0.05, r=0.25, r=0.50, r=0.98, r=-0.05, r=-0.25, r=-0.50 and r= -0.98). This variation in the correlation coefficients was introduced to evaluate how the correction of scores was affected by the relationship between ARS and V, which in real surveys stems from the multiple correlates of ARS, like age, education, cognitive abilities and others (Stukovsky, Palat, & Sedlakova, 1982; Meisenberg & Williams, 2008; Weijters, Cabooter & Schillewaert, 2010). Although it is unlikely that any variable in the real world would yield a very strong correlation with ARS (either positive or negative), this setup contributes to the deeper understanding of the correction of ARS bias.

Using *C1* and *C2* as the underlying latent constructs, items were generated using a Graded Response Model (GRM; Samejima, 1968) as follows. For *C1*, assume that θ_i represents the latent construct of interest for subject i and that $\theta \sim N(0,1)$. There are J items that measure θ through a response scale of $k = 1, 2, 3, \dots, 5$ ordered categories. For each item, the realization of the measurement of item j for subject i is y_{ji}. Then, the probability that a subject with construct level θ_i will endorse category k or higher for item j can be expressed as

$$P(y_{ji} \geq k \mid \theta_i) = \frac{\exp[a_j(\theta_i - b_{kj})]}{1 + \exp[a_j(\theta_i - b_{kj})]} \tag{3.7}$$

Where a_j are the discrimination parameters for each item and b_{kj} are the difficulty thresholds of each category k of each item j. For this study, these parameters were randomly selected following the approach of Jiang, Wang and Weiss (2016), which prevents empty or extreme categories (small or large), which are not very common or even desirable in survey research. The same rationale of the GRM applies for *C2*, and results in the realization of the measurement of *C2* as x_{ji}.

For the computational task of generating all y_{ji}'s and x_{ji}'s from Figure 3.3, the *simIrt* function from the package *catIrt* (Nydick, 2014) in R was used. The inputs of the *simIrt* function are the GRM parameters and a vector representing the latent construct (here *C1* and *C2*), which the function uses to return the discrete response categories. For this study, the output of the *simIrt* function was a set of six items for each substantive construct with five response categories, for which higher response categories indicated higher levels of the construct. This procedure generated the unbalanced version of the scale, and three items were reverse-keyed to create the balanced version of the scale. The items stemming from this procedure were error-free and were used to compute ARS-free scores.

The final step of the data generating process was to introduce ARS to the error-free items for a sample of the generated cases. The percent of cases flagged as "acquiescers" was set to 20% based on the prevalence of ARS reported by Vannette and Krosnick (2014). Cases were selected based on their position in the ARS continuum, which was a uniformly distributed variable correlated with *V*. ARS was introduced for the 20% of cases with the highest values of the ARS construct in each simulation scenario. For the selected cases, a new set of items was obtained using discrimination parameters and difficulty thresholds defined so that higher responses were more likely to occur. None of the items in this final step were reversed-keyed to simulate misresponse due to ARS. The sample size was set to 5,000, and the entire simulation process was replicated 1,000 times, resulting in 1,000 data sets of 5,000 cases each. The syntax of the data generating process of this simulation study is in Appendix A2.

3.4.2 *Data Analysis*

This simulation study focused on construct validity and convergent validity of scale scores. To assess construct validity, the correlation between adjusted scores and *C1*,

$cor(Score, C1)$, was used. In the case of convergent validity, the correlation between scores and the validation variable, $cor(Score, V)$, was used as the primary outcome. Changes in convergent validity were assessed by comparing the values of $cor(Score, V)$ for the scores without ARS, the scores with ARS but with no correction, and the corrected scores. The corrected scores were obtained by implementing the four correction methods as described earlier. The ARS-free and uncorrected scores were computed based on the formula in Equation 3.3, which corresponds to the summation scores and using the ARS-free and uncorrected items respectively. Note that scores were corrected only for *C1* as *C2*'s function in the study was to provide the extra items necessary to fit Billiet & McClendon's CFA model and to increase the pool of items for the agreeableness measure in Equation 3.2.

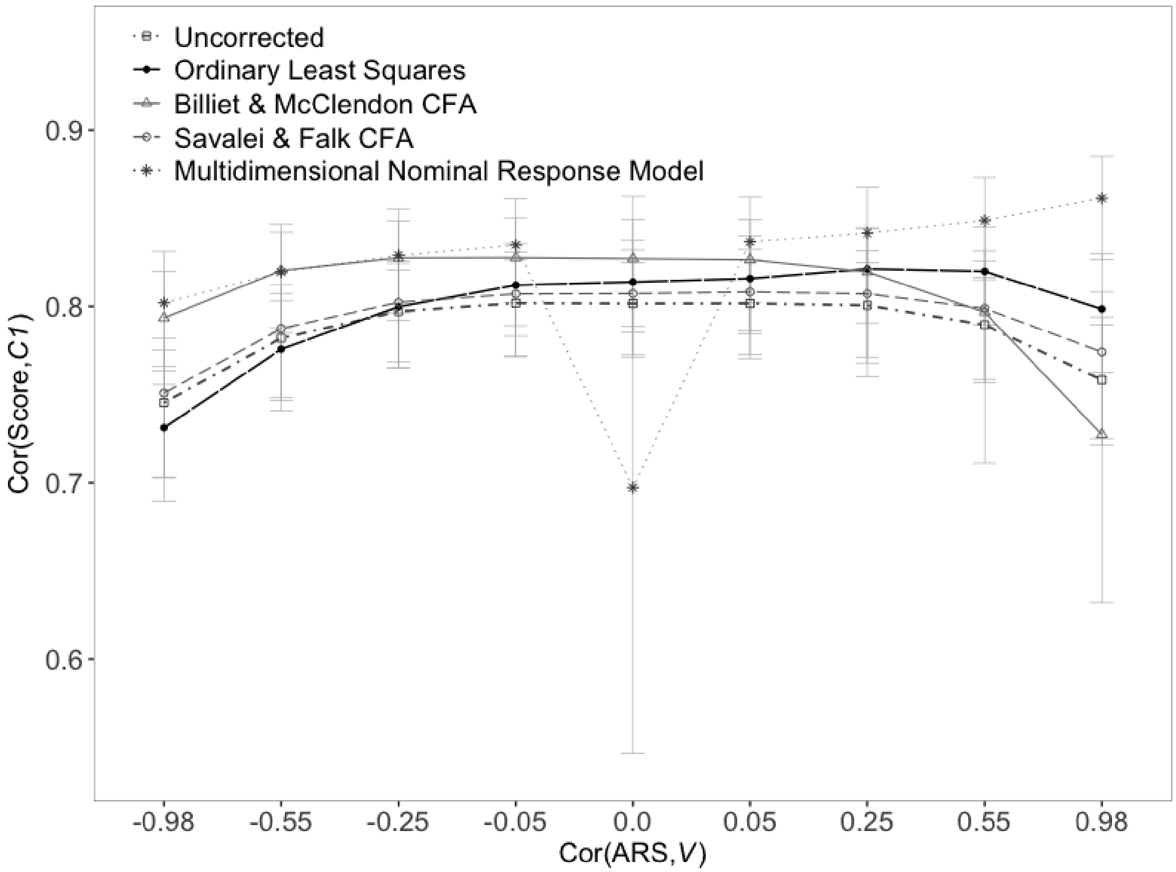
3.4.3 *Results*

Construct validity

Figure 3.4 shows the median, 2.5^{th} and 97.5^{th} percentiles of the distribution of the correlation between the scale scores and the latent construct *C1* for the ARS-free, uncorrected, and corrected scores. In this graph, high correlations imply good construct validity as they indicate that scores closely represent the latent construct. Considering the medians of the distributions, for most of the simulation scenarios, corrected scores yielded better construct validity than uncorrected scores. This was particularly true for MNRM and Billiet and McClendon's CFA adjustment, which yielded the highest construct validity. OLS adjustment followed these methods closely, with slightly lower medians. Savalei and Falk's CFA adjustment yielded the poorest construct validity as the $cor(Score, C1)$ was very close to those using the uncorrected scores. However, it is clear from the error bars in the graph that there was important overlap between the distributions of $cor(Score, C1)$ among correction methods. However, because of the tendencies among the medians, these results offer only partial support the study's hypothesis, which assumes similar efficacy of the adjustment methods.

Even though MNRM showed superior construct validity, there were important problems with the estimation of the parameters of this model. There were convergence problems for some of the simulation replicates except when the correlation between ARS and the validation variable was zero. Because this scenario of independent ARS was also the one with the lowest construct validity, it seems that the high construct validity for the rest of the scenarios was the product of the use of a sample of simulation replications for which the model was successfully estimated and the correction was effective. The rate of model non-convergence was associated with the direction and magnitude of the correlation between the validation variable and ARS. When the correlation between these two variables was positive, the rate of non-convergence was higher for larger correlations and ranged across scenarios from 21% to 40%. When ARS and the validation variable were negatively correlated, the rate of non-convergence was higher when the correlation was smaller and ranged from 5% to 20%.

Figure 3.4 Median, 2.5th percentile and 97.5th percentile of the distribution of the correlation between C1 and both the uncorrected, and corrected scale scores by ARS-correction method

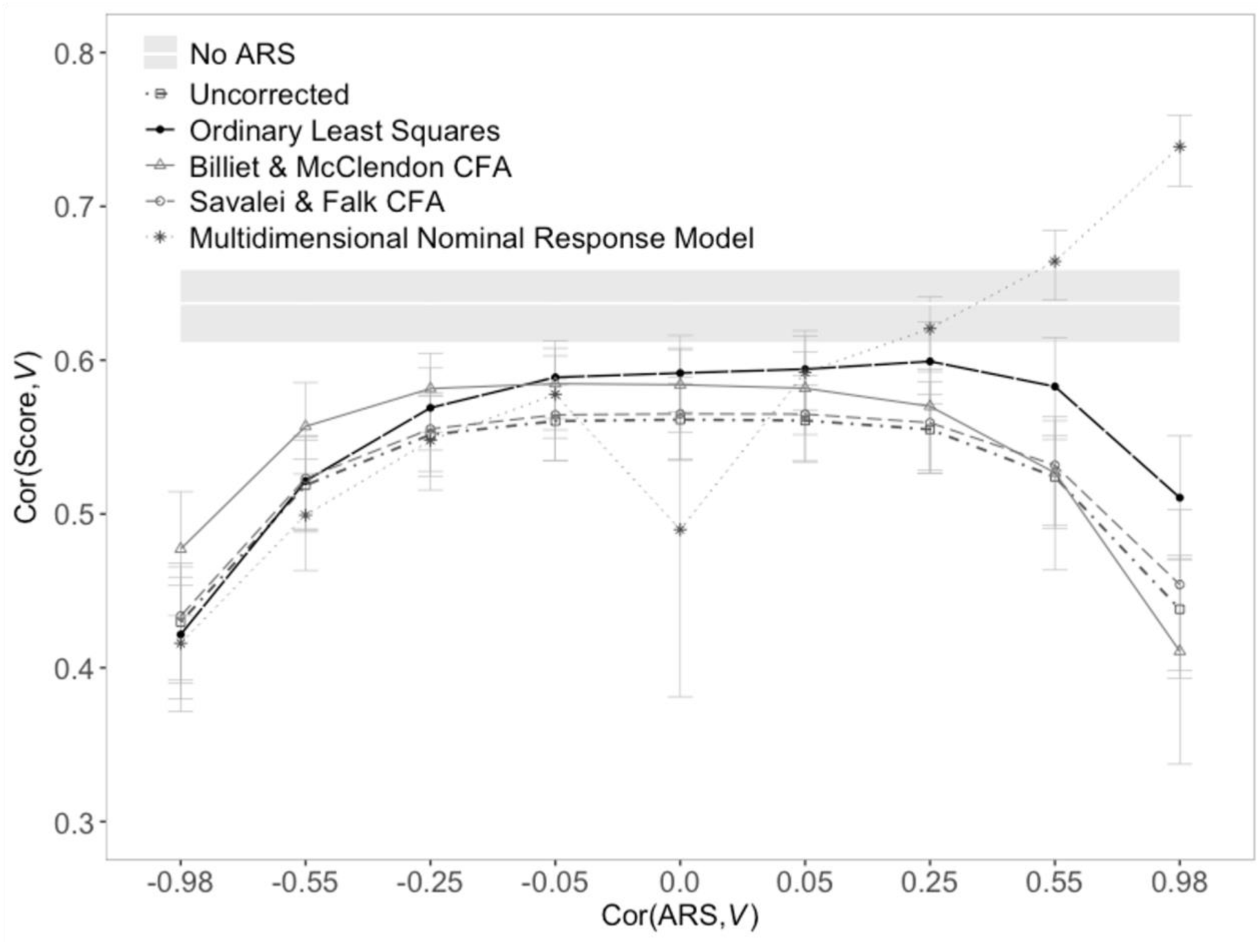


Convergent validity

Figure 3.5 shows the median, 2.5th and 97.5th percentiles of the distribution for $cor(Score, V)$ for the ARS-free, uncorrected, and corrected scores for each adjustment method. The graph shows how ARS affects the correlation between the scores and the validation variable. When ARS was present and not corrected, the correlation decreased. Considering the medians of the distributions, correcting for ARS was most successful for the OLS adjustment and Billiet and McClendon's CFA model, increasing the correlation between the validation variable and the scale scores (although not completely removing the effects of ARS). However, these methods performed best for different simulation scenarios. The OLS adjustment worked best when the correlation between ARS and the validation variable was positive. On the contrary, Billiet and McClendon's CFA performed best when ARS and the validation variable were negatively associated. In addition, Figure 3.5 shows important overlap in the distribution of the correlations of scores and the validation variable among the correction methods. Therefore, these results offer only partial support to the study's hypothesis, as even though important overlap in the distributions across correction methods was observed, the medians of the distributions showed that two methods outperformed the rest in terms of increasing convergent validity.

Note that for MNRM, $cor(Score, V)$ was the lowest when ARS was independent from the validation variable, which corresponds to the only scenario for which there were no convergence issues. Similarly to the previous graph, this suggests that, for those replications for which the model did not converge, the correlation between the scores and the validation variable would have been lower than the median presented in Figure 3.5.

Figure 3.5 Median, 2.5th percentile and 97.5th percentile of the correlation between the validation variable and the ARS-free, uncorrected, and corrected scale scores by correction method



3.4.4 *Discussion*

The aim of this simulation study was to compare different ARS correction methods in their efficacy to return unbiased scale scores and preserve the associations of the scores with other variables. Because of the relevance of scale scores in survey research, this study addresses a critical gap in the ARS literature, which has not compared methods of correction of scale scores. Four correction methods were examined here to adjust scale scores for ARS: OLS regression, CFA using two balanced scales (from Billiet & McClendon, 2000), CFA using one balanced scale (from Savalei & Falk, 2014), and MNRM. To evaluate these methods, the correlation between scores and *C1* and the correlation between scale scores and a validation variable were examined. To determine the effects of the corrections, corrected scores were compared to uncorrected scores and scores without the influence of ARS.

Because previous studies found consistent results for the identification of acquiescent respondents (Fan, 2019; Liu, et al, 2019), the hypothesis of this study was that all methods would produce similar results in terms of correcting scores for ARS. Nonetheless, results of this study did not fully support this hypothesis. Overall, the OLS regression adjustment and Billiet and McClendon's CFA model tended to be more efficacious in correcting for ARS than Savalei and Falk's CFA and MNRM, particularly for correlational analysis. However, it is important to mention that the distributions of the measures of construct and convergent validity showed important overlap across correction methods, implying that differences among these methods are small.

For both construct and convergent validity, the adjustment methods performed the worst for the more extreme cases of the correlation between ARS and the validation variable (0.98 and -0.98) for all correction methods. Researchers should be particularly wary of these scenarios, although it could reasonably be argued that such strong correlations are unlikely to occur in real survey data. Although the methods examined in this study proved to reduce the effects of ARS, these effects were not eliminated completely. Most likely, this is the result of the difficulty in adequately measuring ARS from a relatively small number of items. Future research should focus on improving the existing correction methods through better measurement of ARS to fully address ARS.

3.5 Study 2: Web Survey

To understand how the selected correction methods worked on real survey data, data from a Web survey including multi-item inventories of subjective well-being was analyzed. The aim was to compare the effects of the correction methods on the correlation between the scores of a well-being inventory and two validation variables. In the previous simulation study, the results showed that OLS regression and Billiet and McClendon's CFA model were the most efficacious for reducing the effects of ARS on convergent and construct validity. Therefore, the hypothesis for this application study was that these two methods increase the correlation between the scores of the well-being inventory and the two validation variables more than the other examined ARS correction methods.

3.5.1 *Participants*

Respondents 18 and older were drawn from three different online opt-in panels. Each panel collected data from a different population: (a) Non-Hispanic White respondents in the United States (NHW; n=791), (b) Hispanic respondents in Mexico (HSpa; n=795) and (c) Hispanic respondents in the United States (HUS; n=777). These three populations were selected as they have shown different acquiescent tendencies, therefore providing a better overview of the effects of the adjustment methods on convergent validity. Specifically, Hispanic respondents have consistently showed higher acquiescence when compared to Non-Hispanic White respondents (Aday, Chiu & Andersen, 1980; Marín, Gamba & Marín, 1992), which has been attributed to intrinsic values of the Latino culture that foster acquiescent responding (Davis, Johnson, Lee and Werner, 2019).

3.5.2 *Data collection*

The survey was fielded by Offerwise for Hispanic respondents (in the United States and Mexico) and by Marketing Systems Group (MSG) for the NHW respondents. Incentives were given to respondents in the form of points provided through their respective online panels. To recruit respondents in the United States, panelists were screened for age and race or ethnicity. To recruit panelists in Mexico, panelists were screened for age. Eligibility criteria were communicated to the vendors hired to collect the data and were also checked at the beginning of the questionnaire. Panelists classified as ineligible were not allowed to participate in the study. Quotas by age, education, and gender were used in each panel. The data were collected between February and April 2021. Respondents in the two US panels completed the interview in English, while respondents in Mexico completed the interview in Spanish. All study procedures of this web survey were approved by the Institutional Review Board at the University of Michigan before starting data collection. Informed consent was administered at the beginning of the survey.

3.5.3 *Measures*

The Web survey collected data on multiple measures of subjective well-being. Nonetheless, for this study the main construct of interest was Satisfaction with Life (SWL; Diener, Emmons, Larsen & Griffin, 1985). The original Satisfaction with Life scale contains five items all worded in the same direction. For this study, two items were reverse worded to create a balanced version of the scale. Two constructs known to be correlated with Satisfaction with Life were used to assess convergent validity. The first was Emotional Expressivity, and the second was Depression Symptoms. Emotional Expressivity was measured through 16 items in which respondents rated how emotionally expressive they are of their emotions (Kring, Smith & Neale, 1994). For Depression Symptoms, the Kessler-6 scale was used (Kessler et al, 2002). In addition, to fit Billiet and McClendon's CFA model, a seven-item balanced scale of Purpose in Life (PL; Ryff, 1989) was used as the second substantive construct in the model. All scales in the survey used a 7-point D/A rating scale.

To compute the agreeableness measure required for the OLS regression adjustment (Equation 3.2), four measurement scales included in the web survey were used. These scales were the Affective Orientation Scale (AO; Booth-Butterfield & Booth-Butterfield, 1990), the Simpatía Scale (Davis, Lee, Johnson & Rothschild, 2018), a 7-item scale measuring attitudes toward the novel coronavirus and the Purpose in Life scale (Ryff, 1989). Put together, these scales comprised 42 items. The Simpatía and attitudes toward COVID scales were unbalanced scales, while the AO and PL were balanced. The agreeableness variable serving as a proxy of ARS was created as follows. A score of 1 was assigned if the response was "Slightly agree", a score of 2 if the response was "Agree" and a score of 3 if the response was "Strongly Agree." These codes were assigned prior to reverse coding. For the remaining categories of the rating scale (disagreement and neutral categories) a score of 0 was assigned. Then, by summing these scores across all 42 items for each respondent, the agreeableness variable was created.

3.5.4 *Analysis*

Two types of scores were computed for Satisfaction with Life. The first type were the uncorrected scores, meaning no steps were taken to correct for ARS. Similar to the simulation study, these scores were computed based on Equation 3.3, which corresponds to sum scores. The second type of scores were corrected for ARS. These scores were created using the four correction methods described earlier: OLS regression, Billiet and McClendon's CFA, Savalei and Falk's CFA, and MNRM. An example of the syntax used to compute the corrected scores is in Appendix A3.

The Pearson correlations between Satisfaction with Life, Emotional Expressivity and Depression Symptoms were computed using the corrected and uncorrected scores to evaluate the effects of the correction methods on convergent validity. In addition, Pearson correlations between the agreeableness measure and the ARS factors in the CFA and MNRM adjustments were computed to assess the content of the style factors for these approaches, and to evaluate the content consistency among these ARS factors.

3.5.5 *Results*

Table 3.1 describes the sample for each of the three groups in the study. Overall, the Hispanic group in the US was younger and had more females than the other two respondent groups. The group of Hispanics in Mexico had the biggest percent of college-educated respondents (39%), and the group of Non-Hispanic White respondents had the largest share of respondents with low educational attainment among all groups (17%). The distribution of respondents across income categories was similar for non-Hispanic respondents and Hispanic respondents from Mexico. There were more respondents in the lower category of income for Hispanic respondents in the US than for the other two groups. Finally, Hispanic respondents in the US exhibited more agreeable responses in the proxy variable of ARS, having the largest mean score for this measure (54.3).

