**Examining the Effectiveness of Anchoring Vignettes in Personality Research**

by

**Ni Jie**

A thesis submitted to the graduate faculty in partial fulfillment of the requirements for the degree of MASTER’S ALONG THE WAY

Department of Psychology

Program of Study Committee:

Stephen Stark, Major Professor

Brenton M. Wiernik

Marina A. Bornovalova

University of South Florida

2021

TABLE OF CONTENTS

LIST OF TABLES………………………………………………………………………………...3

LIST OF FIGURES……………………………………………………………………………….4

ABSTRACT………………………………………………………………………….…………...5

CHAPTER 1 INTRODUCTION………………………………………………………………….6

CHAPTER 2 LITERATURE REVIEW…………………………………………………………..9

CHAPTER 3 METHOD………………………………………………………………………....13

CHAPTER 4 RESULTS…………………………………………………………………………17

CHAPTER 5 DISCUSSION……………………………………………………………………..22

REFERENCES…………………………………………………………………………………..27

APPENDIX A. HEALTH BEHAVIOR CHECKLIST………………………………………….44

APPENDIX B. UTRECHT WORK ENGAGEMENT SCALE FOR STUDENTS……………..46

APPENDIX C. BIG FIVE INVENTORY-2……………………………………………………..47

APPENDIX D. SELF-RATED ANCHORING VIGNETTES…………………………………..51

APPENDIX E. DEMOGRAPHIC QUESTIONS………………………………………………..65

TABLE OF TABLES

Table 1. Means, SD, and correlations with CI for control condition………………………….…33

Table 2. Means, SD, and correlations with CI for fixed vignettes condition……………...…….34

Table 3. Means, SD, and correlations with CI for self-rated vignettes condition…………….….35

Table 4. t-test results comparing means between control and fixed vignette condition…...…….36

Table 5. F-test results comparing SDs between control and fixed vignette condition ………….37

TABLE OF FIGURES

Figure 1. The Procedure Flow Chart …………………………………………………………….38

Figure 2a. Censored Ordered Probit Allocation Graph for Conscientiousness ………………....39

Figure 2b. Censored Ordered Probit Allocation Graph for Agreeableness ………………..........40

Figure 2c. Censored Ordered Probit Allocation Graph for Extraversion ……………….............41

Figure 2d. Censored Ordered Probit Allocation Graph for Openness ………………..................42

Figure 2e. Censored Ordered Probit Allocation Graph for Negative Emotionality …………….43

ABSTRACT

The Big Five (Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability) is the most widely used trait taxonomy for personality assessment, and a primary catalyst for ensuing organizational studies and applications (Digman, 1963; Goldberg, 1990). With concerns about the deleterious effects of faking have been allayed, generalizable findings that correlate well with various organizational outcomes were found across contexts and cultures. However, unintentional response distortions including individual differences in response style and frame of reference effect in personality assessment remain concerning for research and practice uses of personality self-assessment (Riese & Waller, 1993; Stark & Drasgow, 2006; Carter et al., 2015; Tay & Ng, 2018). The current study aims to explore different methods for adjusting self-rated personality (fixed vignettes and self-rated anchoring vignettes) relative to a control condition with no adjustments. The results showed that although fixed vignettes failed to improve the reliability and criterion-related validity than the unadjusted scores, personality adjusted based on self-rated vignettes decreased in correlations to relevant criteria.

CHAPTER 1. INTRODUCTION

Personality traits are probabilistic descriptions of relatively stable patterns of emotion, motivation, cognition, and behavior, in response to classes of stimuli that have been present in human cultures over evolutionary time (DeYoung, 2015). As such broad definitions imply, personality test scores should and do predict a wide array of social, clinical, educational, and workplace outcomes. Although research on the validity of personality measures using different response formats and sources of information (self- vs. other- reports) dates back to the 1940s, the emergence of parsimonious factor models in the 1960s, resembling the modern Big Five (Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability (Digman, 1963; Goldberg, 1990), was a primary catalyst for ensuing organizational studies and applications. In addition, seminal meta-analyses in the 1990s put to rest many of the concerns in Mischel’s (1968) critique of trait perspectives by demonstrating that Big Five scores correlate with overall job performance (Barrick & Mount, 1991), and higher correlations may be observed when concentrating on the match between traits and occupational requirements (Tett, Jackson, & Rothstein, 1991).

Today the Big Five is the dominant and most widely used trait taxonomy for personality assessment. It has been “replicated” in different contexts and cultures using a variety of approaches and serves as a benchmark for explaining and evaluating alternative lower- and higher-order representations of personality (Paunonen & Ashton, 2001; Roberts *et al.*, 2005; DeYoung, Quilty, & Peterson, 2007; Ones *et al.*, 2007; Connelly & Ones, 2010; Stark *et al.*, 2014;). A plethora of more recent primary and secondary (meta-analysis) studies have also shown that Big Five scores correlate with organizational outcomes such as overall, citizenship, and counterproductive work performance, training proficiency, satisfaction, retention and a wide array of non-work outcomes. Moreover, longstanding concerns about the deleterious effects of faking (i.e., deliberate response distortion) in organizational contexts have been allayed, to some extent, by validity evidence for traditional (Likert-type) personality measures in employee development and selection contexts (Hogan, Barrett, & Hogan, 2007; Law, Bourdage, & O’Neill, 2016) as well as modern forced-choice tests (Brown & Maydeu-Olivares, 2011; Stark *et al.*, 2005; Stark *et al.*, 2014; Saville, *et al.*, 1996; Boyce et al., 2016) that appear to predict performance (Salgado & Tauriz, 2014) with greater resistance to faking (Cao & Drasgow, 2019).

Nevertheless, important questions remain concerning the most appropriate psychometric models for personality and, more generally, noncognitive responding (Riese & Waller, 1993; Stark & Drasgow, 2006; Carter et al., 2015; Tay & Ng, 2018) and how to most effectively deal with *un*intentional response distortions, or biases, resulting from differences in response style or frame of reference. As Stark *et al.* (2001) noted in a differential item functioning (DIF) study looking at the effects of factors in the testing environment on item responding, examinees may have different and perhaps changing frames of reference as they respond to items in the course of a personality test. For example, how they respond to a conscientiousness item may depend on whether they momentarily envision an organized, responsible supervisor, a colleague who works hard but occasionally misses a meeting or deadline, or a work team that is not goal-oriented and ignores rules it believes are unnecessary. Such idiosyncrasies in the referents undergirding self- and other- ratings of personality may lead to nonintuitive results in DIF, or alternatively, measurement invariance (MI) studies focusing on changes in item parameters across groups or contexts (Stark *et al.*, 2001). Alternatively, it is also possible to model intentional and perhaps unintentional response biases as a temporary change in a respondent (i.e., person parameters) when presented with an item that is more transparent in what it measures and there is a clearly best answer from an organizational standpoint. Additional, more insidious complications may arise from systematic influences, such as preference for inner or extreme categories of a polytomous response scale, or contextual norms and values that affect item responding while leaving item parameters unchanged. Collectively, these influences can distort item- and test-level properties and/or scale/factor means and variances, leading to interpretations that are inconsistent with behavioral observations and other objective data. Therefore, it may be necessary to adjust for differences in respondents’ frame of reference in studies comparing personality scores and predictive utility within as well as across contexts.

CHAPTER 2. LITERATURE REVIEW

In the following sections, I describe a study designed to explore frame of reference effects and three approaches to adjust for them in personality measurement. First, I will explore an approach by Mõttus *et al*. (2012) that transforms categorical item responses into continuous responses, which are adjusted for individuals’ idiosyncrasies, based on their ratings of *anchoring vignettes* exemplifying high, medium, and low levels of a targeted construct - in this case, Big Five personality traits. Second, I will examine a confirmatory factor analysis (CFA) approach to adjust personality scores using the ratings of the anchor vignettes. The advantage of this approach is that it leaves the raw responses to the personality measure unchanged, thus permitting measurement invariance testing when necessary. Third, I will examine an approach that has been used for frame of reference training to improve the accuracy of performance appraisal in organizations. More specifically, rather than having respondents rate the extremity of the attributes portrayed in the vignettes, the vignettes will be presented with their designated numerical anchors prior to completing a personality inventory to help respondents calibrate their self-report judgements. These three adjustment approaches will be compared to a control condition involving unadjusted personality scores. The next sections of this paper describe frame of reference effects, the proposed adjustment methodologies, and an experiment comparing the effects of these adjustments on the psychometric properties and predictive utility of a Big Five personality measure.

**Frame of Reference Effects**

An abundance of studies have examined the impact of response biases in the areas of personality measurement and performance appraisal. For example, ratings of employee performance may be biased by raters’ response styles and motivations. Common rater errors include leniency (providing overly positive ratings of all), severity (providing overly negative ratings of all), and central tendency (providing middling ratings of all). Collectively, these errors impact means, variances, and sometimes correlations with performance predictors. Similarly, self-reported responses to personality items may be influenced by a number of biases. These include acquiescent (the tendency to agree with items), dis-acquiescent (the tendency to disagree with items), socially desirable (the tendency to answer in a way to make oneself look good), extreme (the tendency to endorse the most extreme response categories), and midpoint (the tendency to endorse the middle response category) responding. Differential use of the response format is particularly salient in cross-cultural comparisons of collectivistic and individualistic countries (Chen, Lee, and Stevenson, 1995; He et al., 2014; Lalwani, Shrum, & Chiu, 2009).