Table 3.1 Descriptive characteristics of participants

	Non-Hispanic White (n= 791)	**Hispanic in Mexico (n= 795)**	**Hispanic in the US (n= 777)**
Age in years [Mean (SD)]	48.0 (17.8)	42.7 (14.8)	39.7 (15.2)
Gender (% female)	50.7	52.6	60.9
Married (%)	41.5	49.7	51.07
Education (%)			
Less than High School	16.8	14.5	11.9
High school graduate or equivalent	27.4	16.8	35.0
Some college/ technical/ vocational school	24.7	29.9	21.8
4-year graduate degree and higher	31.1	38.8	31.3
Personal income [a] (%)			
Class 1	32.6	31.1	42.3
Class 2	67.4	68.9	57.7
Language used to complete the survey	English	Spanish	English
Agreeableness to 42 items [Mean (SD)]	47.0 (26.6)	37.8 (19.0)	54.3 (30.5)

[a]In the US, income classes 1 and 2 correspond to yearly income < 15 000 USD and yearly income > 15 000 USD, respectively. In Mexico, income classes 1 and 2 correspond to monthly income < 4250 MXN and monthly income > 4250 MXN. In both countries, class 1 includes individuals earning around minimum wage or less.

Table 3.2 shows the correlations between the corrected scores of Satisfaction with Life and two correlates: Emotional Expressivity and Depression Symptoms. There was a clear trend in the effects of the corrections on convergent validity for all four adjustment methods. Correcting the scores, except for MNRM, increased in magnitude of the correlation between Satisfaction with Life and the correlates, which indicates the reduction of the effects of ARS. In general, using Billiet and McClendon's CFA score adjustment produced the largest increments in the magnitude of correlations when compared to using uncorrected scores. This method was followed by Savalei and Falk's and OLS adjustment in the impact on correlations, as these methods produced somewhat similar results to Billiet and McClendon's method (although changes in the correlations were milder). The MNRM adjustment of scores performed the worst among all methods, consistently lowering the correlations of interest. These results partially support the study hypothesis, as Billiet and McClendon's method increased the magnitude of the correlations more than other adjustment methods.

It is important to note that the magnitude of the changes in the correlations was not the same across populations. The largest changes in the correlations were observed for the Hispanic group in the United States, which had the most acquiescent tendencies according to the agreeableness measure in Table 3.1.

Finally, the correlations between the ARS factors in the CFA and MNRM models and the agreeableness measure described in the measures section were computed to assess the content of the response style factors, as these factors were, by definition, style factors and not necessarily ARS factors. Table A1 in the appendix details all the correlations among the ARS latent factors and the agreeableness measure. Measuring these correlations showed that the ARS factors in Billiet and McClendon's CFA model, Savalei and Falk's CFA model, and MNRM were highly consistent, with an average correlation of 0.73 among the three ARS factors. Furthermore, it was evident these factors measured ARS, as they were moderately correlated with the agreeableness measure (AG; $r_{AG,BM}= 0.58$, $r_{AG, SF}=0.43$, $r_{AG, MNRM}=0.43$).

Table 3.2 Correlation of Satisfaction with Life (SWL) scores with Emotional Expressivity and Depression Symptoms scores by ARS correction method

Correction method	**All respondents (n= 2,363)**	**Non-Hispanic White (n= 791)**	**Hispanic in Mexico (n= 795)**	**Hispanic in the US (n= 777)**
Correlation between SWL scores and Emotional Expressivity scores				
Uncorrected	0.14	0.11	0.20	0.14
Corrected- OLS	0.15	0.11	0.20	0.18
Corrected-Savalei & Falk's CFA	0.15	0.12	0.21	0.20
Corrected-Billiet & McClendon's CFA	0.18	0.15	0.20	0.19
Corrected-MNRM	0.10	0.10	0.19	0.01
Correlation between SWL scores and Depression symptoms scores				
Uncorrected	-0.22	-0.28	-0.27	-0.25
Corrected-OLS	-0.26	-0.32	-0.28	-0.18
Corrected-Savalei & Falk's CFA	-0.30	-0.30	-0.30	-0.26
Corrected-Billiet & McClendon's CFA	-0.35	-0.35	-0.33	-0.32
Corrected-MNRM	-0.07	-0.23	-0.26	0.06

3.5.6 *Discussion*

The Web survey study investigated the effects of four correction methods on real survey data, prone not only to ARS but to other unknown measurement errors. These methods were OLS regression, CFA using two balanced scales (Blliet and McClendon's model specification), CFA using one balanced scale (Savalei and Falk's model specification) and MNRM. Because of the results from Study 1, the hypothesis of this second study was that OLS regression and Billiet and McClendon's CFA model specification would result in higher convergent validity relative to the remaining methods. As there was no "true" score value to compare the corrected scores, only the correlations with other constructs were assessed. To do this, the study compared the corrected vs. uncorrected scores of Satisfaction with Life through their correlations with two other constructs (Emotional Expressivity and Depression Symptoms). Except for MNRM, the correction methods increased the correlation between the substantive constructs. These results partially supported the study's hypothesis of more efficacious correction for the OLS and Billiet and McClendon's adjustment, as the changes in the correlations among constructs were the largest when using Billiet and McClendon's adjustment.

For OLS adjustment, the results showed milder changes in the target correlations when compared to Billiet's and McClendon adjustment, and changes comparable to Savalei and Falk's adjustment. Although there are many possible explanations for this mild effect of the OLS adjustment, most likely this was the result of the relatively low content heterogeneity of items used in this study, which has been considered a suboptimal approach to measure and correct for ARS (De Beuckelaer, Weijters, Rutten, 2010). The Web survey questionnaire included mostly psychosocial and well-being inventories, which, although did not measure the same constructs, measured constructs that are somewhat correlated, complicating the separation of ARS and substantive content. This highlights that, in practice, the heterogeneity recommendation is crucial for the OLS adjustment to work. Although here Billiet and McClendon's adjustment is preferred over the OLS method, any of these two methods will yield more adequate scores than the uncorrected scores, particularly among highly acquiescent respondents.

One key finding of this study was that the magnitude of the corrections was different across the three respondent groups. In particular, for the HUS group the corrections had the largest impact on the magnitude of correlations among all respondents. This group was also the one with the highest agreeable tendencies across the 42 items of the agreeableness measure. This means that the correction methods effectively differentiated among the ARS tendencies of each group, producing the largest corrections for the more acquiescent group while not distorting the correlations for the other groups.

3.6 Limitations

Even though this study provides useful information about the use of adjustment methods to correct for ARS, it is not without limitations. For the simulation study, it is conceivable that the real data generating process is different than the one proposed for the simulation study. In fact, this is likely to be the case, as only one source of measurement error was included in the simulation. Furthermore, only the case in which the validation variable and the latent construct were positively and strongly correlated was explored, and based on the application study, the direction of this correlation can potentially impact the results of the ARS correction. Finally, for the application study, non-probability samples coming from opt-in panels were used, and panelists were recruited from three very specific populations, all of which limits the generalizability of the findings from this study.

3.7 Final remarks

The tendency of systematically endorsing survey items has been a vexing problem for decades, and balanced scales have been proposed as a potential solution for ARS. It was originally believed that the reverse-keying required to estimate scale scores would mitigate the effects of ARS on the scores. However, evidence from the previous chapter of this dissertation showed that the effect of ARS is not completely removed using balanced scales, making post hoc adjustments necessary.

The aim of this research was to compare different methods of ARS correction, and a simulation study and a Web survey were used to this end. The main result of the simulation study was that OLS regression adjustment and Billiet and McClendon's CFA model specification were the most effective methods of correction. When applied to real data, Billiet and McClendon's adjustments showed better results in the correction of ARS than the OLS method. It is plausible that this was the result of a relatively homogenous pool of items in the Web survey, which may have limited the separation of construct and ARS for the OLS method. For survey practice, both the simulation and the application studies indicate that statistical adjustments for ARS are beneficial for the measurement of constructs, can be done simply (e.g., using Billiet and McClendon's correction), and should be the preferred route as any of these methods (except for the MNRM) are better options for representing latent constructs than using uncorrected scores under ARS.

References

Aday, L. A., Chiu, G. Y., & Andersen, R. (1980). Methodological issues in health care surveys of the Spanish heritage population. *American Journal of Public Health*, *70*(4), 367-374.

Bachman, J. G., & O'Malley, P. M. (1984). American Association for Public Opinion Research Yea-Saying , Nay-Saying , and Going to Extremes : Black-White Differences in Response Styles. *Public Opinion Quarterly*, *48*(2), 491–509.

Baron-Epel, O., Kaplan, G., Weinstein, R., & Green, M. S. (2010). Extreme and acquiescence bias in a bi-ethnic population. *European Journal of Public Health*, *20*(5), 543–548.

Baumgartner, H., & Steenkamp, J. B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156.

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, *36*(4), 542-562.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural equation modeling*, *7*(4), 608-628.

Bock, D. R. (1972). Estimating Item Parameters and Latent Ability when Responses are Scored in Two or More Nominal Categories. *Psychometrika*, *37*(1), 29–51.

Bolt, D. M., & Johnson, T. P. (2009). Addressing Score Bias and Differential Item Functioning Due to Individual Differences in Response Style. *Applied Psychological Measurement*, (2001), 335–352.

Booth-Butterfield, M., & Booth-Butterfield, S. (1990). Conceptualizing affect as information in communication production. *Human communication research*, *16*(4), 451-476.

Cloud, J., & Vaughan, G. M. (1970). Using balanced scales to control acquiescence. *Sociometry*, 193-202.

Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among latino survey respondents. *Cross-Cultural Research*, *53*(1), 87-115.

Davis, R. E., Lee, S., Johnson, T. P., & Rothschild, S. K. (2018). Development of a bidimensional simpatía scale for use with Mexican American, Puerto Rican, and Cuban American adults. *Nursing research*, *68*(5), 348-357.

Danner, D., Aichholzer, J., & Rammstedt, B. (2015). Acquiescence in personality questionnaires: Relevance, domain specificity, and stability. *Journal of Research in Personality*, *57*, 119–130.

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761-775.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75.

de la Fuente, J., & Abad, F. J. (2020). Comparing methods for modeling acquiescence in multidimensional partially balanced scales. *Psicothema*.

Falk, C. F., & Cai, L. (2016). A flexible Full-Information Approach to the Modeling of Response Styles. *Psychological Methods*, *21*(3), 328–347. https://doi.org/10.1016/j.cegh.2018.01.004

Falk, C. F., & Ju, U. (2020). Estimation of response styles using the multidimensional nominal response model: A tutorial and comparison with sum scores. *Frontiers in psychology*, *11*, 72.

Fan, Q. (2019). Monte Carlo examination of common approaches for detecting response styles. [Doctoral dissertation, Nanyang Technological University]

Hutton, A. (2017). Assesing Acquiescence in Surveys Using Positively and Negatively Worded Items. [Doctoral dissertation, Virginia Commonwealth University]

Javaras, K. N., & Ripley, B. D. (2007). An “unfolding” latent variable model for likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association*, *102*(478), 454–463.

Jiang, S., Wang, C., & Weiss, D. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, *107*, 109. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00109/full

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine*, *32*(6), 959-976.

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual differences in dispositional expressiveness: development and validation of the Emotional Expressivity Scale. *Journal of personality and social psychology*, *66*(5), 934.

Liu, M., Suzer-Gurtekin, T., Keusch, F., & Lee, S. (2019). Response Styles in Cross-cultural Surveys. In T. P. Johnson, B.-E. Pennell, I. A. L. Stoop, & B. Dorer (Eds.), *Advances in Comparative Survey Methods* (pp. 477–500). New Jersey: Wiley.

Marín, G., Gamba, R. J., & Marín, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of cross-cultural psychology*, *23*(4), 498-509.

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education?. *Personality and Individual Differences*, *44*(7), 1539-1550.

Mirowsky, J., & Ross, C. E. (1991). Eliminating defense and agreement bias from measures of the sense of control: A 2 x 2 index. *Social Psychology Quarterly*, 127-145.

Nydick, S., (2014). catIrt: An R Package for Simulating IRT-Based Computerized Adaptive Tests. R package version 0.5-0. https://CRAN.R-project.org/package=catIrt

Samejima, F. (1968). Estimation of latent ability using a response pattern of graded scores 1. *ETS Research Bulletin Series*, *1968*(1), 1-169.

Savalei, V., & Falk, C. F. (2014). Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches. *Multivariate Behavioral Research*, *49*(5), 407–424.

Stukovsky, R., Palat, M., & Sedlakova, A. (1982). Scoring position styles in the elderly. *Studia Psychologica*, *24*(2), 145.

Primi R., Hauck-Filho N., Valentini F., Santos D. (2020) Classical Perspectives of Controlling Acquiescence with Balanced Scales. In Wiberg M., Molenaar D., González J., Böckenholt U., Kim JS. (Eds.) *Quantitative Psychology. IMPS 2019. Springer Proceedings in Mathematics & Statistics*, vol 322. Springer, Cham.

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145.

Rammstedt, B., Kemper, C., & Borg, I. (2012). Correcting Big Five Personality Measurements for Acquiescence: An 18-Country Cross-Cultural Study. *European Journal of Personality*, *27*, 71–81. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/per.1894

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: A simplified approach to eliminating the problem. *Journal of Service Research*, *13*(2), 230–243.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, *57*(6), 1069.

Vannette, D. L., & Krosnick, J. A. (2014). Answering Questions: A Comparison of Survey Satisficing and Mindlessness. *The Wiley Blackwell Handbook of Mindfulness*, *1–2*, 312–327.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217

Webster, H. (1958). Correcting personality scales for response sets or suppression effects. *Psychological Bulletin*, *55*(1), 62–64. https://doi.org/10.1037/h0048031

Weijters, B. (2006). Response Styles in Consumer Research. [Doctoral dissertation, Ghent University]

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: an integrative model. *Psychological methods*, *18*(3), 320.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247.

Wetzel, E., Böhnke, J. R., & Rose, N. (2016). A Simulation Study on Methods of Correcting for the Effects of Extreme Response Style. *Educational and Psychological Measurement*, *76*(2), 304–324.

Chapter 4 Negated and Polar Opposite Items for Balanced Scale construction: An Empirical Cross-Cultural Assessment

4.1 Abstract

Balanced scales blend items written in different directions, often by linguistically reversing some items, and are hypothesized as a method to control for Acquiescent Response Style (ARS). This study examined the differences in measurement properties between two types of balanced scales designed to control for ARS in a cross-cultural research setting. The first balanced scale type included negated items, which were item reversals formed by inserting a negation, such as, “no” and “not.” The second type included polar opposite items, which used antonyms or opposite terms to reverse the item direction (e.g., “unhappy” as the opposite of “satisfied”). Both types of balanced scales were created for four extant well-being inventories whose items were originally unbalanced. Participants were recruited to a Web survey from three populations with well-documented ARS differences: U.S. non-Hispanic Whites (NHWs; n=1,200), U.S. Hispanics interviewed in English (HUS; n=1,200) and Hispanics respondents in Mexico interviewed in Spanish (HMex; n=1,200). Respondents were randomly assigned to (1) unbalanced, (2) negated balanced or (3) polar opposite balanced scales. No statistical differences were observed between negated and polar opposite scales in fit indices of factor models, reliability measures or convergent validity. However, both types of balanced scales outperformed unbalanced scales in convergent validity, with higher correlations between scale scores and validation variables for balanced than unbalanced scales. These findings suggests that negated

and polar opposite balanced scales are equivalent for ARS control, and that they yield adequate measurement properties. This study suggests that when carefully designed, balanced scales can be a useful tool to measure and correct for ARS.

4.2 Introduction

Balanced scales are formed by combining items that are written in opposite directions of a latent trait. This type of scale has become a well-known tool to identify, control and correct for Acquiescent Response Style (ARS). Nonetheless, there is little empirical guidance on how to create a balanced scale and little empirical information on how their wording structure impacts ARS. There are many ways to word item reversals, and subtle changes to question wording have been shown to affect responses. Therefore, it is natural to examine how the structure of item reversal methods influences the measurement properties of balanced scales in the presence of ARS. The aim of this study was to compare how two strategies for item reversals impacted ARS across three respondent groups with different tendencies to engage in ARS.

ARS is the tendency to agree with a survey item regardless of its content and has been extensively described in the literature (e.g., Cronbach, 1946; Paulhus, 1991; Vaerenbergh & Thomas, 2013). ARS can seriously compromise survey results, as it can produce bias in score estimation (e.g., Baumgartner & Steenkamp, 2001), incoherent correlations and factor structures (e.g., Bentler, Jackson & Messick, 1971; Rammstedt, Goldberg, & Borg, 2010; Rammstedt & Kemper, 2011; Rammstedt & Farmer, 2013; Aichholzer, 2014) and undermine statistical inference (Van Vaerenbergh & Thomas, 2013). Furthermore, ARS tends to vary across different populations and population subgroups (Javeline, 1999; Johnson, Kulesa, Cho, & Shavitt, 2005; Davis, Resnicow & Couper, 2011, Hoffman, Mai, Cristescu, 2013; Rammstedt, Danner & Bosnjak, 2017), adding a layer of complexity to comparative research.

Reversing some of the items to make a measurement scale "balanced" has been brought up in the literature as a solution for measuring and correcting for ARS (e.g., Paulhus, 1991; Paulhus & Vazire, 2007; Rammstedt, Danner & Bosnjak, 2017). While some researchers have argued that balanced scales by themselves mitigate ARS effects on measurement (e.g., Paulhus, 1991, Primi, Hauck-Filho, Valentini & Santos, 2020), others have encouraged the use of balanced scales as a method for the identification and modelling of ARS (e.g., Winkler, Kanouse & Ware, 1982; Savalei & Falk, 2014). This latter approach has become more popular among ARS researchers as they aim to isolate and better comprehend the effects of ARS.

To date, however, there is insufficient empirical guidance on how to write items for a balanced scale for purposes of addressing ARS. Most studies have focused on whether or not to use balanced scales (e.g., Benson & Hocevar, 1985; Roszkowski & Soven, 2010; Menold, 2020) or how to use them to correct for ARS (e.g., Baumgartner & Steenkamp, 2001; Savalei & Falk, 2014; Primi, Hauck-Filho, Valentini & Santos, 2020), and very few have focused on differences in item reversal methods (e.g., Barnette, 2000; Baumgartner, Weijters & Pieters, 2018). Research is needed, therefore, to explore whether different forms of item reversal to address ARS affects scale measurement properties such as factor structure, reliability coefficients and correlations with validation variables. It is important to establish a guidance on whether there is a way to mitigate ARS through the use of balanced scales without sacrificing other measurement properties.

This study attempts to add to the literature around the use of balanced scales for addressing ARS by comparing the potential differences in measurement properties between unbalanced scales and two types of balanced scales. To acknowledge that ARS varies across populations and cultures, this study was conducted with participants from three different populations and two different languages: non-Hispanic White respondents in the United States (NHW), Hispanic respondents in the United States (HUS) interviewed in English and Hispanic respondents in Mexico (HMex) interviewed in Spanish. These three populations represent groups that have shown relatively low (NHW, Marín, Gamba & Marín, 1992) and high (HMex, HUS, Aday, Chiu & Andersen, 1980; Marín, Gamba & Marín, 1992) ARS tendencies.

4.3 Conceptual framework

4.3.1 *Causes of ARS*

There have been multiple theoretical explanations for why ARS occurs in surveys. Cronbach (1946) was one of the first to attempt explaining the mechanisms of ARS in the context of educational psychology. He identified indecisiveness as the cause of acquiescence, arguing that when students were unsure about an answer on a true-false test, more often than not, they would choose the "True" category. Although it is not clear from Cronbach's paper why students would systematically choose "True" instead of "False", it has been theorized that prioritizing more deferential responses could be a plausible explanation for this pattern (Lechner, Partsch, Danner & Rammsted, 2019). Therefore, if all (or most) of the items in a test are keyed so that "True" is the correct answer, a student could successfully guess the answers and obtain a higher score than they deserved. As a result, the items and the test would be less valid. As a solution, Cronbach encouraged researchers to mix the keying of items in tests (creating balanced tests) so that this bias could be prevented.

Gilbert (1991), on the other hand, proposed a dual-process theory of belief acceptance that has been used by some authors to explain ARS (e.g., Knowles & Condon, 1999; Swain, Weathers & Neidrich, 2008). This theory proposes a psychological model in which acceptance (or rejection) of an assertion is represented as a two-step process. First, individuals comprehend and accept the assertion. They then subsequently reexamine their acceptance in order to uphold it or reject it. Gilbert's theory implies that acceptance of an assertion comes easier and faster to an individual than the rejection of that same assertion, meaning that there is asymmetrical processing between acceptance and rejection of a belief (or an item in the case of surveys). The author suggested that when this dual process malfunctions, it results in fewer cognitive resources allocated to the reassessment of the initial belief and, therefore, higher assertion acceptance. This failure in the belief acceptance process most likely occurs at different rates across respondents. For example, cognitive abilities could be intertwined with the functioning of the acceptance process, explaining the higher rate of ARS among those with lower cognitive abilities (Zhou, McClendon & Zhou, 1999; Lechner & Rammstedt, 2015).

There has been some empirical evidence to support the dual-process theory. For example, Knowles and Condon (1999) found that response times to personality self-assessments supported Gilbert's dual process theory. Across three studies asking college students to read 100 adjective markers from the Big Five factor structure, they found that acquiescers were faster to accept personality adjectives as self-descriptive, while rejection of the adjectives took similar time between acquiescers and non-acquiescers.