These and other response biases may stem from social comparison, or more generally, frame of reference effects. Festinger (1954) discussed two social comparison processes by which individuals evaluate their performance or capabilities relative to others. In upward social comparison, individuals evaluate themselves relative to more capable or higher-performing referents, whereas in downward social comparison, individuals evaluate themselves relative to less capable or lower performing referents. In both cases, the comparison other provides a frame of reference for self-evaluation. In the context of personality measurement, one’s witting or unwitting choice of referent can impact his or her trait scores, which can in turn affect group means, variances, and correlations (Bauer et al, 2003; Heine et al., 2008; Mõttus et al., 2012; Schmit *et al*., 1995; Yoon, Schmidt, & Ilies, 2002).

**Adjusting for Frame of Reference Effects**

Over the years, several approaches have been proposed to control or adjust for context when completing noncognitive measures. For example, Bauer *et al.* (2003) found that contextualizing personality items by adding the words ‘at work’ had a positive impact on validity. Although this may help to focus self-evaluations of personality characteristics in the context of work, it does not necessarily improve calibration relative to a comparison other or objective standard. In the context of personnel selection, for example, applicants may envision an ideal employee and tailor their responses to create a more favorable impression for their prospective employer. This adversely affects scale means, variances, and correlations with other scales and work criteria, and is widely considered to be a problem in selection environments. To deal with this effect, researchers have tried using regression to remove the effect of socially desirable responding from trait scores by including scales designed to detect impression management as control variables. Although research in the 1990s showed that this had little, if any, salutary effect on validity (Stark *et al*., 2001; Ones & Viswesvaran, 1998), this idea seems more aligned with recent efforts to calibrate self-report judgements by using *ratings* of anchoring vignettes or, alternatively, *fixed* vignettes exemplifying low, medium, and high levels of traits using short, vivid descriptions of hypothetical persons. Specifically, Mõttus and colleagues proposed an approach that uses individuals’ ratings of anchoring vignettes to adjust for their unique response styles and interpretations of questions in the context of political attitudes. This approach has been used in cross-cultural research to adjust for frame of reference effects created by broad cultural norms (Primi *et al.*, 2016). More recent research has shown that adjustments based on self-ratings of anchoring vignettes can increase reliability, discriminant power, criterion validity, and interpersonal comparability of self-rated personality scores (Weiss & Roberts, 2018). Alternatively, presenting fixed vignettes that explicitly designate low, medium, and high trait levels before completing a personality inventory should crystalize the standards for self-evaluation, in the same way that critical incidents reflecting different levels of employee performance serve as referents in behaviorally anchored rating scales for employee performance appraisal. Examples of anchoring vignettes can be found in Appendix D.

**Current study**

I plan to conduct an experiment comparing different methods for adjusting self-rated personality (fixed vignettes and self-rated anchoring vignettes) relative to a control condition with no adjustments. If the adjustments are effective, the adjusted scores should exhibit higher reliability and criterion-related validity than the unadjusted scores. However, no specific predictions can be made for means and variances, as all participants will be sampled from the same population (i.e., within-country).

The transformation of the raw ordered-categorical responses to personality items using individuals’ ratings of anchoring vignettes via a mathematical procedure called Censored orders probit (COP) developed by King and Wand (2007). COP is a generalization of the standard ordered probit model which can use vectored values in addition to scalar values as dependent variables. This approach has been used with some success in recent research (Mõttus *et al.*, 2012), but has the disadvantage of altering the raw responses, thus complicating any downstream analyses involving, for example, measurement invariance testing across subpopulations.

CHAPTER 3. METHOD

**Participants**

Researchers obtained a convenience sample of 1197 college students from both Amazon Mechanical Turk and the online participant pool program from the university of the southeastern USA. The participants must be over 18 years old to volunteer in this study. The researchers also excluded participants who have failed to select the given option in three attention check questions or those who have reported we should not include their data because of careless response.

**Measures**

**Demographics.** Participants will be asked to report their age, gender, ethnicity, class standing, native language, and parental/household income.

**Self-reported GPA.** Participants will be asked to report their current overall GPA in 4.0 scale.

**Health Behavior Checklist (HBCL).** HBCL is a 40-item measure comprising four subscales: wellness maintenance and enhancement (10 items), accident control behaviors (6 items), traffic risk (7 items), and substance use risk (3 items), with reliabilities ranging from .50 to .77 (Vickers, Conway, & Hervig, 1990). The items are rated using a 5-point Likert-type format from 1= “very uncharacteristic of me” to 5= “very characteristic of me” (see Appendix A). In the current study, the Cronbach’s alpha of HBCL is .77.

**Utrecht Work Engagement Scale for Students (UWES-S).** School engagement will be assessed with the 17-item UWES-S, which includes three subscales: vigor (6 items), dedication (5 items) and absorption (6 items), with reliabilities ranging from .68 to .91 (Schaufeli et al., 2002). The items are rated using a 5-point Likert-type format from 1= “strongly disagree” to 5= “strongly agree” (see Appendix B). In the current study, the Cronbach’s alpha of UWES-S is .93.

**Big Five Inventory-2 (BFI-2).** BFI-2 is a 60-item personality inventory measuring 15 facets that can be arranged to form five broad personality factors reflecting the Big Five: extraversion, agreeableness, conscientiousness, negative emotionality, and open-mindedness. Each broad personality factor is defined by three facets measured by four items each. Extraversion includes sociability, assertiveness, and energy level; agreeableness includes compassion, respectfulness, and trust; conscientiousness includes organization, productiveness, and responsibility; negative emotionality includes anxiety, depression, and emotional volatility; and open-mindedness includes intellectual curiosity, aesthetic sensitivity, and creative imagination (Soto & John, 2017). The internal consistency reliabilities for the 4-item facets scales ranged from .65 to .84 (Soto & John, 2017). The items are rated using a 5-point Likert-type format from 1= “strongly disagree” to 5= “strongly agree” (see Appendix C).

**Anchoring Vignettes.** An anchoring vignette was constructed to represent the high, medium, and low level of each facet of each broad personality factor. These exemplars were reviewed by three subject matter experts for correspondence with the facet definitions. For self-rated vignettes, the participants will rate the vignettes on a 5-point Likert-type scale by judging the level of traits expressed by the hypothetical people; For fixed vignettes, the low, average, and high-level behavioral descriptions will be assigned to 1, 3, and 5 on the 5-point Likert-type scale (see Appendix D).

**Procedure**

All participant data will be collected using an online questionnaire. Participants who sign up for this online study will receive a link to an external Qualtrics survey. First, a consent form will be presented. Then, participants will be randomized into one of three conditions: fixed vignettes (N = 300), self-rated vignettes (N = 300), or control (N = 150). In the fixed condition, participants will be instructed to review the vignettes describing the low, medium, and high levels of a personality trait, followed by the BFI-2 items measuring that trait. In the self-rated condition, respondents will be instructed either to 1) review and rate the vignettes describing each personality trait followed by the BFI-2 items measuring that trait, or 2) answer the BFI-2 items for each trait and then rate the vignettes. In the control condition, the respondent will complete all of the BFI-2 items, followed by the self-rated vignettes. In all conditions, the respondents will then complete the criterion measures and demographics form. One attention check item will be included in the set of items measuring each personality trait to screen out random or inattentive responders. Respondents who miss three or more attention checks will be excluded from the analyses.

This design serves several purposes. First, in the self-rated vignettes condition, presenting the self-rated vignettes before the trait items enables an equivalent comparison with the fixed vignette condition, and thus, the most direct test of the fixed versus statistical adjustment methods. Second, counterbalancing the presentation of the vignettes and BFI-2 trait items in the self-rated vignettes condition provides an opportunity to examine order/priming effects, if any. If there are none, then those data can be collapsed across orderings for comparison with the fixed vignette and control conditions. Third, by including a control condition, we can examine how the methods involving anchoring vignettes alter the psychometric properties of the personality measure and whether they result in any improvements in the criterion-related validities. A flowchart of the experimental design is presented in Figure 1.

**Hypotheses**

Hypothesis 1a: The presentation of fixed vignettes will not alter the psychometric properties (means, variances, reliabilities) of the BFI-2 trait scores.

Hypothesis 1b: The presentation of fixed vignettes improves the correlations of personality scores with the criterion measures relative the control condition.

Hypothesis 2: The use of self-rated vignettes is more effective than fixed vignettes in improving the correlations of personality scores with the criterion measures relative the control condition.

CHAPTER 4. RESULTS

**Preliminary Analyses**

**Missing data.** The researchers excluded participants who have failed to select the given option in either of three attention check questions or those who have reported we should not include their data because of careless response. The sample size of current study after cleaning decreased from 1197 to 906, which includes 552 in control group, 354 in fixed vignettes group, and 292 in self-rated vignettes group. All 292 participants from the self-rated vignettes group were also included into the control group since participants were asked to rate the vignettes items after self-assessment in personality traits as they do in the control group. The sample size for each personality traits may vary a little due to unexpected missing data in certain trait only. The author decided to keep these observations for maximum usage of the available information.