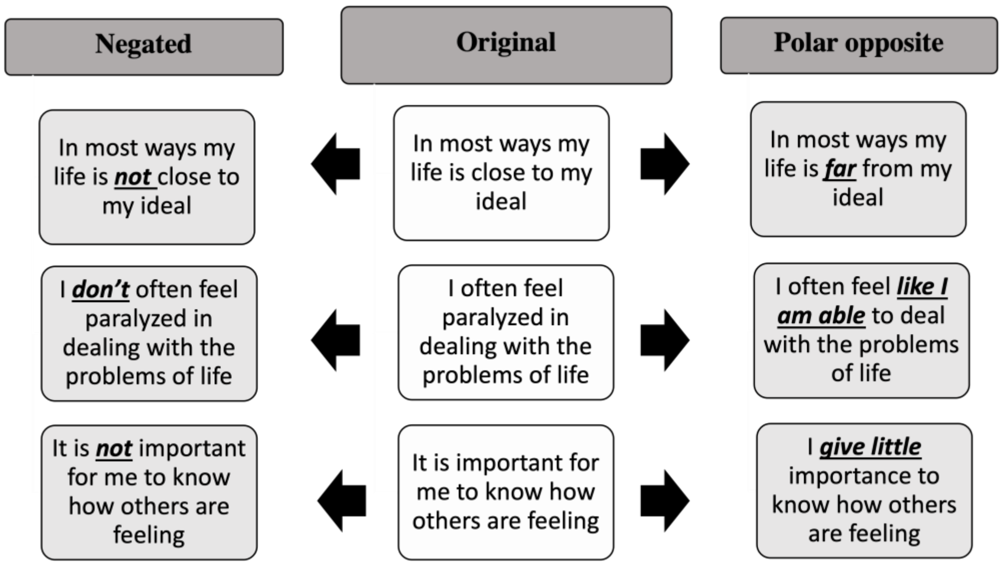
Another theory is that acquiescence is a weak form of satisficing (Krosnick, 1991). When respondents satisfice, they do not go carefully through all stages of the question response process proposed by Tourangeau, Rips and Rasinski (2000), producing lower quality responses. This occurs because respondents do not invest the necessary mental effort to provide the best possible answer. Respondents do not seek an optimal answer but, rather, an acceptable one. Under this theory, choosing an agreement category is seen by respondents as an acceptable response and, therefore, no further efforts are made to find an optimal answer. Krosnick suggests three factors that might foster satisficing (and therefore ARS). First, task difficulty may influence ARS, with more difficult tasks increasing ARS. Second, Krosnick argues that lower cognitive functioning of respondents also increases ARS. Lastly, he reasons that unmotivated respondents will not make sufficient effort to provide optimal answers, leading to higher ARS among respondents with lower rather than higher motivation.

Finally, Lechner et al. (2019) unified these previous theories about ARS. They argue that ARS functions through two mechanisms that generate different correlates of ARS at the respondent, situational and cultural level. These two mechanisms are (1) cognitive processing capacities and (2) differential communication styles. The first mechanism links to Krosnick's theorization of ARS, assuming that adequate cognitive resources are not allocated to the survey response process. The second mechanism is associated with differential communication styles among respondents. For this mechanism, the authors argue that ARS occurs when respondents desire to conform to perceived expectations and deference to hierarchy. Some cultures have been observed to be more acquiescent than others. For example, Latino respondents have consistently been reported to acquiesce more than other ethnic groups throughout different age spectrums in the United States (Aday, Chiu & Andersen, 1980; Marín, Gamba & Marín, 1992), and Davis, Johnson, Lee and Werner (2019) argue that these cross-cultural differences could be the result of values and beliefs associated with Latino culture that encourage ARS.

4.3.2 *Equivalence of reversed and non-reversed items*

There are different ways to word reversed items. Here, two ways identified by Baumgartner, Weijters, and Pieters (2018) in which an item can be linguistically reversed are examined: negated items and polar opposite items. To exemplify how each of these is constructed, consider Figure 4.1. In negated items, there is a negation particle in the statement (such as *not* or *no*). In polar opposite items, the conceptual opposite of the latent construct is measured through a positive statement while avoiding negated wording (such as *far* or *little*).

Figure 4.1 Illustration of item reversal methods



An optimal reversed item should be equivalent in meaning and measurement to its non-reversed counterpart. An important component of that equivalence is symmetrical measurement. For example, when using an agree/disagree (A/D) scale, if a respondent chooses "Strongly agree" for the non-reversed item, she/he should choose "Strongly disagree" for the reversed version of the same item in order to be logically consistent and avoid misresponses. To achieve this, the reversion should include wording to convey the exact opposite of the term it is trying to replace in the original item. This view of reversion is known as the "literalist account of negation" (Colston, 1999; Paradis & Willners, 2006).

However, available empirical evidence suggests that this literalist approach is hard to achieve. Both Colston (1999) and Paradis and Willners (2006) found differences in meaning between positive terms, polar opposites, and negated terms. Schriesheim, Eisenbach and Hill

(1991) investigated the reliability and validity of a scale when including only non-reversed items versus using balanced scales written with negated and polar opposite items. Participants were asked to rate items about an imaginary work supervisor based on a description provided by the researchers. For the sample of undergraduate students, the results showed that the scale containing only non-reversed items and the balanced scale including negated items were comparable in validity. Lower validity was found for the scale containing items with polar opposite terms than scales with negated or solely non-reversed items. However, as this study was based on a sample of college students, it remains unclear what the pattern would be for the general population.

Another concern is the potentially higher cognitive load of reversed items. Negated items, in particular, have been shown to be harder to understand or to require higher reading or cognitive ability (Baker & Ebel, 1982; Weems, Onwuegbuzie & Collins, 2006; Sliter & Zickar; 2014; Gnambs & Schroeders, 2017). Other studies argue that negated items are harder to answer not because of comprehension problems but rather because of difficulties at the judgement stage of the response process (Benson & Hocevar, 1985; Swain, Weathers & Niedrich, 2008; Menold, 2020). Paradis and Willners (2006) observed that negated items took longer to complete, which was interpreted as evidence of difficulty in cognitive processing. Collectively, this evidence shows how the creation of a balanced scale is not trivial. This study aimed to further examine the differences between item reversal strategies.

4.4 Hypotheses

The goal of this study was to compare the measurement properties across (1) balanced scales with negated items, (2) balanced scales with polar opposite items and (3) unbalanced scales. The term “measurement properties” is used to refer to the following scale properties:

Cronbach's alpha as a measure of reliability, CFA factor structure and model fit, and convergent validity.

Some authors have theorized that reversed items function as a mental speedbump (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003, Weijters & Baumgartner, 2012). If reversed items work in this manner, at least some acquiescent behavior will be prevented as respondents will answer more conscientiously. This should lead to lower ARS for balanced scales than for unbalanced scales. Even if reversed items do not remove acquiescent behavior altogether, the reverse coding of balanced scales can contribute to a partial correction of scores (as shown in Chapter 2), potentially improving convergent validity for balanced scales. Regarding factor structure, previous studies have successfully modelled ARS in a CFA setting while preserving model fit (e.g., Billiet & McClendon, 2000, Weijters, Baumgartner & Schillewaert, 2013). For reliability, balanced scales often yield lower reliability coefficients as shown in Chapter 2.

On the other hand, previous studies agree that negations are hard to understand and that they produce more misresponse that may not be associated with ARS (Swain, Weathers & Neidrich, 2008; Weijters & Baumgartner, 2012). Benson and Hocevar (1985) rationalize this as respondents having difficulty conveying their beliefs by disagreeing to a negated item. For example, respondents could find it more cognitively challenging to convey high satisfaction with life by disagreeing to the statement "I am not satisfied with my life" rather than conveying it by agreeing to "I am satisfied with my life". This reasoning is explained by Swain, Weathers and Neidrich (2008), who argue that disagreeing to a negated item requires more cognitive steps at the judgement stage of the response process to match the belief of a respondent with the item that is presented. Based on this, two hypotheses are proposed:

Hypothesis 1. Balanced scale wordings (negated or polar opposite) will yield similar CFA and convergent validity measures but lower reliability than unbalanced scales.

Hypothesis 2. Scales written using polar opposites will yield better measurement properties than scales written using negations.

4.5 Methods

In order to test the hypotheses, an online, randomized scale wording experiment was set up to compare balanced scales with negated items, balanced scales with polar opposite items and unbalanced scales with only non-reversed items. To do this, balanced and unbalanced versions of four well-being inventories were embedded in a Web survey.

4.5.1 *Subjects*

Three groups of respondents were included in the study: NHW respondents in the United States, HUS respondents and HMex respondents. This was done to represent different tendencies in ARS. Respondents 18 and older were drawn from three different online opt-in panels (one per population of interest). To recruit NHW respondents, respondents from a general US online survey panel were screened for race, ethnicity and age. For HMex, respondents from a general population Mexican online panel were screened for age. Finally, respondents in a Hispanic-targeted in the US panel were screened for age and ethnicity to recruit HUS.

For each population, a total of n=1,200 respondents completed the survey. To obtain relatively homogenous samples across the three populations, quotas by age, education and gender were used. The survey was fielded by Offerwise for HUS and HMex respondents and by Marketing Systems Group (MSG) for the NHW respondents. Incentives were given to respondents in the form of points in their respective online panels. Eligibility criteria were communicated to Offerwise and MSG and were also ascertained at the beginning of the questionnaire. Panelists classified as ineligible in this stage were screened out. The data were collected between February and April 2021. The real purpose of the study was not disclosed to participants to avoid respondents being conscientious about their ARS tendencies. They were asked to participate in a health and well-being survey sponsored by the University of Michigan.

4.5.2 *Data collection*

Respondents were randomly assigned to one of three experimental conditions for each measurement scale: unbalanced measurement scales with only non-reversed items, balanced scales that include negated items, or balanced scale wording with a polar opposite. A third of each respondent group was assigned to each experimental condition. Respondents could see the unbalanced version of one scale and a balanced version for another scale.

To control for possible question order effects, the item order was randomized within each measurement scale. However, the scales themselves were presented in the same order across all participants in the study. To control for possible primacy or recency effects, the response scale direction was also randomized at the respondent level, such that respondents were randomly assigned to either an ascending (from Strongly disagree to Strongly agree) or descending (from Strongly agree to Strongly disagree) response scale. Once assigned, the order of categories was consistently displayed throughout the questionnaire.

The median completion times were 19.3 minutes for NHW, 25.9 minutes for HMex and 25.4 minutes for HUS. All study procedures were approved by the Institutional Review Board at the University of Michigan before starting data collection. Informed consent was administered at the beginning of the survey.

4.5.3 ***Measures***

Experimental Scales

For the experiment, four well-established, unbalanced well-being inventories were selected as the control condition. For each inventory, two balanced versions were created: a scale that included negated items and a scale that included polar opposite items. The exact wording for each experimental scale can be found in Table A2 to Table A5 in the appendix.

<u>Satisfaction with Life (SWL)</u>. The SWL scale is a measure of a person's global satisfaction with their life (Diener, Emmons, Larsen & Griffin, 1985). This scale was composed of 5 items using a 7-point A/D scale, which were coded so that higher codes indicated higher satisfaction (i.e., 1= "Strongly disagree"; 7= "Strongly Agree"). The score of the scale was created by reverse coding when necessary and adding the scores of all items. Through structural equation modeling, it has been consistently reported that the SWL scale is one-dimensional (e.g., Diener et al, 1985; Arrindell, Meeuwesen & Huyse, 1991; Neto, 1993; Durak, Senol-Durak & Gencoz, 2010).

<u>Sense of Control (SoC).</u> The SoC scale measures two perceived aspects of sense of control: mastery and constraints. According to Lachman and Weaver (1998), perceived mastery "refers to one's sense of efficacy or effectiveness in carrying out goals", while perceived constraints "indicate to what extent one believes there are obstacles or factors beyond one's control that interfere with reaching goals" (p.765). The SoC scale was comprised of 12 items: four measuring perceived mastery and eight measuring perceived constraints. The SoC scale

used a 7-point A/D rating scale. In this study, scores were computed by reverse-coding when necessary and summing the scores for all items. Higher scores indicated higher mastery and constraints. A two-dimensional factor structure representing the two subscales was observed by the authors of the scale (Lachman & Weaver, 1998). Because this study analyzed the subscales separately a one-dimension solution was expected for each subscale.

Need for Affect (NA). Maio and Esses (2001) define Need for Affect as the "general motivation of people to approach or avoid situations and activities that are emotion inducing for themselves and others" (p. 585). The authors developed a 26-item scale with two subscales: affect approach and affect avoidance. In this study, only the approach subscale (15 items) was used. Respondents rated each item using a 7-point A/D scale. The rating scale was coded so that higher codes would mean higher tendencies to engage with emotional situations. The scale scores were computed using the same process as described for the SoC scale. Maio and Esses (2001) reported a two-factor solution for the entire scale (one factor for approach tendencies and one for avoidance tendencies). Therefore, as only one subscale was administered in this study, a one-factor solution was expected for the approach subscale.

Social Provisions (SP). Here, a short version (10 items) of the 24-item Social Provisions Scale by Cutrona and Russell (1987) is used to measure Social Provisions. Even though it is shorter in length, this short version measures the same six dimensions of social support as the original longer inventory (attachment, social integration, reassurance of worth, reliable alliance, guidance and opportunity for nurturance; Ipachino et al., 2016). Each item used a 7-point A/D response scale, with higher codes meaning higher levels of social support. The scale scores were obtained by adding the scores of each item. The factor structure of the SP has been reported to comprise an overall Social Provisions' factor with six associated factors representing the six social support dimensions (Ipachino et al., 2016).

Convergent Validity Measures

In addition to the experimental scales, other well-established psychometric measures and some factual questions were included in the questionnaire to assess convergent validity. These measures were chosen based on previous empirical evidence of their associations with the four experimental scales (see Table 4.1). The wording for each validation measure is in Appendix A4.

Table 4.1 Variables and scales used to assess convergent validity

Well-being inventory	**Convergent validity measures**	**Number of items in convergent validity measure**	**Observed direction of asociation**
Satisfaction with Life	Self-rated health	1	Higher life satisfaction among those with better self-rated health (Mossey & Shapiro, 1982; Benyamini, Leventhal & Leventhal, 2004)
	Depression symptoms (Kessler et al, 2002)	6	Lower life satisfaction among those with more depression symptoms (Pavot & Diener, 1993; Guney, Kalafat & Boysan, 2010)
	Purpose in life (Ryff, 1989)	7	Higher life satisfaction among those with higher purpose in life (Bronk, Hill, Lapsley, Talib & Finch, 2009)
	Marital status	1	Higher life satisfaction among those who are married (Yang, 2008; Barger, Donoho & Wayment, 2009; Salinas-Jimenez, Artes & Salinas-Jimenez, 2010)
	Home ownership	1	Higher life satisfaction among those who are homeowners (Rohe & Stegman, 1994)

Well-being inventory	**Convergent validity measures**	**Number of items in convergent validity measure**	**Observed direction of asociation**
Sense of control	Depression symptoms (Kessler et al, 2002)	10	Lower sense of control among those with more depression symptoms (Wardle *et al*., 2004)
	Purpose in life (Ryff, 1989)	7	Higher sense of control among those with higher purpose in life (Yarnell, 1971; Shojaee & French, 2014)
	Income	1	Lower sense of control among those with lower income (Lachman & Weaver, 1998)
	Education	1	Lower sense of control among those with less educational attainment (Schieman & Plickert, 2008)
Need for Affect	Emotional expressivity (Leone & Presaghi, 2007)	16	Higher need for affect among those with higher emotional expressivity (Leone & Presaghi, 2007)
	Affective orientation (Maio & Esses, 2001)	19	Higher need for affect among those with higher affective orientation (Maio & Esses, 2001)
	Gender	1	Higher need for affect among women (Maio & Esses, 2001)
	Age	1	Higher need for affect among those who are younger (Maio & Esses, 2001)

Well-being inventory	**Convergent validity measures**	**Number of items in convergent validity measure**	**Observed direction of asociation**
Social Provisions	Depression symptoms (Kessler et al, 2002)	6	More social provisions among those with lower depression symptoms (Ipachino et al, 2016; Orpana, Lang & Yurkowski, 2019)
	Purpose in Life (Ryff, 1989)	7	More social provisions among those with higher life satisfaction (Chiu, Motl & Ditchman, 2016)
	Gender	1	More social provisions among women (Cutrona & Russell, 1987)
	Marital Status	1	More social provisions among those who are married (Sherbourne & Hays, 1990; Wyke & Ford, 1992)

ARS Measures

<u>*ARS - Agreement count.*</u> This measure consisted of counting the number of "agree" responses for each respondent, which is similar to previous studies (e.g., Bachman & O'Malley, 1984). The higher the count of agreeable responses, the more acquiescent the respondent was considered. This procedure has proven to be consistent with other methods for ARS identification (Liu, Suzer-Gurtekin, Keusch & Lee, 2019) but relies on the assumption that the pool of items used to create the count is heterogenous in terms of topics (Van Vaerenbergh & Thomas, 2013). As such, 5 scales (58 items) that were not part of four well-being inventories in the item balancing experiment were used to compute the count indicator. These scales included: the Emotional Expressivity scale (ES; Kring, Smith & Neale, 1994), the Affective Orientation scale (AO; Booth-Butterfield & Booth-Butterfield, 1990), the Simpatía[2] Scale (Davis, Lee, Johnson & Rothschild, 2018), a 7-item scale measuring attitudes toward the novel coronavirus and the Purpose in Life Scale (PL; Ryff, 1989). The *Simpatía* and attitudes toward COVID scales were unbalanced scales, while the AO and EE scales were balanced and contained both polar opposite and negated items. All scales used a 7-point A/D rating scale.

To create the agreement count for the above-mentioned scales, a score of 1 was assigned if the response was "Slightly Agree", a score of 2 if the response was "Agree" and a score of 3 if the response was "Strongly Agree." These codes were assigned prior to reverse coding. All other responses were assigned a score of 0. The ARS count was formed by adding the assigned scores to each item.

---

[2] This construct refers to a "Latino cultural value of being pleasant, agreeable, likable, nonconfrontational, and respectful in interpersonal interactions" (Davis, Johnson, Lee & Werner, 2019, p. 94).

ARS – Balanced item pair. The second method of assessing ARS consisted of counting misresponses to balanced pairs of items, which used responses to items written in the positive (e.g., "I trust my feelings to guide my behavior") and negative (e.g., "I try not to let feelings guide my actions") directions to assess the same construct. This methodology has also been previously used to compute ARS (Rammstedt, Kemper & Borg, 2013; Konstabel et al, 2017). In this study, seven pairs of opposite 7-point A/D items were included to compute ARS. Items were taken from the EE, the AO and the PL scales. As an example, the items: "I try not to let feelings guide my actions" and "I trust my feelings to guide my behavior" from the AO scale formed the first balanced pair. The remaining pairs are shown in Appendix A5. The opposite pair count ranged from 0 (if respondents provided no misresponses responses to any pair of items) to 7 (if respondents provided misresponses to all seven pairs of items).

Note that the EE and AO were part of the ARS measures and used to evaluate convergent validity. However, these scales were never used simultaneously for these two purposes.

4.5.4 *Translation and Adaptation*

An existing translation was used for the Satisfaction with Life scale (Muñoz de Arenillas, Fernandez Borrero, Perez Moreno & Fernandez Mollido, 2010). The author, a bilingual, native Spanish speaker, translated the remaining items into Spanish. All item translations were reviewed by two other bilingual, native Spanish speakers.

4.5.5 *Pilot Test and Pretest*

A pilot study was conducted on SWL, SoC and NA to ensure that the translated scales were understandable and had acceptable measurement properties. The pilot sample consisted of 107 Costa Rican college students who were invited to participate via an email containing the survey link. The Cronbach's alpha coefficients were 0.82 for SWL, 0.81 for SoC, 0.82 for NA. After the pilot, 24 cognitive interviews were conducted with English and Spanish speakers to pretest the items. This pretesting indicated that all items were easily understood, and no changes were made to the items.

4.5.6 *Scale Balancing*

Extensive preparation went into modifying the items to create the two balanced versions of each experimental scale. The English negated items were created by introducing the adverb "no" or "not", while the Spanish negated items were created by introducing the adverb "no." The English polar opposite items were created by using a conceptual polar opposite of the original term (e.g., "far" as the polar opposite of "close"). Terms with "un-", "in-" or "dis-" prefixes were only used if they were among the 10,000 most used words from the Corpus of Contemporary English (COCA; Davies, 2008). The use of terms with these prefixes aimed to simplify wording, as there are terms for which the more natural choice of an opposite term would be a word with a prefix (e.g., like and dislike). For the Spanish polar opposite items, the same logic was used. Polar opposites with "in-" or "des-" prefixes were only introduced if they were among the 10,000 most used words of the Corpus de Referencia de Español Actual (CREA, Real Academia Española, 1997).

4.5.7 *Data Analysis*

All analyses in this study were conducted with Stata 16 ® and R version 4.0.3. Analyses were done independently for all three samples and across scale wordings. Three measurement properties were assessed: reliability, CFA factor structure and model fit parameters, and convergent validity. Reliability was measured using Cronbach's alpha (Cronbach, 1951) and estimated using the built-in *alpha* command in Stata.

CFA was used to explore the factor structure and model fit of the different scales. To fit the models, each scale was assumed to represent a single, one-dimensional latent construct (Figure 4.2a). In the cases where the scale was balanced, a second model that included a style factor was fit (Figure 4.2b). To ensure model identification, style factor loadings were fixed to 1. This bidimensional model configuration was proposed by Savalei and Falk (2014) to control for bias due to ARS, and it was used here with the same purpose. Model estimation was done using Full-information Maximum Likelihood and score estimation of the substantive and ARS factors was done using the regression method (Thurstone, 1935). To assess model fit, four common measures were used: (1) Comparative Fit Index (CFI), (2) Root Mean Square Error of Approximation (RMSEA), (3) Akaike's Information Criterion (AIC) and (4) Bayesian Information Criterion (BIC). The style factor in the bidimensional model was fitted as a general style factor. Therefore, even though it was assumed that the factor represented ARS, it was necessary to check whether it effectively measured this response style. To do this, correlations were calculated between the style factor score for each model and the ARS indicators described earlier. To further examine these relationships, the correlations between the style factors across scales were also estimated to assess the measurement consistency of these factors. CFA models were estimated using the *lavaan* package in R.