**Descriptive Statistics and Correlations.** The sample after cleaning includes 46.29% females, 52.91% males, and .80% individuals identified as non-binary, genderqueer, or transgender. 47.51% are White or Caucasian, 9.30% are Black or African American, 9.64% are Hispanic or Latino, 1.13% are Native American or Alaskan Native, 1.13% are Pacific Islander, 23.81% are Asian or Asian American, and 7.48% identified as others or multiracial. The mean age of this sample is 28.11, and the standard deviation is 9.69.

Among all the individuals, 85.60% reported that English is or is one of their native languages. 12.60% reported their household annual income is lower than $20,000, 16.57% is in between $20,000 and $34,999, 19.41% is in between $35,000 and $49,999, 24.86% is in between $50,000 and $74,999, 16.57% is in between $75,000 and $99,999, 9.99% is over $100,000. The means, standard deviations and correlations among variables of interest for all three method groups were calculated (Table 1: Table 3). As seen in Table 1,

Censored Ordered Probit (COP; King & Wand, 2007) was applied to transform the scalar values of personality scores using vectored value (C) to compare rank order among individuals. In the current study, package ‘anchors’ from R software was used for relevant analysis (Wand, King, & Lau, 2009). See Figure 2a -2e for proportion distribution of personality scores after adjustment. The sample size of adjusted personality scores is 165 after excluding non-scalar observations. The distribution of scores for conscientiousness, agreeableness, extraversion, and openness are majorly normal. The bars on score 1, 3, 5, and 7 are higher than those on score 2, 4, and 6 due to the extension of 5-point Likert scale to 7-point Likert scale for the adjustment purpose. Lastly, the allocation for negative emotionality is positively skewed, which matched the voluntary student sample.

**Main Analyses**

*Hypothesis 1a: The presentation of fixed vignettes will not alter the psychometric properties (means, variances, reliabilities) of the BFI-2 trait scores.*

The 552 participants who directly reported their self-assessment on personality items (*M* = 3.63, *SD* = .73) compared to the 350 participants who were given the fixed vignettes before answering the same personality items (*M* = 3.64, *SD* = .78) demonstrated non-significant difference in conscientiousness score, *t*(710.74) = -.15, *p* = .88; *F*(552, 350) = .88, *p* = .19. This pattern of results haven been repeatedly found in other personality traits as well. There is no significant difference in agreeableness, *t*(760.01) = .95, *p* = .34; *F*(546, 351) = 1.04, *p* = .67 between control (*M* = 3.59, *SD* = .67) and fixed vignettes group (*M* = 3.54, *SD* = .65); extraversion, *t*(752.12) = .93, *p* = .35; *F*(546, 352) = 1.00, *p* = .98 between control

(*M* = 3.19, *SD* = .69) and fixed vignettes group (*M* = 3.15, *SD* = .69); openness, *t*(764.66) = .19, *p* = .85; *F*(550, 354) = 1.03, *p* = .73 between control (*M* = 3.62, *SD* = .76) and fixed vignettes group (*M* = 3.61, *SD* = .74); nor negative emotionality, *t*(714.40) = -.67, *p* = .51;

*F*(550, 354) = .86, *p* = .12 between control (*M* = 2.73, *SD* = .77) and fixed vignettes group

(*M* = 2.77, *SD* = .83). See Table 4 for t-test results comparing control and fixed vignette groups on means of each personality scores, and Table 5 for F-test results comparing control and fixed vignette groups on standard deviations of each personality scores.

Cronbach’s alpha of conscientiousness from control group is .86, from fixed vignettes group is .88; that of agreeableness from control group is .82, from fixed vignettes group is .79; that of negative emotionality from control group is .86, from fixed vignettes group is .88; that of extraversion from control group is .81, from fixed vignettes group is .80. Lastly, the Cronbach’s alpha of openness from control group is .86, from fixed vignettes group is .85.

The independent sample t-test and F-test failed to show any significant result in means nor standard deviations of each personality scores. Besides, the reliabilities of both groups failed to show that one is more efficient than the other. Thus, this hypothesis was supported. Presenting fixed vignettes prior to self-assessment did not alter the psychometrics of personality scores including means, standard deviations, and reliabilities.

*Hypothesis 1b: The presentation of fixed vignettes improves the correlations of personality scores with the criterion measures relative the control condition.*

In the control condition, conscientiousness (*r* = .16, *p* < .01) and negative emotionality

(*r* = -.01, *p* < .05) were significantly correlate with college student’s academic achievement (GPA). Extraversion (*r* = .20, *p* < .01), agreeableness (*r* = .38, *p* < .01), conscientiousness

(*r* = .46, *p* < .01), negative emotionality (*r* = -.38, *p* < .01), and openness (*r* = .22, *p* < .01) were significantly correlate with health behaviors. Besides, extraversion (*r* = .30, *p* < .01), agreeableness (*r* = .12, *p* < .01), conscientiousness (*r* = .26, *p* < .01), and negative emotionality

(*r* = -.33, *p* < .01) were significantly correlate with study engagement. In the fixed vignettes condition, similar patterns were found. Conscientiousness (*r* = .17, *p* < .01) was significantly correlate with GPA. Extraversion (*r* = .36, *p* < .01), agreeableness (*r* = .29, *p* < .01), conscientiousness (*r* = .41, *p* < .01), negative emotionality (*r* = -.37, *p* < .01), and openness

(*r* = .17, *p* < .01) were significantly correlate with health behaviors. Lastly, extraversion

(*r* = .32, *p* < .01), conscientiousness (*r* = .22, *p* < .01), and negative emotionality

(*r* = -.34, *p* < .01) were significantly correlate with study engagement. See Table 1-2 for correlations and their confidence intervals between personality scores and three criterions (GPA, health behavior, and study engagement).

Although the relationships between negative emotionality (*r* = -.08, *ns.*) and GPA, as well as agreeableness (*r* = .10, *ns.*) and study engagement were non-significant in the fixed vignettes group, by comparing the confidence interval of them relative the control condition, the difference is not significant to be acknowledged. Therefore, this hypothesis was not supported. Fixed vignettes failed to improve the correlations of personality scores with the criterion measures relative the personality assessment alone.

*Hypothesis 2: The use of self-rated vignettes is more effective than fixed vignettes in improving the correlations of personality scores with the criterion measures relative the control condition.*

In the self-rated vignettes group, extraversion (*r* = .15, *p* < .05) and conscientiousness

(*r* = .18, *p* < .05) were significantly correlate to GPA. Agreeableness (*r* = .22, *p* < .01), conscientiousness (*r* = .29, *p* < .01), and negative emotionality (*r* = -.19, *p* < .05) were significantly correlate to health behavior. Only negative emotionality (*r* = -.29, *p* < .01) was significantly correlate to study engagement. Comparing to the correlations between personality scores and criterions in the control condition, the correlation between extraversion after adjustment and GPA became significant, while the relationship between adjusted extraversion and both health behavior and study engagement become insignificant. The correlation between adjusted agreeableness and study engagement, conscientiousness and study engagement, negative emotionality and GPA, as well as openness and health behavior all became insignificant. Besides, for the correlations that are significant in both conditions, except the correlation between conscientiousness and GPA increased from .16 to .18, all correlations in the self-rated vignettes group showed to be smaller than the correlations in the control group. Considering the confidence interval of correlation between conscientiousness and GPA was [.08, .25] in the control condition, and [.03, .33] in the self-rated vignettes condition, the difference was not significant to be acknowledged.

Thus, the last hypothesis was not supported by the results. Although the usage of self-rated vignettes changed the correlations between personality scores and relevant criterion (GPA, health behavior, and study engagement), the adjustment lowered the majority of correlations and failed to improve the effectiveness of using self-assessed personalities to predict important outcomes in academic and health behavior.

CHAPTER 5: DISCUSSION

In summary, the present study offers important and novel findings regarding the effectiveness of adjusting methods in personality assessment. The first method is to present fixed vignettes that describe the cognition, affect, and behaviors examples of different levels of personality traits to the respondents. Then, these respondents are asked to assess their own personality by answering the regular personality inventory. The respondents are expected to compare themselves to the examples to ensure a standardized understanding of each personality trait and the levels of scale. The second method is to ask respondents to answer both the regular personality inventory and to rate hypothetical person’s personality based on their own understanding and standard. The response distortions are adjusted using Censored Ordered Probit method (King & Wand, 2007). The expectations are also to standardize the interpretation of scale levels, the understanding of questions by providing vivid examples, and the comparison groups (Weiss & Roberts, 2018).

To compare the effectiveness of using fixed vignettes and control condition, hypothesis 1 was partially supported. First, the means, standard deviations, and reliabilities of each personality traits were not significantly different from the control condition, which suggested using fixed vignettes does not alter the psychometrics of personality scores. Second, the correlations between personality traits and criteria (GPA, health behavior, and school engagement) were also not significantly different from the control condition, which suggested using fixed vignettes failed to improve the effectiveness of personality assessment in predicting important criteria.