Figure 4.2 One-dimensional and Bidimensional CFA models measuring a substantive factor (F) and an ARS style factor (A)

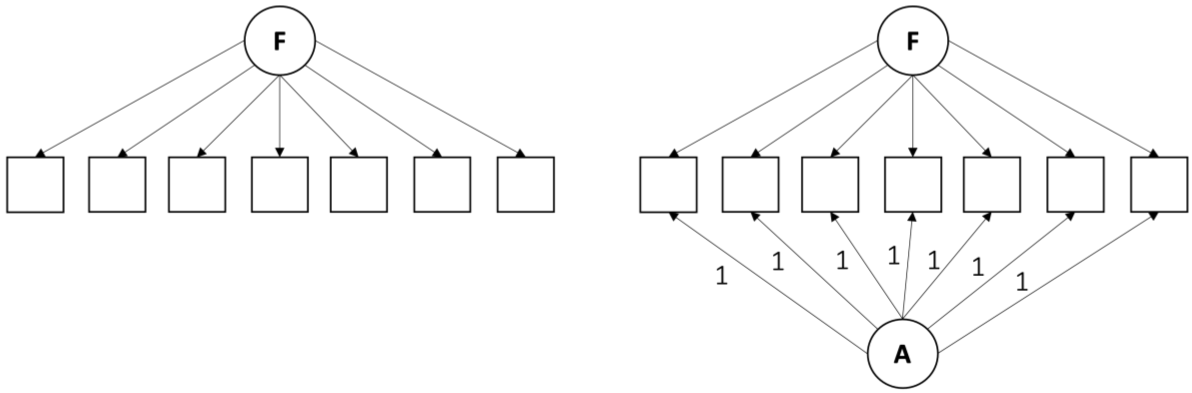


Figure 4.2a. One-dimensional model. Residual terms are not shown but are assumed to exist.

Figure 4.2b. Bidimensional model. The scale representing F is assumed to be balanced. Residual terms are not shown but are assumed to exist.

Convergent validity was assessed using participants' responses to the attitudinal and factual items described earlier (Table 4.1). For attitudinal multi-item inventories, convergent validity was assessed through Pearson correlations between the scores of the well-being inventories in the item balancing experiment and other validation inventories. To assess significant differences in the correlations between reversing methods confidence intervals were computed using Fisher's transformation (Seed, 2001). Differences in coefficients were considered significant when intervals did not intersect. For factual items, the analysis focused on the direction and statistical significance of mean score differences. This is because the magnitude of the mean differences for this type of item has been less documented, depends on the exact items used to measure the construct, and can vary across populations. Differences in score means across groups were assessed using t-tests for two independent samples.

4.6 Results

4.6.1 *Participants*

The demographic composition was different for each of the three groups in the study. The HUS group was on average younger and had more female respondents than the other two groups (Table 4.2). The sample in Mexico had the highest percent of respondents with tertiary education among the three groups (39%), and the sample of NHW had the highest percent of respondents with low educational attainment (18%). More respondents reported a yearly income close to or lower than the minimum wage in the HUS group (43%) than any other group. Hispanic respondents in Mexico and the US reported high use of Spanish to communicate with family (95% and 70%, respectively). For both measures of ARS, the HUS group showed the highest prevalence of ARS among all respondents of the study.

Table 4.2 Descriptive characteristics of participants

	NHW (n=1,200)	**HMex (n=1,200)**	**HUS (n=1,200)**
Age in years [Mean (SD)]	47.8 (18.0)	42.9 (14.7)	39.7 (15.0)
Gender (% female)	49.5	51.4	61.7
Married (%)	42.4	51.1	52.8
Education (%)			
Less than High School	17.8	14.1	12.9
High school graduate or equivalent	26.8	16.0	33.4
Some college/ technical/ vocational school	24.8	30.5	21.8
4-year graduate degree and higher	30.6	39.4	31.9
Personal income [a] (%)			
Class 1	33.4	31.1	42.8
Class 2	66.6	68.9	57.2
Use only or mainly Spanish to communicate with family (%)	1.6	95.2	70.0
Language used to complete the survey	English	Spanish	English
ARS- agreement count [Mean (SD)]	46.7 (26.4)	38.6 (21.2)	55.6 (30.9)
ARS- balanced item pair [Mean (SD)]	1.78 (1.8)	1.98 (1.6)	2.7 (2.1)

[a]In the US, income classes 1 and 2 correspond to yearly income < 15 000 USD and yearly income > 15 000 USD, respectively. In Mexico, income classes 1 and 2 correspond to monthly income < 4250 MXN and monthly income > 4250 MXN. In both countries, class 1 includes individuals earning around minimum wage or less.

4.6.2 *Internal Consistency*

Table 4.3 shows Cronbach's α coefficient for each scale, wording format and population. The highest coefficient for each scale was observed for the unbalanced wording, which supports Hypothesis 1. The change in the coefficient for the balanced versions was slightly different across both scales and populations. For NHW and HMex respondents, although values were lower for the balanced scales, most alpha coefficients remained above or close to the frequently recommended value of 0.7 (except for SoC). In contrast, for HUS, only SP retained values above 0.7 for balanced wordings. In fact, the change in the alphas from an unbalanced to a balanced wording was drastically more evident for the HUS group.

With few exceptions, the coefficients were similar for the two versions of the balanced scales. Furthermore, there was not a trend in the existing small differences. In some instances, the negated scales yielded higher coefficients, while in others, the polar opposite scales resulted in higher alphas. These results do not support Hypothesis 2.

Alpha coefficients were noticeably lower for the balanced versions of both components of SoC for HUS. Further examination of the interitem correlations of the two subscales revealed that all correlations were positive prior to reverse-coding the items. This implies that the reversion was not successful, as respondents (or at least a considerable portion of them) were either not aware of the reversion or did not attend to the direction of the items. Further examination of the interitem correlations for HUS showed that the reversion was also not successful for NA and for one item of the SWL. Because the reversion was ineffective, NA and SoC were excluded from further ARS-related analysis in this study. Considerations are explored in the discussion section.

Table 4.3 Cronbach's alpha coefficient of well-being inventories by item balancing methods and by respondent groups

Well-being inventory	**Respondent Group**		
	NHW	**HMex**	**HUS**
Satisfaction with Life			
Unbalanced	0.90	0.88	0.88
Balanced- Negated	0.77	0.72	0.49
Balanced- Polar Opposite	0.79	0.73	0.46
Sense of Control: Perceived Mastery			
Unbalanced	0.83	0.83	0.81
Balanced- Negated	0.55	0.65	0.21
Balanced- Polar Opposite	0.58	0.60	0.12
Sense of Control: Perceived Constraints			
Unbalanced	0.89	0.88	0.85
Balanced- Negated	0.64	0.56	0.19
Balanced- Polar Opposite	0.74	0.71	0.45
Need for Affect			
Unbalanced	0.92	0.92	0.92
Balanced- Negated	0.82	0.75	0.58
Balanced- Polar Opposite	0.79	0.68	0.55
Social Provisions			
Unbalanced	0.94	0.94	0.93
Balanced- Negated	0.87	0.85	0.79
Balanced- Polar Opposite	0.88	0.86	0.74

4.6.3 *Well-being Factor Structures*

Results of the fit measures for CFA models for SP are in Table 4.4. Results for SWL, NA and both subscales of SoC were similar and shown in Table A6 to A8 in the appendix. However, for the scales with the least number of items (SWL and PM of the SoC), some of the CFA models including the response style factor resulted in a non-positive definite variance-covariance matrices for the estimated parameters. To solve this, either the correlation between the style factor and the well-being measure was specified into the model, or an item was removed from the scale to achieve adequate estimation of models. No patterns in the estimation problems were present regarding scale wording or respondent group.

When no style factor was introduced into the models for the balanced scales, CFI decreased and RMSEA increased. These fluctuations in both measures indicate worse model fit for balanced scales and that a one-dimensional solution was not appropriate. When the style factor was introduced, model fit was improved and, in some instances, surpassed the fit measures for the unbalanced scales. AIC and BIC also reflected this, as these indicators decreased when the style factor was introduced for balanced scales. These results support Hypothesis 1. In contrast, the fit measures between the negated and polar opposite scales were similar with or without the inclusion of the style factor. Thus, no evidence to support Hypothesis 2 was found.

To assess the nature of the style factor in the bidimensional CFA models, the correlations between the ARS factors across models and the ARS measures (ARS- agreement count and ARS- balanced item pair) were compared. All correlations between the style factors and the ARS measures were positive and at least moderate in size (ranging from 0.3 to 0.7), indicating that the style factors represented acquiescent responding. This means that the style factors in the bidimensional solutions effectively measured ARS across all scales and populations. Also, the style factor was consistent across scales, with mostly moderate correlations between the style factors of the experimental scales (ranging from 0.1 to 0.6).

Table 4.4 Fit measures of Social Provisions CFA models by item balancing methods and by respondent groups

Balancing method (Modelling approach)	**Respondent group**		
	NHW	**HMex**	**HUS**[a]
CFI			
Unbalanced (CO)	0.908	0.982	0.974
Negated (CO)	0.709	0.617	0.522
Negated (C+ ARS)	0.949	0.972	0.985
Polar Opposite (CO)	0.662	0.704	0.439
Polar Opposite (C+ ARS)	0.989	0.960	0.986
RMSEA			
Unbalanced (CO)	0.142	0.060	0.065
Negated (CO)	0.191	0.212	0.253
Negated (C+ ARS)	0.082	0.059	0.047
Polar Opposite (CO)	0.236	0.183	0.251
Polar Opposite (C+ ARS)	0.043	0.069	0.041
AIC			
Unbalanced (CO)	12956	10490	12202
Negated (CO)	14717	12887	12822
Negated (C+ ARS)	14300	12335	12176
Polar Opposite (CO)	14949	12646	14726
Polar Opposite (C+ ARS)	14193	12254	13988
BIC			
Unbalanced (CO)	13076	10514	12321
Negated (CO)	14837	13005	12929
Negated (C+ ARS)	14427	12461	12290
Polar Opposite (CO)	15069	12765	14884
Polar Opposite (C+ ARS)	14321	12381	14024

[a] Item 5 measuring SP in the negated format was not included in these analyses as it produced a non-positive definite variance-covariance matrix of the estimated parameters.

CO= Content Factor Only

C+ ARS = Content and ARS Factors

4.6.4 *Convergent Validity*

Table 4.5 shows the correlations between scale scores and other attitudinal correlates. With the exception of the correlation between SP and depression symptoms (for HUS), all correlations across different scale wordings were consistent in direction. However, the magnitude of these correlations varied depending on the direction of the scale. Overall, correlations for the balanced scales were larger in magnitude than for the unbalanced scales (except for SoC Perceived Constraints) supporting Hypothesis 1. Although there was a common trend in the magnitude of the correlations (larger correlations for balanced scales), not all differences were significant. With few exceptions, there were no significant differences between the two types of balanced scales. Thus, Hypothesis 2 was not supported.

Table 4.5 Pearson correlation coefficients between experimental scales and validation variables

Correlation	NHW	HMex	HUS
SWL and Self-Rated Health			
Unbalanced	0.35	0.25	0.34
Negated	0.38	0.25	0.31
Polar Opposite	0.40	0.21	0.34
SWL and Depression			
Unbalanced	-0.33	-0.05	0.00
Negated	-0.27	-0.31[a]	-0.09
Polar Opposite	-0.29	-0.23[a]	-0.28[a,b]
SoC PM and Purpose in Life			
Unbalanced	0.44	0.41	0.33
Negated	0.54	0.51[a]	0.60[a]
Polar Opposite	0.57[a]	0.59[a]	0.53[a]
SoC PM and Depression			
Unbalanced	-0.15	-0.04	-0.06
Negated	-0.38[a]	-0.29[a]	NC
Polar Opposite	-0.31[a]	-0.35[a]	NC
SoC PC and Purpose in Life			
Unbalanced	-0.65	-0.67	-0.58
Negated	-0.51[a]	-0.56[a]	NC
Polar Opposite	-0.60	-0.59	NC
SoC PC with Depression			
Unbalanced	0.63	0.43	0.51
Negated	0.40[a]	0.28[a]	NC
Polar Opposite	0.43[a]	0.40	NC
NA and Emotional Expressivity			
Unbalanced	0.40	0.09	0.05
Negated	0.56[a]	0.35[a]	NC
Polar Opposite	0.48	0.26[a]	NC
NA with Affective Orientation			
Unbalanced	0.59	0.49	0.53
Negated	0.74[a]	0.70[a]	NC
Polar Opposite	0.75[a]	0.58[b]	NC

Correlation	NHW	HMex	HUS
SP and Purpose in Life			
Unbalanced	0.46	0.32	0.31
Negated	0.55	0.55[a]	0.60[a]
Polar Opposite	0.51	0.54[a]	0.58[a]
SP with Depression			
Unbalanced	-0.12	-0.03	0.04
Negated	-0.40[a]	-0.37[a]	-0.48[a]
Opposite	-0.48[a]	-0.39[a]	-0.42[a]

[a] Significant difference with unbalanced wording group, p-value < 0.05

[b] Significant difference with negated wording group, p-value < 0.05

NC= Not computed because the interitem correlations for this scale were positive prior reverse coding.

Table 4.6 shows the mean score differences for each experimental scale between groups defined by validation variables. Results varied across different scales, but the overall direction of the differences was as expected. Regarding inference, more significant differences were observed for the NHW group. No patterns regarding direction or significance of the score differences were observed across scale wordings and for any of the three groups of respondents. Therefore, these results support only Hypothesis 1.

Table 4.6 Differences in scale scores of well-being inventories.

Well-being inventories	**NHW**	**HMex**	**HUS**
Satisfaction with life			
Married – Not married			
Unbalanced	3.44[c]	3.12[c]	1.85[c]
Negated	2.56[c]	1.65[c]	3.06[c]
Polar Opposite	4.26[c]	1.21[c]	2.29[c]
House owners - Non owners			
Unbalanced	2.79[c]	2.22[c]	2.26[c]
Negated	2.31[c]	1.38[c]	2.33[c]
Polar Opposite	2.70[c]	1.48[c]	1.66[c]
Sense of Control: Perceived Mastery			
College – High School or less			
Unbalanced	4.22[c]	2.08[c]	0.68
Negated	1.98	4.29[c]	NC
Polar Opposite	3.98[c]	4.75[c]	NC
Sense of Control: Perceived Constraints			
Personal income class 2 – Personal income class 2			
Unbalanced	7.21[c]	-0.10	5.36[c]
Negated	7.83[c]	4.49[c]	NC
Polar Opposite	4.29[c]	1.83	NC
Need for Affect			
Age 40+ – Age 18 to 39			
Unbalanced	9.26[c]	5.16[c]	1.84
Negated	1.49	1.71	NC
Polar Opposite	-1.80	0.49	NC
Females – Males			
Unbalanced	3.92[c]	-0.32	1.84
Negated	4.46[c]	1.71	NC
Polar Opposite	7.29[c]	2.08[c]	NC
Social Provisions			
Married – Not married			
Unbalanced	3.14	1.79	1.04
Negated	2.87	4.93[c]	5.64[c]
Polar Opposite	6.87[c]	7.34[c]	3.80[c]

Well-being inventories	NHW	HMex	HUS
Social Provisions			
Females – Males			
Unbalanced	0.43	1.37	3.00
Negated	-1.06	-1.84	-1.27
Polar Opposite	6.03[c]	2.11	0.69

[a] Differences are computed as mean of first group listed – mean of second group listed
[b] In the US, income class 1 and 2 correspond to yearly income < 15 000 USD and yearly income > 15 000 USD respectively. In Mexico, income class 1 and 2 correspond to monthly income < 4250 MXN and monthly income > 4250 MXN. In both countries class 1 include individuals earning around minimum wage or less.
[c] Significant difference between the groups of the validation variable, p-value < 0.05
NC= Not computed because the interitem correlations for this scale were positive prior reverse coding.

4.7 Discussion

This study compared the measurement properties of two types of balanced scales and unbalanced scales to assess whether balanced scales can be used to address ARS without affecting measurement properties. The study aimed to provide empirical guidance for the construction of balanced scales to address ARS as the extant literature has focused mainly in whether or not balanced scales should be used and not how they should be written. The measurement properties assessed in this study were CFA fit, reliability and convergent validity. Hypothesis 1 questioned if balanced scales yielded similar measurement properties as unbalanced scales. Hypothesis 2 was concerned with the differences between balanced scales because of the use of negated and polar opposite items. Support for Hypothesis 1 was found, as balanced scales produced similar CFA fit indices once ARS was specified in the model, and higher correlations with validation variables. On the other hand, no support for Hypothesis 2 was found, as there were no differences in the measurement properties between scales written using negations and scales written using polar opposites.

The use of balanced scales to address ARS has remained controversial as loss of factor structure and decreased reliability have been consistently reported in the literature (Schriesheim, Eisenbach & Hill, 1991; Barnette, 2000; Menold, 2020). This study found the same results for both types of balanced scales, however, here these seemingly negative outcomes are interpreted differently. For the one-dimensional CFA model, fit measures were poor for balanced scales, and a second factor emerged. Nonetheless, once this second factor was included in the model, fit measures improved considerably. Additionally, the style factor was found to adequately measure ARS, correlating positively and moderately with the two ARS indicators. The evidence also showed that this second factor was consistent across the well-being inventories, as correlations among style factors across experimental scales were also positive and moderate. These results suggest that for the well-being inventories in the study, balanced scales did not generate a second style factor, but rather enabled the measurement of the response style that would otherwise be undetected. Therefore, the use of balanced scales should be preferred over unbalanced scales in situations where ARS is expected.

Regarding reliability coefficients, alpha coefficients were lower when items were balanced and after reverse coding. Even though this low internal consistency may discourage the use of balanced scales, it is important to consider the rest of results presented in this study and to reflect on the utility and robustness of the alpha coefficient. Multiple studies have criticized the use of Cronbach's alpha and questioned its ability to adequately measure reliability (Komorita & Graham, 1965; Streiner, 2003; Revelle & Zinbarg, 2009; Sijtsma, 2009; Vaske, Beaman, & Sponarski, 2017). Sensitivity to sample size, the number of items in the scale and overall representation of internal consistency instead of reliability all limit the usefulness of the coefficient. Furthermore, correlations between experimental scales and the validation variables were larger for the balanced wordings, which was interpreted as improved convergent validity. Thus, the evidence from this study supports, or at least do not condemn, the use of balanced scales for the purposes of addressing ARS. Therefore, it is reasonable to tolerate the reduction of this coefficient to be able to measure and control for ARS, particularly if it also improves convergent validity. It is important that the discussion moves from whether balancing a scale reduces Cronbach's alpha (because it does) to analyzing and acknowledging the implications of this reduction.

Opposed to expectations, no consistent statistical differences were observed between the two balanced wordings for any of the well-being inventories analyzed in this study. Clearly, this can have many explanations, but here two are explored. First, respondents in this study could be familiarized with the survey completing task as they were recruited from opt-in online panels in which, most likely, they have completed other studies. Thus, it is plausible that because of their previous experience with surveys, they are better at attending different survey cues (e.g., the direction of the items) or better in the use of rating scales. It is also plausible that score differences between the two balancing methods were subtle enough that they did not translate into significant differences in CFA model fit, reliability coefficients and correlations with validation variables. This would be consistent with Paradis and Willners (2006) who found that when respondents rated where adjectives fell on a latent construct spectrum, polar opposites were rated higher than negations, however, these differences were not statistically significant.

In addition, it is important to highlight that the results were consistent across the three group of respondents. This is relevant as ARS has been evident in cross-cultural research (Javeline, 1999; Rammstedt, Kemper & Borg, 2013) and particularly among Hispanic respondents (Davis, Reniscow & Couper, 2011; Davis, Johnson, Lee & Werner, 2019).

However, reversion was not successful for two of the scales for the HUS group. This was evident as interitem correlations were positive for these two scales before items were reverse-keyed. Therefore, for cross-cultural research, it is necessary that balanced scales are carefully tested in all the target populations in the study to make sure balanced scales work for all of them.

Although this research provides valuable information regarding the use of balanced scales, it is not without limitations. In this study, it was not possible to isolate the percent of misresponse due to difficulty in the judgment stage of item responding, which has been suggested as an important source of misresponse (Swain, Weathers & Neidrich, 2008).

Most noticeably, reversion was not successful for two of the scales for the HUS group. There might be many plausible explanations for this failure in reversal, but, as reversal was successful for the other two groups, one possibility is that language proficiency might have affected the understanding of items and resulted in more agreement to items that respondents did not fully comprehend. In this study, 70% of the sample of Hispanic respondents interviewed in English (the HUS group) reported using only or mostly Spanish to communicate with family. Therefore, it is unclear whether the results apply to Hispanics whose main language is English, and more analyses of language use and proficiency are required to understand if this is the cause for this failure in reversal.

In conclusion, the evidence from this study suggests that unbalanced scales did not prevent ARS from occurring, but only hid its effects. For these scales, higher internal consistency along with one-dimensional factor structure seemed to be artifacts of ARS, as convergent validity was lower than for balanced scales. For factor analysis, balanced scales allowed to capture measurement error in the form of a general style factor that was later associated with ARS. Put together, the evidence indicates that ARS was present in all scales, but that it was only possible to capture and control for ARS through balanced scales. The implications of this study for questionnaire design are twofold. First, when ARS is expected unbalanced scales should be avoided as the effects of ARS will be confounded with the measurement of latent constructs, making it impossible to know the extent and consequences of this response style. Lastly, for the purpose of addressing ARS, balanced scales written using negations or polar opposites will yield similar results, and researchers can choose the option that fits their objectives better.

References

Aday, L. A., Chiu, G. Y., & Andersen, R. (1980). Methodological issues in health care surveys of the Spanish heritage population. *American Journal of Public Health*, *70*(4), 367-374.

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality*, *53*, 1-4.

Arrindell, W. A., Meeuwesen, L., & Huyse, F. J. (1991). The Satisfaction With Life Scale (SWLS): Psychometric properties in a non-psychiatric medical outpatients sample. *Personality and individual differences*, *12*(2), 117-123.

Bachman, J. G., & O'Malley, P. M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public opinion quarterly*, *48*(2), 491-509.

Barker, D., & Ebel, R. L. (1982). A comparison of difficulty and discrimination values of selected true-false item types. *Contemporary Educational Psychology*, *7*(1), 35-40.

Barger, S. D., Donoho, C. J., & Wayment, H. A. (2009). The relative contributions of race/ethnicity, socioeconomic status, health, and social relationships to life satisfaction in the United States. *Quality of Life Research*, *18*(2), 179-189.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361-370.

Baumgartner, H., & Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of marketing research*, *38*(2), 143-156.

Baumgartner, H., Weijters, B., & Pieters, R. (2018). Misresponse to survey questions: A conceptual framework and empirical test of the effects of reversals, negations, and polar opposite core concepts. *Journal of Marketing Research*, *55*(6), 869-883.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement*, *22*(3), 231-240.

Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, *76*(3), 186.

Benyamini, Y., Leventhal, H., & Leventhal, E. A. (2004). Self-rated oral health as an independent predictor of self-rated general health, self-esteem and life satisfaction. *Social science & medicine*, *59*(5), 1109-1116.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural equation modeling*, *7*(4), 608-628.

Booth-Butterfield, M., & Booth-Butterfield, S. (1990). Conceptualizing affect as information in communication production. *Human communication research*, *16*(4), 451-476.

Bronk, K. C., Hill, P. L., Lapsley, D. K., Talib, T. L., & Finch, H. (2009). Purpose, hope, and life satisfaction in three age groups. *The Journal of Positive Psychology*, *4*(6), 500-510

Chiu, C. Y., Motl, R. W., & Ditchman, N. (2016). Validation of the Social Provisions Scale in people with multiple sclerosis. *Rehabilitation psychology*, *61*(3), 297.

Colston, H. L. (1999). “Not good” is “bad,” but “not bad” is not “good”: An analysis of three accounts of negation asymmetry. *Discourse Processes*, *28*(3), 237-256.

Cronbach, L. J. (1946). Response sets and test validity. *Educational and psychological measurement*, *6*(4), 475-494.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cutrona, C. E., & Russell, D. W. (1987). The provisions of social relationships and adaptation to stress. *Advances in personal relationships*, *1*(1), 37-67.

Davies, M. (2008). *Corpus of Contemporary American English (COCA).* Retrieved from http://www.americancorpus.org.

Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among latino survey respondents. *Cross-Cultural Research*, *53*(1), 87-115.

Davis, R. E., Lee, S., Johnson, T. P., & Rothschild, S. K. (2018). Development of a bidimensional simpatía scale for use with Mexican American, Puerto Rican, and Cuban American adults. *Nursing research*, *68*(5), 348-357.

Davis, R. E., Resnicow, K., & Couper, M. P. (2011). Survey response styles, acculturation, and culture among a sample of Mexican American adults. *Journal of cross-cultural psychology*, *42*(7), 1219-1236.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71-75.

Durak, M., Senol-Durak, E., & Gencoz, T. (2010). Psychometric properties of the satisfaction with life scale among Turkish university students, correctional officers, and elderly adults. *Social indicators research*, *99*(3), 413-429.

Gnambs, T., & Schroeders, U. (2017). Cognitive Abilities Explain Wording Effects in the Rosenberg Self-Esteem Scale. *Assessment*, *27*(2), 404-418. https://doi.org/10.1177/1073191117746503

Gilbert, D. T. (1991). How mental systems believe. *American psychologist*, *46*(2), 107.

Guney, S., Kalafat, T., & Boysan, M. (2010). Dimensions of mental health: life satisfaction, anxiety and depression: a preventive mental health study in Ankara University students population. *Procedia-Social and Behavioral Sciences*, *2*(2), 1210-1213.

Hoffmann, S., Mai, R., & Cristescu, A. (2013). Do culture-dependent response styles distort substantial relationships?. *International Business Review*, *22*(5), 814-827.

Iapichino, E., Rucci, P., Corbani, I. E., Apter, G., Bollani, M. Q., Cauli, G., ... & Bassi, M. (2016). Development and validation of an abridged version of the Social Provisions Scale (SPS-10) in Italian. *J Psychopathol*, *22*, 157-163.

Javeline, D. (1999). Response effects in polite cultures: A test of acquiescence in Kazakhstan. *Public Opinion Quarterly*, 1-28.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-cultural psychology*, *36*(2), 264-277.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S. L., ... & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological medicine*, *32*(6), 959-976.

Knowles, E. S., & Condon, C. A. (1999). Why people say" yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement*, *25*(4), 987-995.

Konstabel, K., Lönnqvist, J. E., Leikas, S., García Velázquez, R., Qin, H., Verkasalo, M., & Walkowitz, G. (2017). Measuring single constructs by single items: Constructing an even shorter version of the “Short Five” personality inventory. *PloS one*, *12*(8), e0182714.

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual differences in dispositional expressiveness: development and validation of the Emotional Expressivity Scale. *Journal of personality and social psychology*, *66*(5), 934.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied cognitive psychology*, *5*(3), 213-236.

Lachman, M. E., & Weaver, S. L. (1998). The sense of control as a moderator of social class differences in health and well-being. *Journal of personality and social psychology*, *74*(3), 763.

Lechner, C. M., Partsch, M. V., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology*, *72*(3), 426-446.

Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological assessment*, *27*(4), 1301.

Leone, L., & Presaghi, F. (2007). Validity of the need for affect scales: Factorial structure, invariance and validity in the Italian context. *Testing, Psychometrics, Methodology in Applied Psychology*, *14*(3-4), 117-134.

Liu, M., Suzer-Guterkin, T., Keusch, F., & Lee, S. (2019). Response Styles in Cross-cultural Surveys. In Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., & Smith, T. W. (Eds.). (2010). *Survey methods in multinational, multiregional, and multicultural contexts (Vol. 552)*. John Wiley & Sons.

Maio, G. R., & Esses, V. M. (2001). The need for affect: Individual differences in the motivation to approach or avoid emotions. *Journal of personality*, *69*(4), 583-614.

Marín, G., Gamba, R. J., & Marín, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of cross-cultural psychology*, *23*(4), 498-509.

Menold, N. (2020). How Do Reverse-keyed Items in Inventories Affect Measurement Quality and Information Processing?. *Field Methods*, *32*(2), 140-158.

Mossey, J. M., & Shapiro, E. (1982). Self-rated health: a predictor of mortality among the elderly. *American journal of public health*, *72*(8), 800-808.

Muñoz de Arenillas, R. M., Borrero, M. F., Moreno, P. P., & Bollido, M. F. (2010). Evidencias de validez de la adaptación al español de la escala de satisfacción con la vida (SWLS) en una muestra de estudiantes universitarios. *Metodología de encuestas*, *12*(1), 45-62.

Neto, F. (1993). The satisfaction with life scale: Psychometrics properties in an adolescent sample. *Journal of youth and adolescence*, *22*(2), 125-134.

Orpana, H. M., Lang, J. J., & Yurkowski, K. (2019). Original quantitative research-Validation of a brief version of the Social Provisions Scale using Canadian national survey data. *Health promotion and chronic disease prevention in Canada: research, policy and practice*, *39*(12), 323.

Paradis, C., & Willners, C. (2006). Antonymy and negation—The boundedness hypothesis. *Journal of pragmatics*, *38*(7), 1051-1080.

Paulhus, D. L. (1991). *Measurement and control of response bias.* In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of social psychological attitudes, Vol. 1. Measures* *of personality and social psychological attitudes* (p. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L., & Vazire, S. (2007). The self-report method. *Handbook of research methods in personality psychology*, *1*(2007), 224-239.

Pavot, W., & Diener, E. (1993). The affective and cognitive context of self-reported measures of subjective well-being. *Social indicators research*, *28*(1), 1-20.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.

Primi R., Hauck-Filho N., Valentini F., Santos D. (2020) Classical Perspectives of Controlling Acquiescence with Balanced Scales. In Wiberg M., Molenaar D., González J., Böckenholt U., Kim JS. (Eds.) *Quantitative Psychology. IMPS 2019. Springer Proceedings in Mathematics & Statistics*, vol 322. Springer, Cham. https://doi.org/10.1007/978-3-030-43469-4_25

Rammstedt, B., Danner, D., & Bosnjak, M. (2017). Acquiescence response styles: A multilevel model explaining individual-level and country-level differences. *Personality and Individual Differences*, *107*, 190-194.

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological assessment*, *25*(4), 1137.

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factor markers for persons with different levels of education. *Journal of research in personality*, *44*(1), 53-61.

Rammstedt, B., & Kemper, C. J. (2011). Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias. *Journal of Research in Personality*, *45*(1), 121-125.

Rammstedt, B., Kemper, C. J., & Borg, I. (2013). Correcting Big Five personality measurements for acquiescence: An 18–country cross–cultural study. *European Journal of Personality*, *27*(1), 71-81.

Real Academia Española. (1997). *Corpus de referencia del español actual (CREA).* Retrieved from http://corpus.rae.es/creanet.html

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145-154.

Rohe, W. M., & Stegman, M. A. (1994). The effects of homeownership: On the self-esteem, perceived control and life satisfaction of low-income people. *Journal of the American Planning Association*, *60*(2), 173-184.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113-130.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, *57*(6), 1069.

Salinas-Jimenez, M., Artes, J., & Salinas-Jimenez, J. (2010). Income, motivation, and satisfaction with life: An empirical analysis. *Journal of Happiness Studies*, *11*(6), 779-793.

Savalei, V., & Falk, C. F. (2014). Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate behavioral research*, *49*(5), 407-424.

Schieman, S., & Plickert, G. (2008). How knowledge is power: Education and the sense of control. *Social forces*, *87*(1), 153-183.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78.

Sherbourne, C. D., & Hays, R. D. (1990). Marital status, social support, and health transitions in chronic disease patients. *Journal of Health and Social Behavior*, 328-343.

Shojaee, M., & French, C. (2014). The relationship between mental health components and locus of control in youth. *Psychology*, 5, 966-978.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

Sliter, K. A., & Zickar, M. J. (2014). An IRT examination of the psychometric functioning of negatively worded personality items. *Educational and Psychological Measurement*, *74*(2), 214-226.

Seed, P. T. (2001). *sg159: Confidence intervals for correlations*. Stata Technical Bulletin: 59: 27-28.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116-131.
Thurstone, L. L. (1935). *The vectors of mind: Multiple-factor analysis for the isolation of primary traits.* Chicago, IL: University of Chicago Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195-217.

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in Cronbach's alpha. *Leisure Sciences*, *39*(2), 163-173.

Wardle, J., Steptoe, A., Guliš, G., Sartory, G., Sêk, H., Todorova, I., ... & Ziarko, M. (2004). Depression, perceived control, and life satisfaction in university students from Central-Eastern and Western Europe. *International journal of behavioral medicine*, *11*(1), 27-36.

Weems, G. H., Onwuegbuzie, A. J., & Collins, K. M. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, *19*(1), 3-20.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737-747.

Weijters, B., Baumgartner, H., & Schillewaert, N. (2013). Reversed item bias: an integrative model. *Psychological methods*, *18*(3), 320.

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, *67*(5), 555.

Wyke, S., & Ford, G. (1992). Competing explanations for associations between marital status and health. *Social science & medicine*, *34*(5), 523-532.

Yang, Y. (2008). Social inequalities in happiness in the United States, 1972 to 2004: An age-period-cohort analysis. *American sociological review*, *73*(2), 204-226.

Yarnell, T. D. (1971). Purpose-in-life test: further correlates. *Journal of Individual Psychology*. 27(1), 76-79.

Zhou, B., McClendon, M. J., & Zhou, V. (1999). Cognitive ability and acquiescence. In *Proceedings of the survey research methods section* (pp. 1003-1012).

Chapter 5 Conclusions

The purpose of this dissertation was to examine the effectiveness of balanced scales to measure and correct for ARS. To do this, I proposed three main research objectives. The first objective was to investigate the effects of ARS on balanced and unbalanced scales regarding some key measurement properties and the ability of balanced scales to mitigate this response style (Chapter 2). The second objective was to compare statistical methods for ARS correction of scale scores while using balanced scales (Chapter 3). The third objective was to examine the differences in measurement properties of two wording strategies for the creation of balanced scales (Chapter 4). As a whole, this dissertation aimed at providing empirical guidance for practitioners on the use of balanced scales for the correction of ARS.

Chapter 2 examined the impacts of ARS on balanced versus unbalanced scales. More specifically, a simulation study was designed to assess (a) construct validity (through the correlation between scale scores and the true latent construct and bias in score means), (b) convergent validity (through the correlation between scale scores and a validation variable), (c) reliability coefficients (Cronbach's alpha and Greatest Lower Bound), and (d) fit measures of a one-dimensional CFA model. I found differences between balanced and unbalanced scales for all measurement properties. Balanced scales produced better mean estimation for most of the simulation scenarios. Regarding convergent validity, I found higher correlations between balanced scale scores and a validation variable only when ARS and the validation variable were negatively correlated. Reliability measures were impacted for both balanced and unbalanced scales, but with opposite consequences. For unbalanced scales, reliability coefficients were higher, which was consistent with previous evidence (Schriesheim, Eisenbach & Hill, 1991; Barnette, 2000; Eys, Carron, Bray & Brawley, 2007, Roszkowski & Soven, 2010), but this was the result of overestimation of reliability for both Cronbach's alpha and the GLB. This suggests that the apparent superiority in reliability of unbalanced scales may be artificially enhanced by ARS. For balanced scales, both reliability measures decreased, and the drop was larger for Cronbach's alpha than for the GLB. Finally, for CFA, better fit indices were obtained for unbalanced scales when no response style factor was specified in the model. This finding suggests that balanced scales can potentially measure ARS, as the lack of fit for a one-dimensional model suggests that at least two factors are identifiable in the data. The overall conclusion of this chapter was that balanced scales have mixed impacts on measurement properties of constructs. They were helpful for improving mean score estimation and showed the possibility of ARS identification in CFA models. However, reliability coefficients were negatively impacted and correlations with validation variables were improved only under some scenarios. In summary, this chapter demonstrated that balanced scales do not mitigate all effects of ARS on survey data.

Chapter 3 compared four methods for ARS correction using balanced scales: (a) Ordinary Least Squares regression, (b) CFA modelling with two substantive constructs and one style factor (Billiet & McClendon, 2000), (c) CFA modelling with one substantive construct and one style factor (Savalei & Falk, 2014) and (d) Multinomial Nominal Response Modeling. This chapter used a simulation study and an application to real survey data to compare the correction of scores. The simulation study showed that specification of a CFA model with two substantive factors and one style factor (Billiet and McClendon, 2000) and OLS regression adjustment were the most effective in reducing bias in scale scores and/or reducing the attenuation of the correlation between the corrected scores and a validation variable. The subsequent application to real survey data showed that, in general, the corrected scores had stronger correlations with validation variables. In this application study, Billiet and McClendon's model was the method that produced the largest improvements in the correlations between constructs. This chapter showed that the OLS adjustment and Billiet and McClendon's CFA model specification are good alternatives for ARS score correction and should be considered as tools to correct for ARS.

Chapter 4 studied the differences between two types of balanced scales. The first type was formed by using negated statements to reverse items and the second type was formed using polar opposite statements. Negated statements included the particles "no" or "not" and polar opposite statements were formed using antonyms of terms of the non-reversed statements. A randomized wording experiment was embedded in a web survey. Participants were recruited from three populations known to have different acquiescent tendencies: Non-Hispanic White respondents in the United States, Hispanic respondents in Mexico interviewed in Spanish, and Hispanic respondents in the United States interviewed in English. The main hypothesis of this study was that the two types of balanced scales would have different measurement properties. Based on previous evidence (e.g., Baker & Ebel, 1982; Weems, Onwuegbuzie and Collins, 2006; Paradis & Willners, 2006), the expectation was that balanced scales with polar opposite terms would have better measurement properties than negated balanced scales. To test the study's hypothesis, three measurement properties were examined for each type of balanced scales: reliability, CFA model fit and factor structure, and convergent validity. Contrary to expectations, the two types of balanced scales yielded similar measurement properties. Thus, this study concluded that the two reversing strategies were equivalent in terms of construct measurement.

This dissertation aimed at providing an in-depth examination of the use of balanced scales to control for acquiescent responding. Findings from this dissertation showed that balanced scales are far from being the perfect solution to solve ARS; however, I believe they are a good starting point. Balanced scales showed promising results in terms of mean score estimation, convergent validity and CFA model fit (when a response style factor is specified). Based on the results of this dissertation, the recommendation to survey practitioners is to use balanced scales (either with negations or polar opposites) when ARS is expected. Furthermore, scores of these scales should be adjusted as scale balancing does not fully address ARS. Among the methods of score adjustment, Billiet and McClendon's CFA model is recommended, although the OLS adjustment or Savalei and Falk's adjustment will also yield adequate results.

However, it is worth noticing that there are some caveats to the utilization of balanced scales. First, there should be a careful construction of the reversed items, including extensive pretesting with the populations that will eventually answer the questionnaire. Pilot studies with small samples can elucidate any early problems in item reversals. Second, reliability coefficients (such as Cronbach's alpha) worsen when scales are balanced. I argue that this weakening of alpha is an insufficient reason to avoid using balanced scales, as reliability measures, particularly Cronbach's alpha, have their own limitations in the measurement of reliability. In conclusion, I do not recommend against the use of balanced scales, but rather advise that they are used within the context of understanding the attenuation of reliability that they entail. I also argue that even though balanced scales are not without flaws, this does not assure that unbalanced scales are problem-free. I hope that the findings from this dissertation will serve as a starting point for future research on ARS and the use of balanced scales.

References

Barker, D., & Ebel, R. L. (1982). A comparison of difficulty and discrimination values of selected true-false item types. *Contemporary Educational Psychology*, *7*(1), 35-40.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361-370.

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural equation modeling*, *7*(4), 608-628

Eys, M. A., Carron, A. V., Bray, S. R., & Brawley, L. R. (2007). Item wording and internal consistency of a measure of cohesion: The Group Environment Questionnaire. *Journal of Sport and Exercise Psychology*, *29*(3), 395-402.

Paradis, C., & Willners, C. (2006). Antonymy and negation—The boundedness hypothesis. *Journal of pragmatics*, *38*(7), 1051-1080.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113-130.

Schriesheim, C. A., Eisenbach, R. J., & Hill, K. D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78.

Weems, G. H., Onwuegbuzie, A. J., & Collins, K. M. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, *19*(1), 3-20.