To compare the effectiveness of using self-rated vignettes and control condition, hypothesis 2 was rejected. Since the COP adjustment extended the 5-point scale to 7-point scale, and the model transformed the scalar values of each personality trait using the vectored value, in the current study, the author did not directly compare means, standard deviations, and reliabilities to control condition. As shown in Table 1 and Table 3, the correlations between personality traits and criteria decreased after adjustment which contradict to the hypothesis. It has been shown in the literature that discriminant validity and criterion validity increased only when each scale was corrected using its own corresponding vignette set. Moreover, accuracy in rating the vignettes was correlated with language achievement test scores, suggesting that verbal ability has an impact on providing both normative vignette ratings of others and self-reports that are reliable and valid (Primi, Zanon, Santos, De Fruyt, & John, 2016; Weiss & Roberts, 2018). In the current study, the scales were corrected based on the vignette sets of their own factor and the verbal ability should not be a crucial issue for current college students who volunteered in this study.

Another possible explanation is according to the country-level correlational analysis conducted by He and colleagues (2014), when personality traits were assessed with Big Five Inventory, the authors found significant correlations between extreme response style and agreeableness (*r* = .50), conscientiousness (*r* = .42), and openness (*r* = .56); when the same personality traits were assessed with self-reported Revised NEO personality inventory, agreeableness score was also associated with midpoint response style (*r* = .51), conscientiousness score was associated with acquiescent response style (*r* = .53), extreme response style (*r* = .69), midpoint response style (*r* = .50). The current study supported that participants’ responses to personality assessment are influenced by content-irrelevant factors such as the rating scales which overemphasized the relationships between personalities and criteria.

Furthermore, the current study also supported the relationships between Big Five personality traits and college student’s academic behavior, outcome and health behavior. The correlations with or without adjustment still supported that personality constructs are useful for explaining and predicting behavior and performance in educational settings. For example, agreeableness and conscientiousness predicts college students’ health behavior, conscientiousness correlated with GPA, negative emotionality showed negative relationship with health behavior and school engagement.

**Implications for Research and Practice**

These findings have several implications for our understanding of personality assessment and comparison. The proportion of variance explained by unintentional response distortions in personality assessment influenced the consistency with behavioral observations and other objective data, as well as correlations with performance criterion. Providing the fixed vignettes is not an effective method to adjust for response distortions, although the fixed vignettes are presented aside to the personality inventory with a scale to visualize the low, median, and high level of each personality trait, as well as the instruction on the usage of fixed vignettes. The current study raised concern on using fixed anchoring vignettes in practical scenarios including performance assessment and evaluation. Besides, the correlations between personality traits and criteria decreased after using COP to adjust scores based on self-rated vignette sets. This finding supported that unintentional response distortions has influence on predicting performance. There is necessity to correct for the distortions in correlational analysis even under the country level.

**Limitations and Future Directions**

Although the present author asserts the importance of the findings, there were also several limitations to the study which must be acknowledged. Demographically, the sample is limited to convenient student sample who lacks motivation to carefully follow the instruction. The current study tried to minimize the careless responding through excluding any participants who failed attention check question or self-reported as careless response at the end of the survey. However, it was still difficult to ensure that participants compare themselves to the fixed vignettes while answering every personality items. Future research would do well to replicate and extend the results with motivated work population. Besides, the sample size decreased significantly after cleaning and adjustment. For any study that requires more complexed statistical analysis, a larger sample is preferred. Additionally, the current study did not compare the psychometrics between self-rated vignettes condition and control condition, because further transformation on adjusted scores may alter the means and standard deviations of personality scores and lead to inaccurate comparison. However, the author recognizes this limitation and believes the future study should improve on the procedure and analysis.

Moreover, the current study is limited to local population, while research surrounding personality differences across nations and cultures are becoming an urgent issue. In the current cross-national studies, mean levels of self-reported phenomena are often not congruent with more objective criteria. Cultural factors such as reference-group effect, interpretation difference, response style may undermine the validity of average country scores of personality traits. Besides, the argument around whether between-country differences in personality scores represent true score variance or systematic instrument-related biases remains unsolved. Some studies examined that people from different geographic regions were significantly different in certain personality traits (Allik & McCrae, 2004; Schmitt, Allik, McCrae, & Benet-Martínez, 2007; Bartram, 2013), while others provided support for cross-national equivalence of personality (McCrae *et al.*, 2005; Heine, Buchtel, & Norenzayan, 2008). Future research should continue exploring a better method to adjust unintentional response distortions and apply on diverse population to uncover those questions.

Finally, the current study did not fully explain that adjusted personality scores using self-rated vignettes lead to decreased correlations with relevant criteria. Whether participants tend to use extreme scales which overestimated the relationship between personality and performance? Whether the same tendency can be extended to other criteria or other population? Future research can explore further on these questions.

**Conclusions**

Considering the above findings, implications, and limitations of the present study, the author concludes that this study fills a gap in the current literature. By exploring both fixed vignettes and self-rated vignettes method in adjusting for unintentional response distortion in personality assessment, the current finding supported the influence of distortion on correlations between personality and performance criteria. Although fixed vignettes did not alter both psychometrics and correlations relative control condition, self-rated vignettes decreased majority correlations relative control condition. Future research should continue exploring the pattern of distortion as well as extending the findings to diverse population.

References

Allik, J., & McCrae, R. R. (2004). Toward a Geography of Personality Traits: Patterns of Profiles across 36 Cultures. *Journal of Cross-Cultural Psychology*, *35*(1), 13–28. https://doi.org/10.1177/0022022103260382

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10/hjv

Bartram, D. (2008). Global Norms: Towards Some Guidelines for Aggregating Personality Norms Across Countries. *International Journal of Testing*, *8*(4), 315–333. https://doi.org/10.1080/15305050802435037

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363.

Boyce, A. S., Conway, J. S., & Caputo, P. M. (2016). *Development and validation of Aon Hewitt’s personality model and Adaptive Employee Personality Test (ADEPT-15)*. Unpublished technical report.

Brown, A., & Maydeu-Olivares, A. (2011). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, *71*(3), 460–502. <https://doi.org/10.1177/0013164410375112>

Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology*, *104*(11), 1347–1368. <https://doi.org/10.1037/apl0000414>

Carter, E. C., Kofler, L. M., Forster, D. E., & Mccullough, M. E. (2015). A Series of Meta- Analytic Tests of the Depletion Effect : Self-Control Does Not Seem to Rely on a Limited Resource. *Journal of Experimental Psychology: General*, *144*, 796–815.

Chen, C., Lee, S., & Stevenson, H. W. (1995). Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students. *Psychological Science*, *6*(3), 170–175. https://doi.org/10.1111/j.1467-9280.1995.tb00327.x

Connelly, B. S., & Ones, D. S. (2010). An Other Perspective on Personality: Meta-Analytic Integration of Observers’ Accuracy and Predictive Validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10/dqbns8

DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, *56*, 33–58. <https://doi.org/10/33h>

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, *93*(5), 880–896. https://doi.org/10.1037/0022-3514.93.5.880

Digman, J. M. (1963). PRINCIPAL DIMENSIONS OF CHILD PERSONALITY AS INFERRED FROM TEACHERS’ JUDGMENTS. 19.

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Goldberg, L. R. (1990). An Alternative “Description of Personality”: The Big-Five Factor Structure. *Journal of Personality and Social Psychology*. 14.

He, J., Bartram, D., Inceoglu, I., & van de Vijver, F. J. R. (2014). Response Styles and Personality Traits: A Multilevel Analysis. *Journal of Cross-Cultural Psychology*, *45*(7), 1028–1045. https://doi.org/10.1177/0022022114534773

Heine, S. J., Buchtel, E. E., & Norenzayan, A. (2008). What Do Cross-National Comparisons of Personality Traits Tell Us?: The Case of Conscientiousness. *Psychological Science*, *19*(4), 309–313. https://doi.org/10/b429cv

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, *92*(5), 1270–1285. <https://doi.org/10/bvvh38>

King, G., & Wand, J. (2007). Comparing Incomparable Survey Responses: Evaluating and Selecting Anchoring Vignettes. *Political Analysis*, *15*(01), 46–66. https://doi.org/10.1093/pan/mpl011

Lalwani, A. K., Shrum, L. J., & Chiu, C. (2009). Motivated response styles: The role of cultural values, regulatory focus, and self-consciousness in socially desirable responding. *Journal of Personality and Social Psychology*, *96*(4), 870–882. https://doi.org/10.1037/a0014622

Law, S. J., Bourdage, J., & O’Neill, T, A. (2016). To fake or not to fake: Antecedents to interview faking, warning instructions, and its impact on applicant reactions. *Front. Psycol.*, *15*, <https://doi.org/10.3389/fpsyg.2016.01771>

Mischel, W. Personality and assessment. New York Wiley, 1968.

McCrae, R. R., Terracciano, A., & Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*(3), 407–425. https://doi.org/10.1037/0022-3514.89.3.407

Mõttus, R., Allik, J., Realo, A., Pullmann, H., Rossier, J., Zecca, G., … Ng Tseung, C. (2012). Comparability of Self-Reported Conscientiousness Across 21 Countries: Comparability of self-reports across cultures. *European Journal of Personality*, *26*(3), 303–317. https://doi.org/10.1002/per.840

Ones, D. S., & Viswesvaran, C. (1998). The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection. *Human Performance*, *11*(2–3), 245–269. <https://doi.org/10.1080/08959285.1998.9668033>

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, *60*(4), 995–1027. https://doi.org/10/ft4bhz

Paunonen, S. V., & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*(3), 524–539. [https://doi.org/10.1037//0022-3514.81.3.524](https://doi.org/10.1037/0022-3514.81.3.524)

Primi, R., Zanon, C., Santos, D., De Fruyt, F., & John, O. P. (2016). Anchoring Vignettes: Can They Make Adolescent Self-Reports of Social-Emotional Skills More Reliable, Discriminant, and Criterion-Valid? *European Journal of Psychological Assessment*, *32*(1), 39–51. <https://doi.org/10.1027/1015-5759/a000336>

Reise, S. P., & Waller, N. G. (1993). *Traitedness and the Assessment of Response Pattern Scalability.* 9.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). THE STRUCTURE OF CONSCIENTIOUSNESS: AN EMPIRICAL INVESTIGATION BASED ON SEVEN MAJOR PERSONALITY QUESTIONNAIRES. *Personnel Psychology*, *58*(1), 103–139. <https://doi.org/10.1111/j.1744-6570.2005.00301.x>

Salgado, J. F., & Táuriz, G. (2014). The Five-Factor Model, forced-choice personality inventories and performance: A comprehensive meta-analysis of academic and occupational validity studies. *European Journal of Work and Organizational Psychology*, *23*(1), 3–30. <https://doi.org/10.1080/1359432X.2012.716198>

Saville, P., Sik, G., Nyfield, G., Hackston, J., & MacIver, R. (1996). A Demonstration of the Validity of the Occupational Personality Questionnaire (OPQ) in the Measurement of Job Competencies Across Time and in Separate Organisations. *Applied Psychology: An International Review*, *45*(3), 243-262.

Schmit, M. J., & Ryan, A. M. (1995). *Frame-of-Reference Effects on Personality Scale Scores and Schmitt, D. P., Allik, J., McCrae, R. R., & Benet-Martínez, V. (2007). The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. Journal of Cross-Cultural Psychology, 38(2), 173–212. https://doi.org/10.1177/0022022106297299Criterion-Related Validity*. 14.

Stark, S., Chernyshenko, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology*, *86*(5), 943–953. [https://doi.org/10.1037//0021-9010.86.5.943](https://doi.org/10.1037/0021-9010.86.5.943)

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT Approach to Constructing and Scoring Pairwise Preference Items Involving Stimuli on Different Dimensions: The Multi-Unidimensional Pairwise-Preference Model. *Applied Psychological Measurement*, *29*(3), 184–203. <https://doi.org/10.1177/0146621604273988>

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, *91*(6), 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A New Generation of Personality Tests to Support Military Selection and Classification Decisions. *Military Psychology*, *26*(3), 153–164. <https://doi.org/10.1037/mil0000044>

Tay, L., & Ng, V. (2018). Ideal Point Modeling of Non-cognitive Constructs: Review and Recommendations for Research. Frontiers in Psychology, 9. <https://doi.org/10.3389/fpsyg.2018.02423>

Tett, R. P., Jackson, D. N., & Rothstein, M. (2006). PERSONALITY MEASURES AS PREDICTORS OF JOB PERFORMANCE: A META-ANALYTIC REVIEW. *Personnel Psychology*, *44*(4), 703–742. <https://doi.org/10.1111/j.1744-6570.1991.tb00696.x>

Wand J, King G, Lau O (2009). **anchors**: *Software for Anchoring Vignette Data*. R package version 3.0-4, URL http://wand.stanford.edu/anchors/.

Weiss, S., & Roberts, R. D. (2018). Using Anchoring Vignettes to Adjust Self-Reported Personality: A Comparison Between Countries. *Frontiers in Psychology*, *9*. <https://doi.org/10.3389/fpsyg.2018.00325>

Yoon, K., Schmidt, F., & Ilies, R. (2002). Cross-Cultural Construct Validity of the Five-Factor Model of Personality among Korean Employees. *Journal of Cross-Cultural Psychology*, *33*(3), 217–235. https://doi.org/10.1177/0022022102033003001







Table 4

*t-test Results Comparing Control and Fixed Vignette Groups on Means of Personality Measures.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Control | | Fixed Vignettes | | *t-test* | *df* | *p* |
| M | SD | M | SD |
|  |  |  |  |  |  |  |  |
| Conscientiousness | 3.63 | 0.73 | 3.64 | 0.78 | -0.15 | 710.74 | 0.88 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Agreeableness | 3.59 | 0.67 | 3.54 | 0.65 | 0.95 | 760.01 | 0.34 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Extraversion | 3.19 | 0.69 | 3.15 | 0.69 | 0.93 | 752.12 | 0.35 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Openness | 3.62 | 0.76 | 3.61 | 0.74 | 0.19 | 764.66 | 0.85 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| Negative Emotionality | 2.73 | 0.77 | 2.77 | 0.83 | -0.67 | 714.40 | 0.51 |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

*Note.* \* indicates *p* < .05. \*\* indicates *p* < .01.

Table 5

*F-test Results Comparing Control and Fixed Vignette Groups on Variances of Personality Measures.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Control | | Fixed Vignettes | | *F-test* | *num df* | *p* |
| M | SD | M | SD | *denom df* |
|  |  |  |  |  |  |  |  |
| Conscientiousness | 3.63 | 0.73 | 3.64 | 0.78 | 0.88 | 552 | 0.19 |
|  |  |  |  |  |  | 350 |  |
|  |  |  |  |  |  |  |  |
| Agreeableness | 3.59 | 0.67 | 3.54 | 0.65 | 1.04 | 546 | 0.67 |
|  |  |  |  |  |  | 351 |  |
|  |  |  |  |  |  |  |  |
| Extraversion | 3.19 | 0.69 | 3.15 | 0.69 | 1.00 | 546 | 0.98 |
|  |  |  |  |  |  | 352 |  |
|  |  |  |  |  |  |  |  |
| Openness | 3.62 | 0.76 | 3.61 | 0.74 | 1.03 | 550 | 0.73 |
|  |  |  |  |  |  | 354 |  |
|  |  |  |  |  |  |  |  |
| Negative Emotionality | 2.73 | 0.77 | 2.77 | 0.83 | 0.86 | 550  354 | 0.12 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

*Note.* \* indicates *p* < .05. \*\* indicates *p* < .01.