Appendices

Appendix A. Additional Materials for Chapter 2

A1 Code used to generate the simulation data for Chapter 2

```
###########################################################
#Data generating for Chapter 2
#To simplify computation procedures,
#this code generates 1000 replicates of the simulation
#to reach the 10,000 replicates, the code was run 10 times
#using different seed numbers
############################################################

rm(list = ls(all.names = TRUE))
library(truncnorm); library(stats); library(car);
library(dplyr);library(MASS);
library(knitr); library(simstudy); library(catIrt); library(utils);
library(lavaan);
library(ppcor); library(psych); library(rlist)

seed <- 04081992
set.seed(seed) #setting seed so it can be replicated
sample.size<- 5000
replicates <- 1000

#generating multiple datasets of correlated
#latent construct (y) and validation variable (x)
data <-replicate(n = replicates,  as.data.frame(mvrnorm(n=sample.size, mu= c(0,0),
                                                         Sigma = 
matrix(c(1,0.7,0.7,1), ncol = 2))),
                 simplify = FALSE )
colnames <- c("x", "y")
data <- lapply(data, setNames, colnames)
```

```
#generating the uniform distributions (that represent ARS) correlated to x
seed <- seed +1
set.seed(seed)

#To facilitate the generation of the 8 scenarios of ARS, the scenarios
were generated
#in sets of 2.

#SET 1: cor(ARS & validation variable)= 0.25 and 0.98
#step 1. generate a vector of random numbers from a normal distribution
X <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X <- mapply(cbind, data,  X, SIMPLIFY= F)
lapply(X, cor)
X <-lapply(X,function(x) { x["y"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1   ,  0.25,   0.98,
         0.25,    1,   0.1,
         0.98 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y <- lapply( X, function(x) { as.matrix(x) %*% C })
lapply(Y, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y <- lapply(Y, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y, cor)


#SET 2: cor(ARS & validation variable)= -0.25 and -0.98
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)
X2 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
```

```
X2 <- mapply(cbind, data,  X2, SIMPLIFY= F)
lapply(X2, cor)
X2 <-lapply(X2,function(x) { x["y"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1   ,  -0.25,  -0.98,
         -0.25 ,    1,   0.1,
         -0.98 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y2 <- lapply( X2, function(x) { as.matrix(x) %*% C })
lapply(Y2, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y2 <- lapply(Y2, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y2, cor)


#SET 3: cor(ARS & validation variable)= 0.05 and 0.55
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)

X3 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X3 <- mapply(cbind, data,  X3, SIMPLIFY= F)
lapply(X3, cor)
X3 <-lapply(X3,function(x) { x["y"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1   ,  0.05,  0.55,
         0.05 ,    1,   0.1,
         0.55 ,  0.1,    1 )
dim(M) <- c(3, 3)
```

```
#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y3 <- lapply( X3, function(x) { as.matrix(x) %*% C })
lapply(Y3, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y3 <- lapply(Y3, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y3, cor)


#SET 4: cor(ARS & validation variable)= -0.05 and -0.55
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)

X4 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X4 <- mapply(cbind, data,  X4, SIMPLIFY= F)
lapply(X4, cor)
X4 <-lapply(X4,function(x) { x["y"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1   ,  -0.05,  -0.55,
         -0.05 ,    1,   0.1,
         -0.55 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}
```

```
#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y4 <- lapply( X4, function(x) { as.matrix(x) %*% C })
lapply(Y4, cor)


#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y4 <- lapply(Y4, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y4, cor)

#changing names to columns in data set to represent the
#continous ARS latent constructs
Y <- lapply(Y, function(x){data.frame(x)})
colnames <- c("x", "cont.ars2", "cont.ars4")
Y <- lapply(Y, setNames, colnames)

Y2 <- lapply(Y2, function(x){data.frame(x)})
colnames <- c("x", "cont.ars6", "cont.ars8")
Y2 <- lapply(Y2, setNames, colnames)

Y3 <- lapply(Y3, function(x){data.frame(x)})
colnames <- c("x", "cont.ars1", "cont.ars3")
Y3 <- lapply(Y3, setNames, colnames)

Y4 <- lapply(Y4, function(x){data.frame(x)})
colnames <- c("x", "cont.ars5", "cont.ars7")
Y4 <- lapply(Y4, setNames, colnames)


#saving ARS latent constructs in original data set and generating ARS
indicators
#20% of the cases are selected to show acquiescent tendencies
data<- mapply(cbind, data,  Y3, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars1 = ifelse(x[["cont.ars1"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars3 = ifelse(x[["cont.ars3"]]
<= 0.8, 0, 1)) })
data <- lapply(data, function(x) { x[,3] <- NULL; x } )

data<- mapply(cbind, data,  Y, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars2 = ifelse(x[["cont.ars2"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars4 = ifelse(x[["cont.ars4"]]
<= 0.8, 0, 1)) })
data <- lapply(data, function(x) { x[,7] <- NULL; x } )


data<- mapply(cbind, data,  Y4, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars5 = ifelse(x[["cont.ars5"]]
<= 0.8, 0, 1)) })
```

```
data <-lapply(data,function(x) { cbind(x, ars7 = ifelse(x[["cont.ars7"]]
<= 0.8, 0, 1)) })
data <- lapply(data, function(x) { x[,11] <- NULL; x } )

data<- mapply(cbind, data,  Y2, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars6 = ifelse(x[["cont.ars6"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars8 = ifelse(x[["cont.ars8"]]
<= 0.8, 0, 1)) })
data <- lapply(data, function(x) { x[,15] <- NULL; x } )

#generate ars indicator for independent ARS
seed <- seed +1
set.seed(seed)
data <-lapply(data, function(x) { cbind(x, random.ars =
rbinom(sample.size, 1, 0.2) ) } ) #20% of the sample to be acquiescers

#Generating the parameters of the Graded Response Model
items <- 6
k <- 5 #number of categories for each item
seed<- seed +1
set.seed(seed)
#generating the difficulty thresholds
#more than needed are generated in order
#to select those with acceptable differences between them according
#to Jiang, Wang and Weiss(2016)
r <-replicate(n = replicates, cbind.data.frame(X1=runif(10000,-1.5, -
0.75), X2=runif(10000,-0.75, 0),
X3=runif(10000,0, 0.75),
X4=runif(1000, 0.75,1.5)), simplify= FALSE)

#identifying adequate generated thresholds
r <-lapply(r,function(x) {
  cbind(x, d12 = abs(x[["X2"]]- x[["X1"]]) < 0.5 | abs(x[["X2"]]-
x[["X1"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d23 = abs(x[["X3"]]- x[["X2"]]) < 0.5 | abs(x[["X3"]]-
x[["X2"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d34 = abs(x[["X4"]]- x[["X3"]]) < 0.5 | abs(x[["X4"]]-
x[["X3"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, adequate = ifelse(x[["d12"]]== FALSE & x[["d23"]]== FALSE &
x[["d34"]]== FALSE, 1, 0)) })

r <- lapply(r, function(x) {
  x <- x[x$adequate==1,]})

lapply(r, function(x) {dim(x)})

seed<- seed +1
```

```
set.seed(seed)
#randomly selecting from the adequate thresholds
sample <- lapply(r, function(x) {sample(1:dim(x)[1], items)})

r <-lapply(r,function(x) {
  cbind(x, id =seq_len(dim(x)[1]) ) })

r2 <- list()
for (i in 1:replicates) {
  r2[[i]] <-cbind(r[[i]], sample2= r[[i]]$id%in%sample[[i]])
}

r2 <- lapply(r2, function(x) {
  x <- x[x$sample2==TRUE,]})
r2 <- lapply(r2, function(x) {x[, 1:4]})

#generating discrimination parameters
a <- list()
for (i in 1:replicates){
  a[[i]]<- runif(dim(r2[[i]])[1],1.1,2.8)
}

#combining all model parameters
g.params <- list()
for (i in 1:replicates){
  g.params[[i]] <- cbind(a[[i]], r2[[i]])}
colnames <- c("a","X1", "X2", "X3", "X4")
g.params <- lapply(g.params, setNames, colnames)


#generating items, appending items to original data and reversing 3 items
g.mod<- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod[[i]] <- simIrt(theta= as.vector(data[[i]]$y), params =
as.matrix(g.params[[i]]), mod = "grm")
  data[[i]]$item1 <- g.mod[[i]]$resp[,1]
  data[[i]]$item2 <- g.mod[[i]]$resp[,2]
  data[[i]]$item3 <- g.mod[[i]]$resp[,3]
  data[[i]]$item4 <- g.mod[[i]]$resp[,4]
  data[[i]]$item5 <- g.mod[[i]]$resp[,5]
  data[[i]]$item6 <- g.mod[[i]]$resp[,6]
  data[[i]]$item3.rev <- car::recode(data[[i]]$item3, "1=5; 2=4; 3=3; 4=2;
5=1")
  data[[i]]$item4.rev <- car::recode(data[[i]]$item4, "1=5; 2=4; 3=3; 4=2;
5=1")
  data[[i]]$item5.rev <- car::recode(data[[i]]$item5, "1=5; 2=4; 3=3; 4=2;
5=1")
}

#generate items under ARS
seed <- seed +1
set.seed(seed)
```

```
a<- c(runif(items,2, 2.8))
r1<- c(a[1],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r2<- c(a[2],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r3<- c(a[3],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r4<- c(a[4],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r5<- c(a[5],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r6<- c(a[6],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))

g.params.ars <- rbind(r1, r2,r3,r4,r5, r6)
g.mod.ars <- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod.ars[[i]] <- simIrt(theta= as.vector(data[[i]]$y), params =
as.matrix(g.params.ars), mod = "grm")
  data[[i]]$ars.item1 <- g.mod.ars[[i]]$resp[,1]
  data[[i]]$ars.item2 <- g.mod.ars[[i]]$resp[,2]
  data[[i]]$ars.item3 <- g.mod.ars[[i]]$resp[,3]
  data[[i]]$ars.item4 <- g.mod.ars[[i]]$resp[,4]
  data[[i]]$ars.item5 <- g.mod.ars[[i]]$resp[,5]
  data[[i]]$ars.item6 <- g.mod.ars[[i]]$resp[,6]
}


#create variables containing the items with ARS for flagged cases
data <-lapply(data, function(x) { cbind(x, i1_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ran_norev = ifelse
(x[["random.ars"]]==1, x[["ars.item6"]], x[["item6"]]))})


data <- lapply(data, function(x) {cbind(x, i1_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <- lapply(data, function(x) {cbind(x, i2_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <- lapply(data, function(x) {cbind(x, i3_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <- lapply(data, function(x) {cbind(x, i4_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <- lapply(data, function(x) {cbind(x, i5_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <- lapply(data, function(x) {cbind(x, i6_ran_rev =
ifelse(x[["random.ars"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item1"]], x[["item1"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i2_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars1_norev = ifelse
(x[["ars1"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars1_rev = ifelse
(x[["ars1"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars2_norev = ifelse
(x[["ars2"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars2_rev = ifelse
(x[["ars2"]]==1, x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item1"]], x[["item1"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i2_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars3_norev = ifelse
(x[["ars3"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars3_rev = ifelse
(x[["ars3"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars4_norev = ifelse
(x[["ars4"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars4_rev = ifelse
(x[["ars4"]]==1, x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item1"]], x[["item1"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i2_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars5_norev = ifelse
(x[["ars5"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars5_rev = ifelse
(x[["ars5"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars6_norev = ifelse
(x[["ars6"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars6_rev = ifelse
(x[["ars6"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item2"]], x[["item2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i3_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars7_norev = ifelse
(x[["ars7"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars7_rev = ifelse
(x[["ars7"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars8_norev = ifelse
(x[["ars8"]]==1, x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars8_rev = ifelse
(x[["ars8"]]==1, x[["ars.item6"]], x[["item6"]]))})

#computing scale scores
data <- lapply(data, function(x) {cbind(x, score = x[["item1"]] +
x[["item2"]] + x[["item3"]] + x[["item4"]] + x[["item5"]] +
x[["item6"]])})
```

```
data <- lapply(data, function(x) {cbind(x, score_rev = x[["item1"]] +
x[["item2"]] + 6-x[["item3.rev"]] + 6- x[["item4.rev"]] + 6-
x[["item5.rev"]] + x[["item6"]])})

data <- lapply(data, function(x) {cbind(x, sran_norev =
x[["i1_ran_norev"]] + x[["i2_ran_norev"]] + x[["i3_ran_norev"]] +
x[["i4_ran_norev"]] + x[["i5_ran_norev"]] + x[["i6_ran_norev"]])})
data <- lapply(data, function(x) {cbind(x, sran_rev = x[["i1_ran_rev"]] +
x[["i2_ran_rev"]] + 6-x[["i3_ran_rev"]] + 6- x[["i4_ran_rev"]] + 6-
x[["i5_ran_rev"]] + x[["i6_ran_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars1_norev =
x[["i1_ars1_norev"]] + x[["i2_ars1_norev"]] + x[["i3_ars1_norev"]] +
x[["i4_ars1_norev"]] + x[["i5_ars1_norev"]] + x[["i6_ars1_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars1_rev = x[["i1_ars1_rev"]]
+ x[["i2_ars1_rev"]] + 6-x[["i3_ars1_rev"]] + 6- x[["i4_ars1_rev"]] + 6-
x[["i5_ars1_rev"]] + x[["i6_ars1_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars2_norev =
x[["i1_ars2_norev"]] + x[["i2_ars2_norev"]] + x[["i3_ars2_norev"]] +
x[["i4_ars2_norev"]] + x[["i5_ars2_norev"]] + x[["i6_ars2_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars2_rev = x[["i1_ars2_rev"]]
+ x[["i2_ars2_rev"]] + 6-x[["i3_ars2_rev"]] + 6- x[["i4_ars2_rev"]] + 6-
x[["i5_ars2_rev"]] + x[["i6_ars2_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars3_norev =
x[["i1_ars3_norev"]] + x[["i2_ars3_norev"]] + x[["i3_ars3_norev"]] +
x[["i4_ars3_norev"]] + x[["i5_ars3_norev"]] + x[["i6_ars3_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars3_rev = x[["i1_ars3_rev"]]
+ x[["i2_ars3_rev"]] + 6-x[["i3_ars3_rev"]] + 6- x[["i4_ars3_rev"]] + 6-
x[["i5_ars3_rev"]] + x[["i6_ars3_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars4_norev =
x[["i1_ars4_norev"]] + x[["i2_ars4_norev"]] + x[["i3_ars4_norev"]] +
x[["i4_ars4_norev"]] + x[["i5_ars4_norev"]] + x[["i6_ars4_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars4_rev = x[["i1_ars4_rev"]]
+ x[["i2_ars4_rev"]] + 6-x[["i3_ars4_rev"]] + 6- x[["i4_ars4_rev"]] + 6-
x[["i5_ars4_rev"]] + x[["i6_ars4_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars5_norev =
x[["i1_ars5_norev"]] + x[["i2_ars5_norev"]] + x[["i3_ars5_norev"]] +
x[["i4_ars5_norev"]] + x[["i5_ars5_norev"]] + x[["i6_ars5_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars5_rev = x[["i1_ars5_rev"]]
+ x[["i2_ars5_rev"]] + 6-x[["i3_ars5_rev"]] + 6- x[["i4_ars5_rev"]] + 6-
x[["i5_ars5_rev"]] + x[["i6_ars5_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars6_norev =
x[["i1_ars6_norev"]] + x[["i2_ars6_norev"]] + x[["i3_ars6_norev"]] +
x[["i4_ars6_norev"]] + x[["i5_ars6_norev"]] + x[["i6_ars6_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars6_rev = x[["i1_ars6_rev"]]
+ x[["i2_ars6_rev"]] + 6-x[["i3_ars6_rev"]] + 6- x[["i4_ars6_rev"]] + 6-
x[["i5_ars6_rev"]] + x[["i6_ars6_rev"]])})
```

```
data <- lapply(data, function(x) {cbind(x, sars7_norev =
x[["i1_ars7_norev"]] + x[["i2_ars7_norev"]] + x[["i3_ars7_norev"]] +
x[["i4_ars7_norev"]] + x[["i5_ars7_norev"]] + x[["i6_ars7_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars7_rev = x[["i1_ars7_rev"]]
+ x[["i2_ars7_rev"]] + 6-x[["i3_ars7_rev"]] + 6- x[["i4_ars7_rev"]] + 6-
x[["i5_ars7_rev"]] + x[["i6_ars7_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars8_norev =
x[["i1_ars8_norev"]] + x[["i2_ars8_norev"]] + x[["i3_ars8_norev"]] +
x[["i4_ars8_norev"]] + x[["i5_ars8_norev"]] + x[["i6_ars8_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars8_rev = x[["i1_ars8_rev"]]
+ x[["i2_ars8_rev"]] + 6-x[["i3_ars8_rev"]] + 6- x[["i4_ars8_rev"]] + 6-
x[["i5_ars8_rev"]] + x[["i6_ars8_rev"]])})

#saving data file
save(data, file= "data.RData")
```

Appendix B. Additional Materials for Chapter 3

A2 Code used to generate the simulation data for Chapter 3

```
###########################################################
#Data generating for Chapter 3
#To simplify computation procedures,
#this code generates 500 replicates of the simulation
#to reach the 1000 replicates, the code was run 2 times
#using different seed numbers
############################################################

rm(list = ls(all.names = TRUE))
library(truncnorm); library(stats); library(car);
library(dplyr);library(MASS);
library(knitr); library(simstudy); library(catIrt); library(utils);
library(lavaan);
library(ppcor); library(psych); library(rlist); library(gtools)

seed <- 01091990
set.seed(seed) #setting seed so it can be replicated
sample.size<- 5000
replicates <- 500

#generating multiple datasets of correlated
#latent constructs (y1 and y2) and validation variable (x)
data <-replicate(n = replicates,  as.data.frame(mvrnorm(n=sample.size, mu=
c(0,0,0),
                                                        Sigma =
matrix(c(1,0.1,0.7, 0.1,1,0.5, 0.7, 0.5, 1), ncol = 3))), simplify = FALSE
)
colnames <- c("y1", "y2", "x" )
data <- lapply(data, setNames, colnames)


#generating the uniform distributions (that represent ARS) correlated to x
seed <- seed +1
set.seed(seed)

#To facilitate the generation of the 8 scenarios of ARS, the scenarios
were generated
```

```
#in sets of 2.

#SET 1: cor(ARS & validation variable)= 0.25 and 0.98
#step 1. generate a vector of random numbers from a normal distribution
X <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X <- mapply(cbind, data,  X, SIMPLIFY= F)
lapply(X, cor)
X <-lapply(X,function(x) { x["y1"] <- NULL; x })
X <-lapply(X,function(x) { x["y2"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1   ,  0.25,   0.98,
         0.25,    1,   0.1,
         0.98 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y <- lapply( X, function(x) { as.matrix(x) %*% C })
lapply(Y, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y <- lapply(Y, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y, cor)

#SET 2: cor(ARS & validation variable)= -0.25 and -0.98
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)
X2 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X2 <- mapply(cbind, data,  X2, SIMPLIFY= F)
lapply(X2, cor)
X2 <-lapply(X2,function(x) { x["y1"] <- NULL; x })
X2 <-lapply(X2,function(x) { x["y2"] <- NULL; x })
```

```
#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1    ,  -0.25,  -0.98,
         -0.25 ,    1,   0.1,
         -0.98 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y2 <- lapply( X2, function(x) { as.matrix(x) %*% C })
lapply(Y2, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y2 <- lapply(Y2, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y2, cor)


#SET 3: cor(ARS & validation variable)= 0.05 and 0.55
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)

X3 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X3 <- mapply(cbind, data,  X3, SIMPLIFY= F)
lapply(X3, cor)
X3 <-lapply(X3,function(x) { x["y1"] <- NULL; x })
X3 <-lapply(X3,function(x) { x["y2"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1    ,  0.05,  0.55,
         0.05 ,    1,   0.1,
         0.55 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)for (i in 1:3){
```

```
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
#and checking correlations
C <- chol(M)
Y3 <- lapply( X3, function(x) { as.matrix(x) %*% C })
lapply(Y3, cor)

#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y3 <- lapply(Y3, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y3, cor)

#SET 4: cor(ARS & validation variable)= -0.05 and -0.55
#step 1. generate a vector of random numbers from a normal distribution
seed <- seed +1
set.seed(seed)

X4 <- replicate(n= replicates, as.data.frame(array(rnorm(sample.size*2),
dim = c(sample.size, 2))), simplify= FALSE)
X4 <- mapply(cbind, data,  X4, SIMPLIFY= F)
lapply(X4, cor)
X4 <-lapply(X4,function(x) { x["y1"] <- NULL; x })
X4 <-lapply(X4,function(x) { x["y2"] <- NULL; x })

#step 2: specifying the desired correlations (rho) with validation
variable
M <- c(  1    ,  -0.05,  -0.55,
        -0.05 ,    1,   0.1,
        -0.55 ,  0.1,    1 )
dim(M) <- c(3, 3)

#step 3: adjust the correlation to match the correlation under a Gaussian
copula
#adj.rho= 2*sin(rho*pi/6)
for (i in 1:3){
  for (j in max(i, 2):3){
    if (i != j){
      M[i, j] <- 2 * sin(pi * M[i, j] / 6)
      M[j, i] <- 2 * sin(pi * M[j, i] / 6)
    }
  }
}

#Step 4: inducing the correlations from step 3 between the
#betweem ARS variable and validation variable
```

```
#and checking correlations
C <- chol(M)
Y4 <- lapply( X4, function(x) { as.matrix(x) %*% C })
lapply(Y4, cor)


#Step 5: transform normal distributions of ARS into uniform
#distributions, check correlations
Y4 <- lapply(Y4, function(x) {x[, 2:3] <- pnorm(as.matrix(x[, 2:3])); x})
lapply(Y4, cor)

#changing names to columns in data set to represent the
#continous ARS latent constructs
Y <- lapply(Y, function(x){data.frame(x)})
colnames <- c("x", "cont.ars2", "cont.ars4")
Y <- lapply(Y, setNames, colnames)

Y2 <- lapply(Y2, function(x){data.frame(x)})
colnames <- c("x", "cont.ars6", "cont.ars8")
Y2 <- lapply(Y2, setNames, colnames)

Y3 <- lapply(Y3, function(x){data.frame(x)})
colnames <- c("x", "cont.ars1", "cont.ars3")
Y3 <- lapply(Y3, setNames, colnames)

Y4 <- lapply(Y4, function(x){data.frame(x)})
colnames <- c("x", "cont.ars5", "cont.ars7")
Y4 <- lapply(Y4, setNames, colnames)


#saving ARS latent constructs in original data set and generating ARS
indicators
#20% of the cases are selected to show acquiescent tendencies
data<- mapply(cbind, data,  Y3, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars1 = ifelse(x[["cont.ars1"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars3 = ifelse(x[["cont.ars3"]]
<= 0.8, 0, 1)) })

data<- mapply(cbind, data,  Y, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars2 = ifelse(x[["cont.ars2"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars4 = ifelse(x[["cont.ars4"]]
<= 0.8, 0, 1)) })


data<- mapply(cbind, data,  Y4, SIMPLIFY= F)
data <-lapply(data,function(x) { cbind(x, ars5 = ifelse(x[["cont.ars5"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars7 = ifelse(x[["cont.ars7"]]
<= 0.8, 0, 1)) })

data<- mapply(cbind, data,  Y2, SIMPLIFY= F)
```

```
data <-lapply(data,function(x) { cbind(x, ars6 = ifelse(x[["cont.ars6"]]
<= 0.8, 0, 1)) })
data <-lapply(data,function(x) { cbind(x, ars8 = ifelse(x[["cont.ars8"]]
<= 0.8, 0, 1)) })

#generate ars indicator for independent ARS
seed <- seed +1
set.seed(seed)
data <-lapply(data, function(x) { cbind(x, random.ars =
rbinom(sample.size, 1, 0.2) ) } ) #20% of the sample to be acquiescers


########## GENERATING ITEMS FOR CONSTRUCT #1 (Y1) ################
#Generating the parameters of the Graded Response Model
items <- 6
k <- 5 #number of categories for each item
seed<- seed +1
set.seed(seed)
#generating the difficulty thresholds
#more than needed are generated in order
#to select those with acceptable differences between them according
#to Jiang, Wang and Weiss(2016)
r <-replicate(n = replicates,  cbind.data.frame(X1=runif(10000,-1.5, -
0.75),
                                                X2=runif(10000,-0.75, 0),
                                                X3=runif(10000,0, 0.75),
                                                X4=runif(1000, 0.75,1.5)),
simplify= FALSE)

#identifying adequate generated thresholds
r <-lapply(r,function(x) {
  cbind(x, d12 = abs(x[["X2"]]- x[["X1"]]) < 0.5 | abs(x[["X2"]]-
x[["X1"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d23 = abs(x[["X3"]]- x[["X2"]]) < 0.5 | abs(x[["X3"]]-
x[["X2"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d34 = abs(x[["X4"]]- x[["X3"]]) < 0.5 | abs(x[["X4"]]-
x[["X3"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, adequate = ifelse(x[["d12"]]== FALSE & x[["d23"]]== FALSE &
x[["d34"]]== FALSE, 1, 0)) })

r <- lapply(r, function(x) {
  x <- x[x$adequate==1,]})

lapply(r, function(x) {dim(x)})

seed<- seed +1
set.seed(seed)
#randomly selecting from the adequate thresholds
```

```
sample <- lapply(r, function(x) {sample(1:dim(x)[1], items)})

r <-lapply(r,function(x) {
  cbind(x, id =seq_len(dim(x)[1]) ) })

r2 <- list()
for (i in 1:replicates) {
  r2[[i]] <-cbind(r[[i]], sample2= r[[i]]$id%in%sample[[i]])
}

r2 <- lapply(r2, function(x) {
  x <- x[x$sample2==TRUE,]})
r2 <- lapply(r2, function(x) {x[, 1:4]})

#generating discrimination parameters
a <- list()
for (i in 1:replicates){
  a[[i]]<- runif(dim(r2[[i]])[1],1.1,2.8)
}

#combining all model parameters
g.params <- list()
for (i in 1:replicates){
  g.params[[i]] <- cbind(a[[i]], r2[[i]])}
colnames <- c("a","X1", "X2", "X3", "X4")
g.params <- lapply(g.params, setNames, colnames)

#generating items, appending items to original data and reversing 3 items
g.mod<- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod[[i]] <- simIrt(theta= as.vector(data[[i]]$y1), params =
as.matrix(g.params[[i]]), mod = "grm")
  data[[i]]$item1 <- g.mod[[i]]$resp[,1]
  data[[i]]$item2 <- g.mod[[i]]$resp[,2]
  data[[i]]$item3 <- g.mod[[i]]$resp[,3]
  data[[i]]$item4 <- g.mod[[i]]$resp[,4]
  data[[i]]$item5 <- g.mod[[i]]$resp[,5]
  data[[i]]$item6 <- g.mod[[i]]$resp[,6]
  data[[i]]$item3.rev <- car::recode(data[[i]]$item3, "1=5; 2=4; 3=3; 4=2;
5=1")
  data[[i]]$item4.rev <- car::recode(data[[i]]$item4, "1=5; 2=4; 3=3; 4=2;
5=1")
  data[[i]]$item5.rev <- car::recode(data[[i]]$item5, "1=5; 2=4; 3=3; 4=2;
5=1")
}

#generate items under ARS
seed <- seed +1
set.seed(seed)
a<- c(runif(items,2, 2.8))
r1<- c(a[1],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r2<- c(a[2],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
```

```
r3<- c(a[3],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r4<- c(a[4],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r5<- c(a[5],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r6<- c(a[6],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))

g.params.ars <- rbind(r1, r2,r3,r4,r5, r6)
g.mod.ars <- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod.ars[[i]] <- simIrt(theta= as.vector(data[[i]]$y1), params =
as.matrix(g.params.ars), mod = "grm")
  data[[i]]$ars.item1 <- g.mod.ars[[i]]$resp[,1]
  data[[i]]$ars.item2 <- g.mod.ars[[i]]$resp[,2]
  data[[i]]$ars.item3 <- g.mod.ars[[i]]$resp[,3]
  data[[i]]$ars.item4 <- g.mod.ars[[i]]$resp[,4]
  data[[i]]$ars.item5 <- g.mod.ars[[i]]$resp[,5]
  data[[i]]$ars.item6 <- g.mod.ars[[i]]$resp[,6]
}

########## GENERATING ITEMS FOR CONSTRUCT #2 (Y2) ######################
#Generating the parameters of the Graded Response Model
items <- 6
k <- 5 #number of categories for each item
seed<- seed +1
set.seed(seed)
#generating the difficulty thresholds
#more than needed are generated in order
#to select those with acceptable differences between them according
#to Jiang, Wang and Weiss(2016)
r <-replicate(n = replicates,  cbind.data.frame(X1=runif(10000,-1.5, -
0.75), X2=runif(10000,-0.75, 0),
X3=runif(10000,0, 0.75),
X4=runif(1000, 0.75,1.5)), simplify= FALSE)

#identifying adequate generated thresholds
r <-lapply(r,function(x) {
  cbind(x, d12 = abs(x[["X2"]]- x[["X1"]]) < 0.5 | abs(x[["X2"]]-
x[["X1"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d23 = abs(x[["X3"]]- x[["X2"]]) < 0.5 | abs(x[["X3"]]-
x[["X2"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, d34 = abs(x[["X4"]]- x[["X3"]]) < 0.5 | abs(x[["X4"]]-
x[["X3"]]) > 2) })

r <-lapply(r,function(x) {
  cbind(x, adequate = ifelse(x[["d12"]]== FALSE & x[["d23"]]== FALSE &
x[["d34"]]== FALSE, 1, 0)) })

r <- lapply(r, function(x) {
  x <- x[x$adequate==1,]})
```

```
lapply(r, function(x) {dim(x)})

seed<- seed +1
set.seed(seed)
#randomly selecting from the adequate thresholds
sample <- lapply(r, function(x) {sample(1:dim(x)[1], items)})

r <-lapply(r,function(x) {
  cbind(x, id =seq_len(dim(x)[1]) ) })

r2 <- list()
for (i in 1:replicates) {
  r2[[i]] <-cbind(r[[i]], sample2= r[[i]]$id%in%sample[[i]])
}

r2 <- lapply(r2, function(x) {
  x <- x[x$sample2==TRUE,]})
r2 <- lapply(r2, function(x) {x[, 1:4]})
#generating discrimination parameters
a <- list()
for (i in 1:replicates){
  a[[i]]<- runif(dim(r2[[i]])[1],1.1,2.8)
}
#combining all model parameters
g.params <- list()
for (i in 1:replicates){
  g.params[[i]] <- cbind(a[[i]], r2[[i]])}
colnames <- c("a","X1", "X2", "X3", "X4")
g.params <- lapply(g.params, setNames, colnames)

#generating items, appending items to original data and reversing 3 items
g.mod<- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod[[i]] <- simIrt(theta= as.vector(data[[i]]$y2), params =
as.matrix(g.params[[i]]), mod = "grm")
  data[[i]]$item1_y2 <- g.mod[[i]]$resp[,1]
  data[[i]]$item2_y2 <- g.mod[[i]]$resp[,2]
  data[[i]]$item3_y2 <- g.mod[[i]]$resp[,3]
  data[[i]]$item4_y2 <- g.mod[[i]]$resp[,4]
  data[[i]]$item5_y2 <- g.mod[[i]]$resp[,5]
  data[[i]]$item6_y2 <- g.mod[[i]]$resp[,6]
  data[[i]]$item3.rev_y2 <- car::recode(data[[i]]$item3, "1=5; 2=4; 3=3;
4=2; 5=1")
  data[[i]]$item4.rev_y2 <- car::recode(data[[i]]$item4, "1=5; 2=4; 3=3;
4=2; 5=1")
  data[[i]]$item5.rev_y2 <- car::recode(data[[i]]$item5, "1=5; 2=4; 3=3;
4=2; 5=1")
}

#generate items under ARS
seed <- seed +1
```

```
set.seed(seed)
a<- c(runif(items,2, 2.8))
r1<- c(a[1],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r2<- c(a[2],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r3<- c(a[3],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r4<- c(a[4],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r5<- c(a[5],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))
r6<- c(a[6],qnorm(.03), qnorm(.10), qnorm(.20), qnorm(.40))

g.params.ars <- rbind(r1, r2,r3,r4,r5, r6)
g.mod.ars <- list()
for (i in 1:replicates){
  seed<- seed +1
  set.seed(seed)
  g.mod.ars[[i]] <- simIrt(theta= as.vector(data[[i]]$y2), params =
as.matrix(g.params.ars), mod = "grm")
  data[[i]]$ars.item1_y2 <- g.mod.ars[[i]]$resp[,1]
  data[[i]]$ars.item2_y2 <- g.mod.ars[[i]]$resp[,2]
  data[[i]]$ars.item3_y2 <- g.mod.ars[[i]]$resp[,3]
  data[[i]]$ars.item4_y2 <- g.mod.ars[[i]]$resp[,4]
  data[[i]]$ars.item5_y2 <- g.mod.ars[[i]]$resp[,5]
  data[[i]]$ars.item6_y2 <- g.mod.ars[[i]]$resp[,6]
}

#create variables containing the items with ARS for flagged cases

##### CONSTRUCT 1 #######
data <-lapply(data, function(x) { cbind(x, i1_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item6"]], x[["item6"]]))})

data <-lapply(data, function(x) { cbind(x, i1_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ran_rev = ifelse
(x[["random.ars"]]==1 , x[["ars.item6"]], x[["item6"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i1_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars2_norev = ifelse
(x[["ars2"]]==1 ,  x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i6_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item1"]],  x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item6"]],  x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i4_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item2"]],  x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item2"]], x[["item2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i3_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item5"]],  x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item3"]],  x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item1"]], x[["item1"]]))})
data <-lapply(data, function(x) { cbind(x, i2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item3"]], x[["item3"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item4"]], x[["item4"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item5"]], x[["item5"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item6"]], x[["item6"]]))})


data <-lapply(data, function(x) { cbind(x, i1_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item1"]], x[["item1"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item2"]], x[["item2"]]))})
data <-lapply(data, function(x) { cbind(x, i3_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item3"]], x[["item3.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i4_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item4"]], x[["item4.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i5_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item5"]], x[["item5.rev"]]))})
data <-lapply(data, function(x) { cbind(x, i6_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item6"]],  x[["item6"]]))})

######################  CONSTRUCT 2 ####################################
data <-lapply(data, function(x) { cbind(x, i1y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ran_norev = ifelse
(x[["random.ars"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <- lapply(data, function(x) {cbind(x, i1y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <- lapply(data, function(x) {cbind(x, i2y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <- lapply(data, function(x) {cbind(x, i3y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <- lapply(data, function(x) {cbind(x, i4y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <- lapply(data, function(x) {cbind(x, i5y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <- lapply(data, function(x) {cbind(x, i6y2_ran_rev =
ifelse(x[["random.ars"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars1_norev = ifelse
(x[["ars1"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i1y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars1_rev = ifelse
(x[["ars1"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars2_norev = ifelse
(x[["ars2"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars2_rev = ifelse
(x[["ars2"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars3_norev = ifelse
(x[["ars3"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i1y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars3_rev = ifelse
(x[["ars3"]]==1 , x[["ars.item6_y2"]],  x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars4_norev = ifelse
(x[["ars4"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars4_rev = ifelse
(x[["ars4"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i6y2_ars5_norev = ifelse
(x[["ars5"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars5_rev = ifelse
(x[["ars5"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars6_norev = ifelse
(x[["ars6"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})

data <-lapply(data, function(x) { cbind(x, i1y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars6_rev = ifelse
(x[["ars6"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
```

```
data <-lapply(data, function(x) { cbind(x, i5y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars7_norev = ifelse
(x[["ars7"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars7_rev = ifelse
(x[["ars7"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


data <-lapply(data, function(x) { cbind(x, i1y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item3_y2"]], x[["item3_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item4_y2"]], x[["item4_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item5_y2"]], x[["item5_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars8_norev = ifelse
(x[["ars8"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})

data <-lapply(data, function(x) { cbind(x, i1y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item1_y2"]], x[["item1_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i2y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item2_y2"]], x[["item2_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i3y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item3_y2"]], x[["item3.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i4y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item4_y2"]], x[["item4.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i5y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item5_y2"]], x[["item5.rev_y2"]]))})
data <-lapply(data, function(x) { cbind(x, i6y2_ars8_rev = ifelse
(x[["ars8"]]==1 , x[["ars.item6_y2"]], x[["item6_y2"]]))})


#################### SCORES CONSTRUCT 1 ##################
data <- lapply(data, function(x) {cbind(x, score = x[["item1"]] +
x[["item2"]] + x[["item3"]] + x[["item4"]] + x[["item5"]] +
x[["item6"]])})
data <- lapply(data, function(x) {cbind(x, score_rev = x[["item1"]] +
x[["item2"]] + 6-x[["item3.rev"]] + 6- x[["item4.rev"]] + 6-
x[["item5.rev"]] + x[["item6"]])})
```

```
data <- lapply(data, function(x) {cbind(x, sran_norev =
x[["i1_ran_norev"]] + x[["i2_ran_norev"]] + x[["i3_ran_norev"]] +
x[["i4_ran_norev"]] + x[["i5_ran_norev"]] + x[["i6_ran_norev"]])})
data <- lapply(data, function(x) {cbind(x, sran_rev = x[["i1_ran_rev"]] +
x[["i2_ran_rev"]] + 6-x[["i3_ran_rev"]] + 6- x[["i4_ran_rev"]] + 6-
x[["i5_ran_rev"]] + x[["i6_ran_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars1_norev =
x[["i1_ars1_norev"]] + x[["i2_ars1_norev"]] + x[["i3_ars1_norev"]] +
x[["i4_ars1_norev"]] + x[["i5_ars1_norev"]] + x[["i6_ars1_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars1_rev = x[["i1_ars1_rev"]]
+ x[["i2_ars1_rev"]] + 6-x[["i3_ars1_rev"]] + 6- x[["i4_ars1_rev"]] + 6-
x[["i5_ars1_rev"]] + x[["i6_ars1_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars2_norev =
x[["i1_ars2_norev"]] + x[["i2_ars2_norev"]] + x[["i3_ars2_norev"]] +
x[["i4_ars2_norev"]] + x[["i5_ars2_norev"]] + x[["i6_ars2_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars2_rev = x[["i1_ars2_rev"]]
+ x[["i2_ars2_rev"]] + 6-x[["i3_ars2_rev"]] + 6- x[["i4_ars2_rev"]] + 6-
x[["i5_ars2_rev"]] + x[["i6_ars2_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars3_norev =
x[["i1_ars3_norev"]] + x[["i2_ars3_norev"]] + x[["i3_ars3_norev"]] +
x[["i4_ars3_norev"]] + x[["i5_ars3_norev"]] + x[["i6_ars3_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars3_rev = x[["i1_ars3_rev"]]
+ x[["i2_ars3_rev"]] + 6-x[["i3_ars3_rev"]] + 6- x[["i4_ars3_rev"]] + 6-
x[["i5_ars3_rev"]] + x[["i6_ars3_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars4_norev =
x[["i1_ars4_norev"]] + x[["i2_ars4_norev"]] + x[["i3_ars4_norev"]] +
x[["i4_ars4_norev"]] + x[["i5_ars4_norev"]] + x[["i6_ars4_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars4_rev = x[["i1_ars4_rev"]]
+ x[["i2_ars4_rev"]] + 6-x[["i3_ars4_rev"]] + 6- x[["i4_ars4_rev"]] + 6-
x[["i5_ars4_rev"]] + x[["i6_ars4_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars5_norev =
x[["i1_ars5_norev"]] + x[["i2_ars5_norev"]] + x[["i3_ars5_norev"]] +
x[["i4_ars5_norev"]] + x[["i5_ars5_norev"]] + x[["i6_ars5_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars5_rev = x[["i1_ars5_rev"]]
+ x[["i2_ars5_rev"]] + 6-x[["i3_ars5_rev"]] + 6- x[["i4_ars5_rev"]] + 6-
x[["i5_ars5_rev"]] + x[["i6_ars5_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars6_norev =
x[["i1_ars6_norev"]] + x[["i2_ars6_norev"]] + x[["i3_ars6_norev"]] +
x[["i4_ars6_norev"]] + x[["i5_ars6_norev"]] + x[["i6_ars6_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars6_rev = x[["i1_ars6_rev"]]
+ x[["i2_ars6_rev"]] + 6-x[["i3_ars6_rev"]] + 6- x[["i4_ars6_rev"]] + 6-
x[["i5_ars6_rev"]] + x[["i6_ars6_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars7_norev =
x[["i1_ars7_norev"]] + x[["i2_ars7_norev"]] + x[["i3_ars7_norev"]] +
x[["i4_ars7_norev"]] + x[["i5_ars7_norev"]] + x[["i6_ars7_norev"]])})
```

```
data <- lapply(data, function(x) {cbind(x, sars7_rev = x[["i1_ars7_rev"]]
+ x[["i2_ars7_rev"]] + 6-x[["i3_ars7_rev"]] + 6- x[["i4_ars7_rev"]] + 6-
x[["i5_ars7_rev"]] + x[["i6_ars7_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars8_norev =
x[["i1_ars8_norev"]] + x[["i2_ars8_norev"]] + x[["i3_ars8_norev"]] +
x[["i4_ars8_norev"]] + x[["i5_ars8_norev"]] + x[["i6_ars8_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars8_rev = x[["i1_ars8_rev"]]
+ x[["i2_ars8_rev"]] + 6-x[["i3_ars8_rev"]] + 6- x[["i4_ars8_rev"]] + 6-
x[["i5_ars8_rev"]] + x[["i6_ars8_rev"]])})

###################### SCORES CONSTRUCT 2 ######################
data <- lapply(data, function(x) {cbind(x, score_y2 = x[["item1_y2"]] +
x[["item2_y2"]] + x[["item3_y2"]] + x[["item4_y2"]] + x[["item5_y2"]] +
x[["item6_y2"]])})
data <- lapply(data, function(x) {cbind(x, score_rev_y2 = x[["item1_y2"]]
+ x[["item2_y2"]] + 6-x[["item3.rev_y2"]] + 6- x[["item4.rev_y2"]] + 6-
x[["item5.rev_y2"]] + x[["item6_y2"]])})

data <- lapply(data, function(x) {cbind(x, sran_norev_y2 =
x[["i1y2_ran_norev"]] + x[["i2y2_ran_norev"]] + x[["i3y2_ran_norev"]] +
x[["i4y2_ran_norev"]] + x[["i5y2_ran_norev"]] + x[["i6y2_ran_norev"]])})
data <- lapply(data, function(x) {cbind(x, sran_rev_y2 =
x[["i1y2_ran_rev"]] + x[["i2y2_ran_rev"]] + 6-x[["i3y2_ran_rev"]] + 6-
x[["i4y2_ran_rev"]] + 6- x[["i5y2_ran_rev"]] + x[["i6y2_ran_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars1_norev_y2 =
x[["i1y2_ars1_norev"]] + x[["i2y2_ars1_norev"]] + x[["i3y2_ars1_norev"]] +
x[["i4y2_ars1_norev"]] + x[["i5y2_ars1_norev"]] +
x[["i6y2_ars1_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars1_rev_y2 =
x[["i1y2_ars1_rev"]] + x[["i2y2_ars1_rev"]] + 6-x[["i3y2_ars1_rev"]] + 6-
x[["i4y2_ars1_rev"]] + 6- x[["i5y2_ars1_rev"]] + x[["i6y2_ars1_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars2_norev_y2 =
x[["i1y2_ars2_norev"]] + x[["i2y2_ars2_norev"]] + x[["i3y2_ars2_norev"]] +
x[["i4y2_ars2_norev"]] + x[["i5y2_ars2_norev"]] +
x[["i6y2_ars2_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars2_rev_y2 =
x[["i1y2_ars2_rev"]] + x[["i2y2_ars2_rev"]] + 6-x[["i3y2_ars2_rev"]] + 6-
x[["i4y2_ars2_rev"]] + 6- x[["i5y2_ars2_rev"]] + x[["i6y2_ars2_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars3_norev_y2 =
x[["i1y2_ars3_norev"]] + x[["i2y2_ars3_norev"]] + x[["i3y2_ars3_norev"]] +
x[["i4y2_ars3_norev"]] + x[["i5y2_ars3_norev"]] +
x[["i6y2_ars3_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars3_rev_y2 =
x[["i1y2_ars3_rev"]] + x[["i2y2_ars3_rev"]] + 6-x[["i3y2_ars3_rev"]] + 6-
x[["i4y2_ars3_rev"]] + 6- x[["i5y2_ars3_rev"]] + x[["i6y2_ars3_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars4_norev_y2 =
x[["i1y2_ars4_norev"]] + x[["i2y2_ars4_norev"]] + x[["i3y2_ars4_norev"]] +
x[["i4y2_ars4_norev"]] + x[["i5y2_ars4_norev"]] +
x[["i6y2_ars4_norev"]])})
```

```
data <- lapply(data, function(x) {cbind(x, sars4_rev_y2 =
x[["i1y2_ars4_rev"]] + x[["i2y2_ars4_rev"]] + 6-x[["i3y2_ars4_rev"]] + 6-
x[["i4y2_ars4_rev"]] + 6- x[["i5y2_ars4_rev"]] + x[["i6y2_ars4_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars5_norev_y2 =
x[["i1y2_ars5_norev"]] + x[["i2y2_ars5_norev"]] + x[["i3y2_ars5_norev"]] +
x[["i4y2_ars5_norev"]] + x[["i5y2_ars5_norev"]] +
x[["i6y2_ars5_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars5_rev_y2 =
x[["i1y2_ars5_rev"]] + x[["i2y2_ars5_rev"]] + 6-x[["i3y2_ars5_rev"]] + 6-
x[["i4y2_ars5_rev"]] + 6- x[["i5y2_ars5_rev"]] + x[["i6y2_ars5_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars6_norev_y2 =
x[["i1y2_ars6_norev"]] + x[["i2y2_ars6_norev"]] + x[["i3y2_ars6_norev"]] +
x[["i4y2_ars6_norev"]] + x[["i5y2_ars6_norev"]] +
x[["i6y2_ars6_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars6_rev_y2 =
x[["i1y2_ars6_rev"]] + x[["i2y2_ars6_rev"]] + 6-x[["i3y2_ars6_rev"]] + 6-
x[["i4y2_ars6_rev"]] + 6- x[["i5y2_ars6_rev"]] + x[["i6y2_ars6_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars7_norev_y2 =
x[["i1y2_ars7_norev"]] + x[["i2y2_ars7_norev"]] + x[["i3y2_ars7_norev"]] +
x[["i4y2_ars7_norev"]] + x[["i5y2_ars7_norev"]] +
x[["i6y2_ars7_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars7_rev_y2 =
x[["i1y2_ars7_rev"]] + x[["i2y2_ars7_rev"]] + 6-x[["i3y2_ars7_rev"]] + 6-
x[["i4y2_ars7_rev"]] + 6- x[["i5y2_ars7_rev"]] + x[["i6y2_ars7_rev"]])})

data <- lapply(data, function(x) {cbind(x, sars8_norev_y2 =
x[["i1y2_ars8_norev"]] + x[["i2y2_ars8_norev"]] + x[["i3y2_ars8_norev"]] +
x[["i4y2_ars8_norev"]] + x[["i5y2_ars8_norev"]] +
x[["i6y2_ars8_norev"]])})
data <- lapply(data, function(x) {cbind(x, sars8_rev_y2 =
x[["i1y2_ars8_rev"]] + x[["i2y2_ars8_rev"]] + 6-x[["i3y2_ars8_rev"]] + 6-
x[["i4y2_ars8_rev"]] + 6- x[["i5y2_ars8_rev"]] + x[["i6y2_ars8_rev"]])})

#saving data file
save(data, file= "data.RData")
```