Figure 1: The Procedure Flow Chart

750 US Student sample

method one group

N=150

method two group

N=300

method three group

N=300

demographics

self-rated vignettes

BFI-2

GPA, HBCL, & UWES-S

demographics

fixed vignettes

BFI-2

GPA, HBCL, & UWES-S

demographics

self-rated vignettes

BFI-2

GPA, HBCL, & UWES-S

Figure 2a. Censored Ordered Probit Allocation Graphs

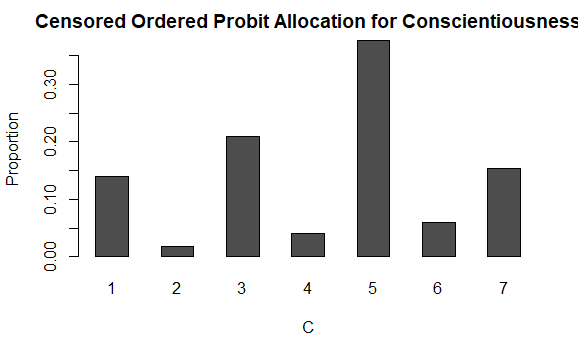


Figure 2b. Censored Ordered Probit Allocation Graphs

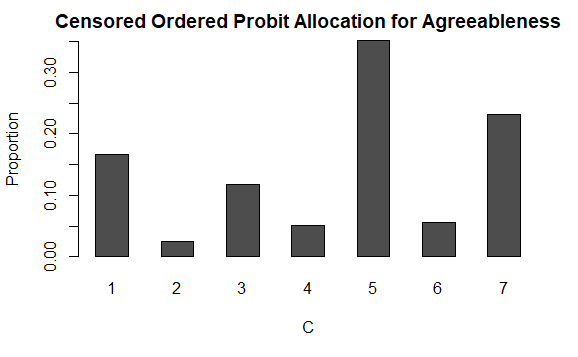


Figure 2c. Censored Ordered Probit Allocation Graphs

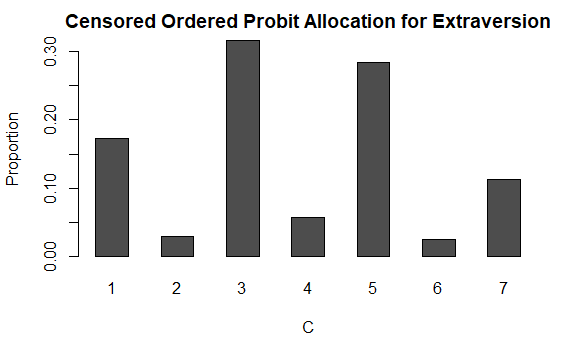


Figure 2d. Censored Ordered Probit Allocation Graphs

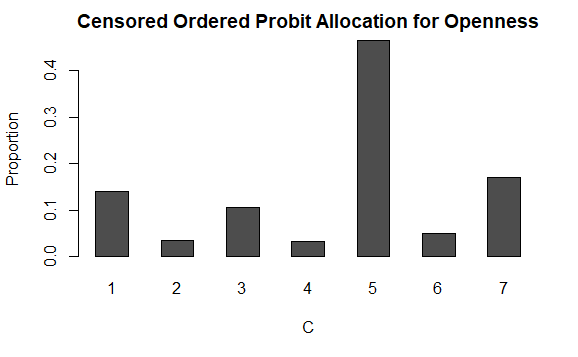
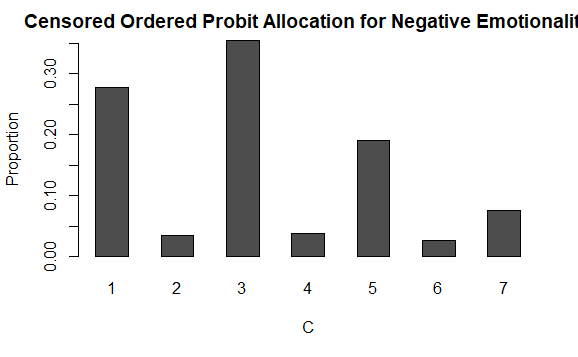


Figure 2e. Censored Ordered Probit Allocation Graphs



**Appendix A: Health Behavior Checklist**

**Wellness Maintenance and Enhancement**

1. I exercise to stay healthy.

2. I gather information on things that affect my health by watching television and reading books, newspapers, or magazine articles.

3. I see a doctor for regular checkups.

4. I see a dentist for regular checkups.

5. I discuss health with friends, neighbors, and relatives.

6. I limit my intake of foods like coffee, sugar, fats, etc.

7. I use dental floss regularly.

8. I watch my weight.

9. I take vitamins.

10. I take health food supplements (e.g. protein additives, wheat germ, bran, lecithin).

**Accident Control**

1. I keep emergency numbers near the phone.

2. I destroy old or unused medicines.

3. I have a first aid kit in my home.

4. I check the condition of electrical appliances, the car, etc. to avoid accidents.

5. I fix broken things around my home right away.

6. I learn first aid techniques.

**Traffic Risk**

1. I cross busy streets in the middle of the block.

2. I take more chances doing things than the average person.

3. I speed while driving.

4. I take chances when crossing the street.

5. I carefully obey traffic rules, so I won't have accidents. [reverse scored]

6. I cross the street against the stop light.

7. I engage in activities or hobbies where accidents are possible (e.g. motorcycle riding, skiing, using power tools, sky or skin diving, hang-gliding, etc.).

**Substance Use Risk**

1. I do not drink alcohol. [reverse scored]

2. I don't take chemical substances which might injure my health (e.g. food additives, drugs, stimulants). [reverse scored]

3. I don't smoke. [reverse scored]

**Appendix B: Utrecht Work Engagement Scale for Students**

**Vigor (VI)**

1. When I get up in the morning, I feel like going to class.

2. When I’m doing my work as a student, I feel bursting with energy.

3. As far as my studies are concerned, I always persevere, even when things do not go well.

4. I can continue studying for very long periods at a time.

5. I am very resilient, mentally, as far as my studies are concerned.

6. I feel strong and vigorous when I’m studying or going to class.

**Dedication (DE)**

1. To me, my studies are challenging.

2. My study inspires me.

3. I am enthusiastic about my studies.

4. I am proud of my studies.

5. I find my studies full of meaning and purpose.

**Absorption (AB)**

1. When I am studying, I forget everything else around me.

2. Time flies when I am studying.

3. I get carried away when I am studying.

4. It is difficult to detach myself from my studies.

5. I am immersed in my studies.

6. I feel happy when I am studying intensely.

**Appendix C: Big Five Inventroy-2**

**Extraversion**

**Sociability items**

1. Tends to be quiet. [reverse scored]
2. Is talkative.
3. Is outgoing, sociable.
4. Is sometimes shy, introverted. [reverse scored]

**Assertiveness items**

1. Is dominant, acts as a leader.
2. Has an assertive personality.
3. Prefers to have others take charge. [reverse scored]
4. Finds it hard to influence people. [reverse scored]

**Energy Level items**

1. Is full of energy.
2. Shows a lot of enthusiasm.
3. Rarely feels excited or eager. [reverse scored]
4. Is less active than other people. [reverse scored]

**Agreeableness**

**Compassion items**

1. Is compassionate, has a soft heart.
2. Can be cold and uncaring. [reverse scored]
3. Is helpful and unselfish with others.
4. Feels little sympathy for others. [reverse scored]

**Respectfulness items**

1. Is respectful, treats others with respect.
2. Is polite, courteous to others.
3. Is sometimes rude to others. [reverse scored]
4. Starts arguments with others. [reverse scored]

**Trust items**

1. Assumes the best about people.
2. Has a forgiving nature.
3. Tends to find fault with others. [reverse scored]
4. Is suspicious of others’ intentions. [reverse scored]

**Conscientiousness**

**Organization items**

1. Tends to be disorganized. [reverse scored]
2. Is systematic, likes to keep things in order.
3. Keeps things neat and tidy.
4. Leaves a mess, doesn’t clean up. [reverse scored]

**Productiveness items**

1. Is efficient, gets things done.
2. Is persistent, works until the task is finished.
3. Tends to be lazy. [reverse scored]
4. Has difficulty getting started on tasks. [reverse scored]

**Responsibility items**

1. Can be somewhat careless. [reverse scored]
2. Sometimes behaves irresponsibly. [reverse scored]
3. Is reliable, can always be counted on.
4. Is dependable, steady.

**Negative Emotionality**

**Anxiety items**

1. Is relaxed, handles stress well. [reverse scored]
2. Worries a lot.
3. Rarely feels anxious or afraid. [reverse scored]
4. Can be tense.

**Depression items**

1. Often feels sad.
2. Tends to feel depressed, blue.
3. Feels secure, comfortable with self. [reverse scored]
4. Stays optimistic after experiencing a setback. [reverse scored]

**Emotional Volatility items**

1. Is emotionally stable, not easily upset. [reverse scored]
2. Is temperamental, gets emotional easily.
3. Keeps emotions under control. [reverse scored]
4. Is moody, has up and down mood swings.

**Open-Mindedness**

**Intellectual Curiosity items**

1. Has little interest in abstract ideas. [reverse scored]
2. Is complex, a deep thinker.
3. Avoids intellectual, philosophical discussions. [reverse scored]
4. Is curious about many different things.

**Aesthetic Sensitivity items**

1. Is fascinated by art, music, or literature.
2. Has few artistic interests. [reverse scored]
3. Values art and beauty.
4. Thinks poetry and plays are boring. [reverse scored]

**Creative Imagination items.**

1. Has little creativity. [reverse scored]
2. Is inventive, finds clever ways to do things.
3. Is original, comes up with new ideas.
4. Has difficulty imagining things. [reverse scored]

**Appendix D: Anchoring Vignettes**

**Conscientiousness:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Facet** | **Low (1)** | **Average (3)** | **High (5)** |
| **Vignette set 1** | *Organization* | Addison often feels it is extremely hard to map out activities or follow the plan. | Hollis likes to keep anything in need at hands but sometimes would forget to clean things up once the job is done. | Parker has a strict daily schedule and can follow it very well. |
| *Productiveness* | Addison never starts on tasks before the last minute and often fails to finish them. | Hollis tends to take some time but generally manages to get things done. | Parker is able to start on a new task anytime if needed while keeping others on track. |
| *Responsibility* | Addison’s coworkers have to cover Addison’s duty from time to time because Addison could not get things done on time or would mess things up. | Hollis’s partner feels at ease to leave some housework but needs to remind Hollis to do them from time to time. | Friends said Parker is the most dependable person they have ever seen. |
| **Vignette set 2** | *Organization* | Adlai never cleans their workspace and leaves all paperwork mixed on the desk. | Riley cleans their studio regularly but sometimes cannot finds their things in the original place. | Cleo is a clean freak. |
| *Productiveness* | Adlai can be distracted from duty quite easily and it takes a long time for Adlai to regain their focus. | Riley always gets things done before the deadline while having a problem to start a new task early. | Even if Cleo has a whole month to work on an essay, Cleo will still finish it by the first week. |
| *Responsibility* | Adlai’s boss always finds spelling and grammar mistakes in their work-related emails. | Riley never makes major mistakes at work but is also known to be careless with minor things. | Their supervisor considers Cleo as the first choice to be the one responsible for managing an event. |
| **Vignette set 3** | *Organization* | Eli never cleans their apartment until it is too dirty to live in. | Sawyer cleans the living room frequently, but Sawyer’s bedroom is a mess most of the time. | Tobin makes a list of housework for the week and follows the daily tasks strictly to keep the apartment tidy and clean. |
| *Productiveness* | Eli is used to taking a long lunch and to leaving the office as early as possible. | Sawyer’s Coworkers think Sawyer is generally productive but not initiative enough. | Tobin is the person who assigns the duty to each team member and keeps checking in on everyone’s progress regularly. |
| *Responsibility* | Eli seldom replies to coworkers’ work-related messages and emails. | Sawyer can get things done but always make last minute changes to the group projects. | Tobin’s friends always ask Tobin to remind them in case they forgot something. |
| **Vignette set 4** | *Organization* | Tanner has difficulties to find the emails people sent a few days before since Tanner never separates important emails from the rest when they receive them. | Jacky’s personal space is not tidy but at least Jacky keeps everything important in a certain place. | Yael likes to keep every document in a clear-categorized folder. Yael can find useful documents they received years ago without a problem. |
| *Productiveness* | Tanner needs a very hands-on advisor to constantly tell Tanner what to do within a certain amount of time. Otherwise, Tanner never finishes a single job. | Jacky is lazy in personal life but finishes jobs efficiently. | Yael enjoys following the to-do list and checks off each task they have done. Yael can finish almost every task they planned to do for that day. |
| *Responsibility* | Tanner always makes up excuses so that they do not have to take responsibility for tasks that are not directly assigned to Tanner. | Jacky only cares about jobs that interest them. Sometimes, Jacky will get jobs done half-heartedly as long as it does not lead to major issues. | Yael always ask other people to proof-read their important emails and writings to avoid any small mistake. |
| **Vignette set 5** | *Organization* | There are numbers of folders on the desktop of Raphael’s laptop, and Raphael feels anxious when trying to find the file they want. | Piper is a college student who is usually clean and neat. However, he is a mess and would not even remember to take out the trash during final weeks. | Quinn is the person who would rather clean the common space frequently for their roommates than bearing with the messy living space. |
| *Productiveness* | Raphael is constantly on Facebook during work hours. | Piper can keep along with a tight work schedule at the beginning of the semester but gradually loosens up at the end of the semester. | Quinn can finish double the work within the same amount of time compared to their coworkers. |
| *Responsibility* | Raphael always asks for a lot of duties but fails to finish a single one of them at the end. | Piper is a responsible group mate when there is only one thing at hand. Piper will miss group deadlines when they are multi-tasking. | Quinn prioritizes group projects over their own tasks. Quinn said they hate to let other people wait because of one’s own issues. |

**Agreeableness:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Facet** | **Low (1)** | **Average (3)** | **High (5)** |
| **Vignette set 1** | *Compassion* | People generally think Ivy is mean after the first encounter. | Joey is very caring in front of family and friends but not so much towards strangers. | Jean does charity on a regular basis. |
| *Respectfulness* | Ivy generally does not care about what other people say and feel. | Joey is very polite to elderly, authorities, and those who Joey respects. Joey is less considerate to peers sometimes. | Jean respects other’s effort even if the outcomes are imperfect. |
| *Trust* | Ivy believes that everyone is selfish all the time because this is the human nature. | Joey is a little picky about others’ mistakes and is suspicious about people’s intentions when they consistently make the same mistake. | Jean believes that people all have good nature but sometimes make mistakes due to external factors they cannot control over. |
| **Vignette set 2** | *Compassion* | Sage believes that having a soft heart is a weakness. | Hayden helps friends when they are truly in need and when they are affordable because Hayden believes one can only help limited people. | Rio’s friends and peers feel comfortable to tell Rio their concerns and difficulties because Rio is always willing to help and can understand their feelings. |
| *Respectfulness* | Sage wouldn’t feel bad to start arguments with others no matter the circumstances. | Hayden respects people who are hard-working but look down on people Hayden thinks are lazy. | Rio is willing to take others’ perspectives even though Rio does not agree with some of their ideas. |
| *Trust* | Sage believes it is best to finish the most important tasks by themselves to prevent other people messing them up. | Hayden trusts people who Hayden has been working with for a while. However, Hayden tends to keep an eye on newcomers. | Rio always makes sure to compliment every teammate even when some of them made mistakes. |
| **Vignette set 3** | *Compassion* | Rudy cannot understand why bother with killing livestock in a more humane way. | Shiloh is a customer server who does the job to help the customer well, but Shiloh is not a very caring person off the job. | Skyler has three adopted dogs because Skyler cannot allow them to be put down. |
| *Respectfulness* | Rudy never says “thank you” to a customer server because Rudy thinks they are paid to do their job. | Shiloh makes sure to use polite languages with customers, while sometimes is a little rude towards coworkers. | Skyler always offer compliments before giving critiques because Skyler wants to acknowledge people’s work. |
| *Trust* | Rudy feels it is hard to forgive others because people should take responsibility for their own mistakes. | Shiloh believes that most people are born to be good while a few is born to be evil. There are some people that just cannot be trusted. | Skyler feels comfortable to ask strangers sitting next to them to look after their belongings when leaving to the restroom. |
| **Vignette set 4** | *Compassion* | Lennox generally does not care about others’ business. | Val has no problem with saying “no” to people if Val wants to. | Emerson cries when watching documentaries about disasters and wars. |
| *Respectfulness* | Lennox thinks it is important to make their own points clear than making other people feel comfortable during the conversation. | Val sometimes argues with people just for showing them Val is right. | Emerson never curses in public but seldom does when being alone or with close friends. |
| *Trust* | Lennox often comments on people’s shortcomings because it is easier to see faults with others. | Val would only let people to borrow a small amount of money without a receipt. | Emerson feels sorry for the criminals and thinks they should be given a second chance. |
| **Vignette set 5** | *Compassion* | Most people do not want to travel with Kylar because it is hard to make Kylar compromise to other people’s needs and interests. | Sean likes to help others because Sean wants them to be thankful and to think Sean is a moral person. | Sam always asks friends to send them a message when get home safely if leaving at night. |
| *Respectfulness* | Kylar is very cautious to avoid giving negative comments to prestigious people but very rude toward peers and subordinates. | Sean is a nice person in general but sometimes makes inappropriate jokes on others only because Sean finds it funny. | Sam bows or nods whenever they meet their professors. |
| *Trust* | Kylar feels selfless people are hypocritical or fake. | Sean assumes the best about people when Sean is in a good mood. When Sean is anxious, Sean cannot help but to think coworkers are trying to drag them down. | Sam thinks people have no reason to hurt others, but if they do, there must be a reason behind that forced them to do so. |

**Open-Mindedness:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Facet** | **Low (1)** | **Average (3)** | **High (5)** |
| **Vignette set 1** | *Intellectual Curiosity* | Reese believes that higher education is a waste of time because Reese doesn’t care about learning things that are unnecessary in daily life. | Billie sometimes engages in conversations about future technology or history with friends but has little interest to go deep into those topics. | Xoan enjoys deep discussions and always initiates them with friends. |
| *Aesthetic Sensitivity* | Reese hates poetry and abstract arts. | Billie played piano as a child. Billie has stopped playing piano for a very long time now. | Xoan can keep talking about music and novels for hours. |
| *Creative Imagination* | Compared to coming up with new ideas, Reese feels easier to follow the old ways. | Billie sometimes daydreams, but thinks daydreaming is a waste of time. | Xoan works as a graphic designer in the advertisement industry which allows Xoan to express their creativity. |
| **Vignette set 2** | *Intellectual Curiosity* | Ali knows almost nothing about philosophical ideas because they are not interested in them. | River is curious about things relevant to River’s daily life, but not so much about international political issues which River has no control over. | Indigo took several extra philosophy courses in college and enjoyed them very much. |
| *Aesthetic Sensitivity* | Ali cannot name any famous artists and their master works. | River can name a few artists and their styles but has no opinion on which one is better. | During travel, Indigo always goes to local museums and theaters. |
| *Creative Imagination* | Compared to finding a clever way to do things, as long as things get done, Ali prefer to stick to the old ways. | River is willing to find new ways to do things when River is asked to do so or has nothing else to do. | Indigo enjoys finding multiple ways to reach the same results. |
| **Vignette set 3** | *Intellectual Curiosity* | Harley never seriously thinks about life goals, self-identity, or reflects on the past experiences. | Greer likes self-reflective thinking but avoids thinking deeply about religious belief. | Lyric writes reflective notes while reading a book. |
| *Aesthetic Sensitivity* | Harley thinks it is meaningless to value art and beauty. | Greer thinks arts are good for leisure but not as essential as math and science. | When Lyric was in college, they registered for several literature courses despite of majoring in chemistry. |
| *Creative Imagination* | Harley thinks that new ideas are generally bad ideas. | When Greer comes up with new ideas, they often hesitate to apply them unless no other way works. | Lyric enjoys the process of figuring things out more than the solution itself. |
| **Vignette set 4** | *Intellectual Curiosity* | Justice never openly talks about political and religious beliefs with other people. | Jael is curious to learn new things only when it is easy and effortless to do. | Harlow’s parents said Harlow is always very curious like a child. |
| *Aesthetic Sensitivity* | Justice believes that arts are for a small amount of people with special talents only. | Jael only practice drawing in leisure time, although they enjoy it very much, Justice still does not think being a painter is an appropriate occupation. | Harlow enjoys going camping because Harlow enjoys the beauty of nature. |
| *Creative Imagination* | Justice never consider themselves as a creative person. | Jael likes creative ideas from others but seldom comes up with one by themselves. | Harlow really likes fantasy novels. |
| **Vignette set 5** | *Intellectual Curiosity* | Casey never hangs out with nerds because philosophical discussions are boring. | Drew enjoys listening to others’ philosophical ideas but feels it is difficult to come up with original thoughts. | Eden writes self-reflected journals every day. |
| *Aesthetic Sensitivity* | Casey cannot differentiate the work of a professional photographer and a layman. | Drew enjoys popular music and famous paintings, but never tries to discover niche singers or painters. | Eden admires the beauty of wildflowers and landscapes. |
| *Creative Imagination* | Casey has difficulty thinking of a new solution without others telling Casey what to do. | Drew daydreamed a lot as a kid but seldom imagines things after growing up. | Eden keeps a notebook at hands for scrawling and writing down random ideas. |