A3 Syntax for the use of ARS correction methods

```
##########################################################
#Example of use of correction methods for ARS correction
#Data used in this example: sample of respondents from Mexico
#relevant variables:
#swl_bal: scale score for Satisfaction with Life (SWL)
#count: count of agreements variable (proxy of ARS)
#swl_bal1-swl_bal5: individual items of the SWL scale
#pl1-pl7: individual items of the Purpose in Life scale
#ee: scale score for Emotional Expressivity scale
#dep: scale score for Depression Symptoms scale
########################################################

rm(list=ls(all=T))
library(lavaan); library(stats); library(mirt)

mex_data = read.csv("mexico_cfa_data_ch2.csv",header=T)

#OLS adjustment
#Step 1. Estimating the model
sum.mod <- summary(glm(swl_bal ~ count, data = mex_data))

#Step 2. Computing corrected scores (swl_ols_adj)
mex_data$swl_ols_adj <- mex_data$swl_bal-
sum.mod$coefficients[2]*mex_data$count2


#CFA adjustments
#Savalei & Falk adjustment
#Step 1. Estimating the model
swl_bal1 <- 'swl =~ swl_bal2 + swl_bal4 + swl_bal3  + swl_bal1
             ars =~ 1*swl_bal2 + 1*swl_bal4 + 1*swl_bal3 + 1*swl_bal1'
fit1 <- cfa(swl_bal1, data= mex_data, missing= "ML")

#Step 2. Computing factor scores for SWL (swl_cfaSF) and ARS (ars_cfaSF)
mex_data$swl_cfaSF <- predict(cfa(fit2, data= mex_data, missing=
"ML"))[,1]
mex_data$ars_cfaSF <- predict(cfa(fit2, data= mex_data, missing=
"ML"))[,2]

#Billiet & McClendon adjustment
#Step 1. Estimating the model
swl_bal2 <- 'swl =~ swl_bal2 + swl_bal4 + swl_bal3 + swl_bal5 + swl_bal1
            pll =~ pl1 + pl2 + pl3 + pl4 + pl5 + pl6 +pl7
            ars =~ 1*swl_bal2 + 1*swl_bal4 + 1*swl_bal3 + 1*swl_bal5 +
1*swl_bal1 +
                     1*pl1 +1*pl2 +1*pl3 +1*pl4 + 1*pl5 + 1*pl6 + 1*pl7
            swl~~pll'
fit3 <- cfa(swl_bal2, data= mex_data, missing= "ML")

#Step 2. Computing factor scores for SWL (swl_cfaBM) and ARS (ars_cfaBM)
mex_data$swl_cfaBM <- predict(cfa(fit3, data= mex_data, missing=
"ML"))[,1]
```

```
mex_data$ars_cfaBM <- predict(cfa(fit3, data= mex_data, missing=
"ML"))[,3]


#Multidimensional Nominal Response Model
#Step 1. Create the scoring function for each item and reverse the codes
#when necessary
sf <-list()
sf[[1]]<-matrix(
  c(0,1,2,3,4,5,6, # swl
    0,0,0,0, 1,1,1 # ars
  ), 7, 2)

sf[[2]]<-matrix(
  c(0,1,2,3,4,5,6, # swl
    0,0,0,0, 1,1,1 # ars
  ), 7, 2)

sf[[3]]<-matrix(
  c(6,5,4,3,2,1,0, # swl (reversed item)
    0,0,0,0, 1,1,1 # ars
  ), 7, 2)

sf[[4]]<-matrix(
  c(0,1,2,3,4,5,6, # swl
    0,0,0,0, 1,1,1 # ars
  ), 7, 2)


sf2[[5]]<-matrix(
  c(6,5,4,3,2,1,0, # swl (reversed item)
    0,0,0,0, 1,1,1 # ars
  ), 7, 2)

#Step 2. Define the mode with all items loading on both dimensions (SWL
#and ARS). The numbers 1 to 5 indicate the number of the column for each
#item (e.g., 1= column 1)
mod.swl1<-"
swl2 = 1-5
ars = 1-5
COV = swl*ars
"

#Step 3. Create data set with only the needed items and in the order
specified in the model
#and the scoring functions
mex.mod1 <- subset(mex_data, select= c(swl_bal1, swl_bal2, swl_bal3,
swl_bal4, swl_bal5, ee, id, dep))

#Step 4. Estimate the model
fit.mod1 <-
mirt(mex.mod1,mod.swl1,itemtype="gpcm",gpcm_mats=sf,technical=list(NCYCLES
=1000, removeEmptyRows=TRUE))
```

```
#Step 5. Compute factor scores for SWL (mnrm_adj) and ARS (mnrm_ars)
mex.mod1$mnrm_adj <- fscores(fit.mod1, method= "EAP", full.scores = T,
full.scores.SE = F)[,1]
mex.mod1$mnrm_ars <- fscores(fit.mod1, method= "EAP", full.scores = T,
full.scores.SE = F)[,2]
```

Table A 1 Correlations between the Agreeableness measure and ARS latent factors from Billiet and McCledon's CFA model, Savalei and Falk's CFA model and MNRM.

All respondents (n= 2,363)				
	Agreeableness variable	ARS factor -Billiet & McClendon's CFA	ARS factor - Savalei & Falk's CFA	ARS factor - MNRM
Agreeableness variable	1.000			
ARS factor -Billiet and McClendon's CFA	0.624	1.000		
ARS factor - Savalei and Falk's CFA	0.472	0.850	1.000	
ARS factor- MNRM	0.451	0.678	0.814	1.000
Non-Hispanic White (n= 791)				
	Agreeableness variable	ARS factor -Billiet & McClendon's CFA	ARS factor - Savalei & Falk's CFA	ARS factor - MNRM
Agreeableness variable	1.000			
ARS factor -Billiet & McClendon's CFA	0.642	1.000		
ARS factor - Savalei & Falk's CFA	0.526	0.850	1.000	
ARS factor- MNRM	0.489	0.669	0.741	1.000
Hispanic in Mexico (n= 795)				
	Agreeableness variable	ARS factor -Billiet & McClendon's CFA	ARS factor - Savalei & Falk's CFA	ARS factor - MNRM
Agreeableness variable	1.000			
ARS factor -Billiet & McClendon's CFA	0.492	1.000		
ARS factor - Savalei & Falk's CFA	0.302	0.762	1.000	
ARS factor- MNRM	0.341	0.579	0.670	1.000
Hispanic in the US (n= 777)				

	Agreeableness variable	ARS factor -Billiet & McClendon's CFA	ARS factor - Savalei & Falk's CFA	ARS factor - MNRM
Agreeableness variable	1.000			
ARS factor -Billiet & McClendon's CFA	0.602	1.000		
ARS factor - Savalei & Falk's CFA	0.435	0.746	1.000	
ARS factor- MNRM	0.455	0.685	0.745	1.000

Appendix C. Additional Materials for Chapter 4

Table A 2 Scale experimental wording: Satisfaction with life

Item number	**Control**	**Negated**	**Opposite**
Item 1	In most ways my life is close to my ideal	In most ways my life is not close to my ideal	In most ways my life is far from my ideal
Item 2	The conditions of my life are excellent		
Item 3	I am satisfied with my life	I am not satisfied with my life	I am unhappy with my life
Item 4	So far I have gotten the important things I want in life		
Item 5	If I could live my life over, I would change almost nothing		

Table A 3 Scale experimental wording: Sense of Control

Item number	**Control**	**Negated**	**Opposite**
Perceived Mastery			
Item 1	I can do just about anything I really set my mind to		
Item 2	When I really want to do something, I usually find a way to succeed at it		
Item 3	Whether I am able to get what I want is in my own hands	Whether I am able to get what I want is not in my own hands	Whether I am able to get what I want is out of my own hands

Item 4	What happens to me in the future mostly depends on me	What happens to me in the future does not depend on me	What happens to me in the future is beyond my control
Perceived Constraints			
Item 1	There is little I can do to change the important things in my life		
Item 2	I often feel paralyzed in dealing with the problems of life	I don't often feel paralyzed in dealing with the problems of life	I usually feel like I am able to deal with the problems of life
Item 3	Other people determine most of what I can do	Other people don't determine most of what I can do	I myself determine what I can do
Item 4	What happens in my life is often beyond my control	What happens in my life is not often beyond my control	What happens in my life is often within my control
Item 5	There are many things that interfere with what I want to do	There are not many things that interfere with what I want to do	There are few things that interfere with what I want to do
Item 6	I have little control over the things that happen to me		
Item 7	I sometimes feel I am being pushed around my life		

Table A 4 Scale experimental wording: Need for Affect

Item number	**Control**	**Negated**	**Opposite**
Item 1	It is important for me to be in touch with my feelings		
Item 2	I think that it is important to explore my feelings		
Item 3	I am a very emotional person		
Item 4	It is important for me to know how others are feeling	It is not important for me to know how others are feeling	I give little importance to know how others are feeling
Item 5	Emotions help people get along in life		
Item 6	Strong emotions are generally helpful	Strong emotions are not generally helpful	Strong emotions are generally harmful
Item 7	I feel that I need to experience strong emotions regularly		
Item 8	I approach situations in which I expect to experience strong emotions	I don’t approach situations in which I expect to experience strong emotions	I avoid situations in which I expect strong emotions
Item 9	I am comfortable with experiencing strong emotions	I am not comfortable with experiencing strong emotions	I feel uncomfortable with experiencing strong emotions
Item 10	I enjoy experiencing strong emotions		
Item 11	I feel like I need a good cry every now and then		
Item 12	I like to dwell on my emotions	I don’t like to dwell on my emotions	I dislike dwelling on my emotions
Item 13	We should indulge ourselves in experiencing our emotions	We should not indulge ourselves in experiencing our emotions	We should limit ourselves in experiencing our emotions

Item 14	I like decorating my spaces with lots of pictures or other images of things that are emotionally important to me		
Item 15	The experience of emotions promotes human survival		

Table A 5 Scale experimental wording: Social Provisions

Item number	**Control**	**Negated**	**Opposite**
Item 1	There are people I can depend on to help me if I really need it		
Item 2	There are people who enjoy the same social activities I do		
Item 3	I have close relationships that provide me with a sense of emotional security and well-being	I don't have close relationships that provide me with a sense of emotional security and well-being	I lack close relationships that provide me with a sense of emotional security and well-being
Item 4	There is someone I could talk to about important decisions in my life		
Item 5	I have relationships where my competence and skill are recognized	I don't have relationships where my competence and skill are recognized	I lack relationships where my competence and skill are recognized
Item 6	There is a trustworthy person I could turn to for advice if I were having problems	There is not a trustworthy person I could turn to for advice if I were having problems	I lack a trustworthy person I could turn to for advice if I were having problems
Item 7	I feel part of a group of people who share my	I don't feel part of a group of people who share my	I feel alone in my attitudes and beliefs

	attitudes and beliefs	attitudes and beliefs	
Item 8	I feel a strong emotional bond with at least one other person	I don't feel a strong emotional bond with even one other person	I lack a strong emotional bond with even one other person
Item 9	There are people who admire my talents and abilities		
Item 10	There are people I can count on in an emergency		

A 4 Wording for validation variables and scales

a) Self-Rated Health

Now we want to ask about your health. In general, would you say your health is...

Excellent

Very Good

Good

Fair

Poor

b) Depression symptoms

The following questions ask about how you have been feeling since March 2020. For each question, please circle the number that best describes how often you had this feeling. Since March 2020, about how often has the coronavirus and its impacts made you feel...

1) … nervous?
2) … hopeless?
3) … restless or fidgety?

4) … so depressed that nothing could cheer you up?

5) … that everything was an effort?

6) … worthless?

Response categories: Strongly Disagree, Slightly Disagree, Disagree, Neither Agree nor Disagree, Agree, Slightly Agree, Strongly Agree.

c) Purpose in Life

1) I enjoy making plans for the future and working to make them a reality.

2) My daily activities often seem trivial and unimportant to me.

3) I am an active person in carrying out the plans I set for myself.

4) I don't have a good sense of what it is I'm trying to accomplish in life.

5) I sometimes feel as if I've done all there is to do in life.

6) I live life one day at a time and don't really think about the future.

7) I have a sense of direction and purpose in my life.

Response categories: Strongly Disagree, Slightly Disagree, Disagree, Neither Agree nor Disagree, Agree, Slightly Agree, Strongly Agree.

d) Income

What was your annual personal income in 2020? By income we mean income from your own wages, salary, or other sources, before taxes.

Response categories: Less than $30 000, $30 000 to $39 999, $40 000 to $49 999, $50 000 to $59 999, $60 000 to $69 000, $70 000 to $79 999, $80 000 to $89 999, $90 000 to $99 999, $100 000 or more.

e) Education

What is the highest grade or degree that you have completed?

Response categories: None, preschool or kindergarten, Elementary/primary school (grades 1-5), Middle school/junior high (grades 6-8), Highschool diploma or equivalent (GED), Trade school/vocational school certificate, 2-year university or Associate degree (e.g., AA, AS), Some university/college no degree, 4-year university/college graduate, Graduate or professional degree, Other.

f) Emotional expressivity

1) I think of myself as emotionally expressive.
2) People think of me as an unemotional person.
3) I keep my feelings to myself.
4) I am often considered indifferent by others.
5) People can read my emotions.
6) I display my emotions to other people.
7) I don’t like to let other people see how I am feeling.
8) I am able to cry in front of other people.
9) Even if I am feeling very emotional, I don’t let other see my feelings.
10) Other people aren’t easily able to observe what I am feeling.
11) Even when I am experiencing strong feelings, I don’t express them outwardly.
12) I can’t hide the way I am feeling.
13) Other people believe me to be very emotional.
14) I don’t express my emotions to other people.
15) The way I feel is different from how others think I feel.

16) I hold my feelings in.

Response categories: Strongly Disagree, Slightly Disagree, Disagree, Neither Agree nor Disagree, Agree, Slightly Agree, Strongly Agree.

g) Affective Orientation

1) I am very aware of my feelings.
2) I use my feelings and emotions to determine what I should do in situations.
3) My feelings and emotions are very important to me.
4) I listen to what my “gut” or “hear” says in many situations.
5) My emotions tell me what to do in many cases.
6) I try not to let feelings guide my actions.
7) I trust my feelings to guide my behavior.
8) I don’t pay much attention to my emotions most of the time.
9) My feelings tell me a lot about how to act in a given situation.
10) The intensity of my emotions does not change much from situation to situation.
11) I use my feelings to determine whether to trust another person.
12) I learn a lot about myself on the basis of my feelings.
13) I am not usually aware of my feelings at any given moment.
14) Feelings are a valuable source of information.
15) My feelings don’t seem to be very intense or strong.
16) I use feelings to guide me more than most people do.
17) Feelings only interfere with behavior.

18) My emotions have many different levels of intensity; I can be very angry, for example, or very angry.

19) I seem to have just few basic emotions.

Response categories: Strongly Disagree, Slightly Disagree, Disagree, Neither Agree nor Disagree, Agree, Slightly Agree, Strongly Agree.

h) Gender

What is your gender?

Response categories: Male, Female, Other.

i) Marital status

Are you currently married, separated, divorced, widowed, or have you never been married?

Response categories: Married, Separated, Divorced, Widowed, Never married, Other.

A 5 Items used for the balanced pair measure

Pair two (Affective Orientation):

"I am very aware of my feelings"

"I am not usually aware of my feelings at any given moment"

Pair three (Affective Orientation):

"The intensity of my emotions does not change much from situation from situation"

"My emotions have many different levels of intensity; I can be angry, for example, or very angry"

Pair four (Emotional expressivity):

"I display my emotions to other people"

"I don't express my emotions to other people"

Pair five (Emotional expressivity):

"People think of me as an unemotional person"

"Other people believe me to be very emotional"

Pair six (Emotional expressivity):

"I can't hide the way I am feeling"

"Even if I am feeling very emotional, I don't let others see my feelings"

Pair seven (Purpose in Life):

"I don't have a good sense of what it is I am trying to accomplish in life"

"I have a sense of direction and purpose in my life"

Table A 6 CFA fit measures for Satisfaction with Life models

Measure	**NHW**[a]	**HMex**[b]	**HUS**[c]
CFI			
Unbalanced	0.992	0.995	0.991
Negated	0.815	0.861	0.690
Negated + ARS	1.000	1.000	0.996
Polar Opposite	0.861	0.850	0.740
Polar Opposite + ARS	0.998	1.000	0.997
RMSEA			
Unbalanced	0.070	0.050	0.066
Negated	0.233	0.166	0.233
Negated + ARS	0.000	0.000	0.033
Polar Opposite	0.207	0.172	0.205
Polar Opposite + ARS	0.000	0.000	0.030
AIC			
Unbalanced	7020.8	6190.3	6839.4
Negated	7523.3	6799.4	7636.8
Negated + ARS	7418.5	6746.2	7532.5
Polar Opposite	7855.1	6911.8	7954.2
Polar Opposite + ARS	6177.3	6852.4	7872.6
BIC			
Unbalanced	7080.8	6250.1	6899.3
Negated	7582.8	6859.2	7696.5
Negated + ARS	7485.9	6813.9	7600.2
Polar Opposite	7915.2	6971.8	8014.2
Polar Opposite + ARS	6233.4	6920.5	7940.5

[a] For the Polar Opposite + ARS model, item 5 of the scale was not included in the analysis as it produced a non-positive definite variance-covariance matrix of the estimated parameters.
[b]For the Negated + ARS model, the correlation between ARS and SWL was specified in the model and freely estimated.
[c] For the Polar Opposite + ARS model, the correlation between ARS and SWL was specified in the model and freely estimated.

Table A 7 CFA fit measures for Sense of Control: Perceived Control models

Measure	**NHW**[a]	**HMex**
CFI		
Unbalanced	0.936	0.977
Negated	0.701	0.549
Negated + ARS	0.958	0.975
Polar Opposite	0.618	0.623
Polar Opposite + ARS	0.964	0.978
RMSEA		
Unbalanced	0.136	0.070
Negated	0.204	0.201
Negated + ARS	0.087	0.051
Polar Opposite	0.229	0.207
Polar Opposite + ARS	0.076	0.054
AIC		
Unbalanced	10117.8	8649.4
Negated	9117.8	9753.3
Negated + ARS	8990.9	9543.0
Polar Opposite	10569.7	9411.6
Polar Opposite + ARS	10301.2	9178.7
BIC		
Unbalanced	10201.2	8732.6
Negated	9189.6	9836.9
Negated + ARS	9070.8	9634.7
Polar Opposite	10653.9	9496.2
Polar Opposite + ARS	10393.4	9271.3

[a] Item 5 of the SoC PC **negated** scale was not included in these analyses as it produced a non-positive definite variance-covariance matrix of the estimated parameters.

Table A 8 CFA fit measures for Need for Affect models

Measure	**NHW**	**HMex**
CFI		
Unbalanced	0.860	0.813
Negated	0.653	0.597
Negated + ARS	0.852	0.830
Polar Opposite	0.587	0.417
Polar Opposite + ARS	0.870	0.793
RMSEA		
Unbalanced	0.106	0.124
Negated	0.141	0.136
Negated + ARS	0.093	0.090
Polar Opposite	0.167	0.155
Polar Opposite + ARS	0.094	0.094
AIC		
Unbalanced	20074.1	18471.7
Negated	21775.4	19560.1
Negated + ARS	21367.9	19179.7
Polar Opposite	22073.9	20179.1
Polar Opposite + ARS	21388.6	19616.1
BIC		
Unbalanced	20253.5	18651.2
Negated	21955.2	19739.3
Negated + ARS	21555.6	19366.9
Polar Opposite	22110.8	20359.2
Polar Opposite + ARS	21576.3	19804.2