**Extraversion:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Facet** | **Low (1)** | **Average (3)** | **High (5)** |
| **Vignette set 1** | *Sociability* | Dylan’s friends often complain that Dylan is too quiet in groups. | Spencer can be talkative if he needs to. | Finley is always energetic and talkative. |
| *Assertiveness* | When Dylan speaks up, Dylan feels most people don’t take their words seriously. | Spencer sometimes takes leadership roles but is relieved if someone else would like to take the responsibility. | Finley is very certain about personal opinions and is capable to convince others. |
| *Energy Level* | Dylan is not a passionate person. | Spencer actively work on the projects that feels interesting but put less energy on the ones Spencer doesn’t care that much about. | Finley always wakes up with a list of things to do and is enthusiastic about every day. |
| **Vignette set 2** | *Sociability* | Blair is the person who plays with their phone the whole time at parties, | Avis needs some alone time after going out with friends. | Friends would always invite Finley to parties and group events because they always say yes to going. |
| *Assertiveness* | Blair always makes others to make the final decision. | Avis knows what to do but still prefers to have others take charge. | Finley always makes final decisions to group events/club activities. |
| *Energy Level* | Blair cannot think of a life goal or something Blair really wants. | Avis often feels excited but seldom shows their enthusiasm in front of other people. | Finley is very passionate about their job and is willing to work extra hours on the task until finishing it. |
| **Vignette set 3** | *Sociability* | Kelly always speaks in a low volume and tends to avoid eye contact when talking with people. | Ellery is only talkative around family and close friends. | Joe needs to have lunch with coworkers every day, otherwise Joe feels lonely. |
| *Assertiveness* | Kelly tends to be the follower in a group and seldom speaks up. | Although Ellery likes taking initiatives, Ellery does not mind stepping back and taking people’s suggestions. | People consider Joe as a very persuasive person. |
| *Energy Level* | Kelly does not have any favorite sports or outdoor activities. | Ellery tries to avoid any social events that are later than 9 PM. | Joe is always eager to try out new things. |
| **Vignette set 4** | *Sociability* | Kaden seldom initiates conversation with people. | Merit can be talkative around the right groups of people but keeps quiet when Merit thinks there is nothing to talk about. | Kennedy is always the spokesperson in a group and finds it easy to share their opinions. |
| *Assertiveness* | Kaden changes their ideas or opinion easily when someone else says something differently. | Merit has strong opinions but feels it is hard to convince other people. | Kennedy has a strong opinion on their profession and never steps back when facing criticism. |
| *Energy Level* | Kaden always looks tired. | Merit is enthusiastic over a lot of things such as learning a new language, painting, and playing violin. However, Merit can only keep enthusiastic for a short period of time. | Kennedy often works overnight on the projects they are enthusiastic about. |
| **Vignette set 5** | *Sociability* | Lumi always answers others’ question within one or two sentences. | Ode only talks when the topic is important or when Ode feels it is safe to speak up. | Frankie gains energy when hanging out with people. |
| *Assertiveness* | Lumi does not know what to do next and always tries to follow what the majority is doing. | Ode has clear opinions but prefers to follow other people just to avoid conflicts. | Frankie always volunteers to be the leader. |
| *Energy Level* | Lumi never voluntarily tries new things. Lumi prefers to watch other people exploring things. | Ode likes to be around enthusiastic people so Ode can be motivated by them. | Frankie gets excited over little things easily and cannot hold back from jumping or moving around. |

**Negative Emotionality:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Facet** | **Low (1)** | **Average (3)** | **High (5)** |
| **Vignette set 1** | *Anxiety* | Merle is very calm even under a stressful situation. | Zane feels very nervous before giving speeches. | Jordan worries about every little thing that can go wrong. |
| *Depression* | Merle is satisfied with the current life and can move on from a setback very quickly. | Zane thinks permissively from time to time, and it takes several days to recover from a failed exam. | Jordan feels sad very easily and can hardly think optimistically about little failures in life. |
| *Emotional Volatility* | They are comfortable with making kindly jokes because Merle never acts too sensitive. | Zane would let friends know when they feels moody, and friends understand it because Zane only shows extreme negative emotions seldomly. | Jordan can hardly keep calm because Jordan is too sensitive to people’s words and expressions. |
| **Vignette set 2** | *Anxiety* | London seldom feels tense even during the job interview. | Moriah would feel worry when starting to speak in public but quickly calms down. | West sets multiple alarms every day to avoid being late. |
| *Depression* | London’s friends like hanging out with London because of London’s positive and sunny personality. | Moriah is a cheerful person especially when spending time with family and friends. However, Moriah feels quite upset alone and Moriah has to call someone to ease the feeling. | West does not like the appearance of themselves and often thinks about going on a diet. |
| *Emotional Volatility* | Friends almost never see London being moody. | Moriah is able to keep emotions under control to the extent that the mood swings never lead to impulsive behaviors. | West’s friends feel they have to be cautious when saying negative things because West can be easily upset. |
| **Vignette set 3** | *Anxiety* | Nicky never worries about things to the extent that they cannot fall asleep. | Cody cannot stop worrying about their young daughter who is very shy. | Alex is a college student who always feels stressful during the final weeks which causes temporary sleep problems. |
| *Depression* | Nicky sees themselves as imperfect but definitely worth others’ love. | Cody has no problem with staying positive after experiencing a setback only if supported by family and friends. | Alex can easily feel insecure. Alex feels people are constantly judging them. |
| *Emotional Volatility* | Nicky feels it is easy to control their emotions under different circumstances if needed. | Cody’s family and close friends never see Cody cry even though Cody sometimes hides and cries alone. | Alex will cry if people point out their mistakes in public. |
| **Vignette set 4** | *Anxiety* | Cypress’s partner sometimes complains that Cypress is too carefree all the time and it seems Cypress never cares about anything. | Pat can be tense over little things such as talking to the boss, but Pat can quickly calm themselves down once the conversation starts to flow naturally. | Perry keeps worrying about forgetting to lock the door or feed the cat even though Perry is always very careful. |
| *Depression* | Cypress can easily find something positive out of a bad situation. | Pat can feel depressed from time to time, but Pat knows how to get out of the negative emotions. For example, Pat goes to the zoo to distract themselves. | Perry sometimes believes that they are incapable of doing what they love doing. Perry feels they have no talent to achieve success. |
| *Emotional Volatility* | Cypress can take overdone jokes from their friends. | Pat sometimes feels moody but believes it is normal to have the mood swings that are not strong enough to affect daily activities. | Perry does not know how to control their emotions. Perry hates to cry in front of people but always end up with doing it. |
| **Vignette set 5** | *Anxiety* | Shae believes being worrisome does not change the result so there is no need to be anxious under any circumstances. | Wyatt worries a lot about if things may go wrong if Wyatt is not fully prepared for their duty. | Tate constantly worries about anything that can go wrong. Tate prefers not to drive because of the anxious feeling during driving. |
| *Depression* | Shae is confident with who they are because everyone has shortcomings. | Wyatt sometimes feels sad during evenings or early mornings, but this feeling would not last very long. | When Tate is in the low mood, Tate cannot get up off the bed and end up skipping meals. |
| *Emotional Volatility* | Shae is able to give warnings before expressing anger or frustration towards others. | Wyatt is generally a stable person but can be very moody when traveling with friends. | Tate always feels regret because Tate cannot keep their emotions under control and always ends up shouting at Tate’s parents. |

**Appendix E: Demographic Questions**

1. Age
2. Gender (Woman, Man, No binary/Genderqueer/Trans)
3. Ethnicity (White/Caucasian, Black/African American, Hispanic/Latino, Native American/Alaskan Native, Pacific Islander, Asian/Asian American, Other/Multiracial)
4. Class standing (current overall 4-point GPA)
5. Native language (English/including English, Non-English)
6. Parental/household income (unit: $1000): <20, 20-34.999, 35-49.999, 50-74.999, 75-99.999, >100